CIVIL ENGINEERING

Hydrosystems Engineering Reliability Assessment and Risk Analysis

✔ Tools required to design and predict the safety of various hydrosystems

✔ First book to integrate reliability analysis and risk management

✔ Dual units used in all problems

Dr. Yeou-Koung Tung, Dr. Ben-Chie Yen and Charles S. Melching

Hydrosystems Engineering Reliability Assessment and Risk Analysis

Yeou-Koung Tung, Ph.D.

Department of Civil Engineering
Hong Kong University of Science and Technology
Kowloon, Hong Kong

Ben-Chie Yen, Ph.D.

Late Professor
Department of Civil and Environmental Engineering
University of Illinois at Urbana–Champaign
Urbana, Illinois

Charles S. Melching, Ph.D.

Department of Civil and Environmental Engineering
Marquette University
Milwaukee, Wisconsin

McGraw Hill **Professional**

Want to learn more?

We hope you enjoy this McGraw-Hill eBook! If you'd like more information about this book, its author, or related books and websites, please click here.

To humanity and human welfare

God understands the way to wisdom and He alone knows where it dwells.

— Job 28:23

Contents

This page intentionally left blank

Preface

Failures of major engineering systems always raise public concern on the safety and reliability of engineering infrastructure. Decades ago quantitative evaluations of the reliability of complex infrastructure systems were not practical, if not impossible. Engineers had to resort to the use of a safety factor mainly determined through experience and judgment. The contribution of human factors to structural safety still remains elusive for analytical treatment. The main areas of concern and application in this book are hydrosystems and related environmental engineering.

Without exception, failures of hydrosystem infrastructure (e.g., dams, levees, and storm sewers) could potentially pose significant threats to public safety and inflict enormous damage on properties and the environment. The traditional approach of considering occurrence frequency of heavy rainfalls or floods, along with an arbitrarily chosen safety factor, has been found inadequate for assessing the reliability of hydrosystem infrastructure and for risk-based cost analysis and decision making. In the past two decades or so, there has been a steady growth in the development and application of reliability analysis in hydrosystems engineering and other disciplines. The main objective of the book is to bring together some of these developments and applications in one volume and to present them in a systematic and understandable manner to the water resource related engineering profession. Through this book it is hoped to demonstrate how to integrate involved physical processes, along with some knowledge in mathematics, probability, and statistics, to perform reliability assessment and risk analysis of hydrosystem engineering problems. An accompanying book, *Hydrosystems Engineering Uncertainty Analysis*, provides treatments and quantifications of various types of uncertainty, which serve as essential information needed for the reliability assessment and risk analysis of hydrosystems.

Hydrosystems is the term used to describe collectively the technical areas of hydrology, hydraulics, and water resources. The term has now been widely used to encompass various water resource systems including surface water storage, groundwater, water distribution, flood control, drainage, and others. In many hydrosystem infrastructural engineering and management problems, both quantity as well as quality aspects of water and other environmental

issues have to be addressed simultaneously. Due to the presence of numerous uncertainties, the ability of the system to achieve the goals of design and management decisions cannot be assessed definitely. It is almost mandatory for an engineer involved in major hydrosystem infrastructural design or hazardous waste management to quantify the potential risk of failure and the associated consequences.

Application of reliability analysis to hydrosystems engineering covers a wide scope of subfields, ranging from data collection and gauging network design to turbulence loading on structures; and from inland surface water to groundwater to coastal water. In terms of the system scale, it could involve entire river basins containing many components, or a large dam and reservoir, or a single culvert or pipe. Depending on the objective, the application could be for designing the geometry and dimension of hydraulic facilities, for planning of a hydraulic project, for determining operation procedure or management strategy, for risk-cost analysis, or for risk-based decision making.

The book is not intended to be a review of literature, but is an introduction for upper level undergraduate and graduate students to methods applicable for reliability analysis of hydrosystem infrastructure. Most of the principles and methodologies presented in the book can equally be applied to other civil engineering disciplines. The book presents relevant theories of reliability analysis in a systematic fashion and illustrates applications to various hydrosystem engineering problems. Although more advanced statistical and mathematical skills are occasionally required, the great majority of the problems can be solved with basic knowledge of probability and statistics. Illustrations in the book bring together the use of probability and statistics, along with knowledge of hydrology, hydraulics, water resources, and operations research for the reliability analysis and optimal reliability-based design of various hydrosystem engineering problems. The book provides added dimensions to water resource engineers beyond conventional frequency analysis.

The book consists of eight chapters. In each chapter of the book, ample examples are given to illustrate the methodology for enhancing the understanding of the materials. The book can serve as an excellent reference book not only for engineers, planners, system analysts, and managers in area of hydrosystems, but also other civil engineering disciplines. In addition, end-of-chapter problems are provided for practice and homework assignments for classroom teaching.

The book focuses on integration of reliability analysis with knowledge in hydrosystems engineering with applications made to hydraulics, hydrology, water resources, and occasionally, to environmental and water quality management related problems. Since many good books on basic probability, statistics, and hydrologic frequency analysis have been written, background in probability, statistics, and frequency analysis that are relevant to reliability analysis are summarized in Chapters 2 and 3, respectively. The book, instead of dwelling on the subject of data analysis, focuses on how to perform reliability analysis of hydrosystem engineering problems once relevant statistical data analysis has been conducted. As real-life hydrosystems generally involve

various uncertainties other than just inherent natural randomness of hydrologic events, the book goes beyond conventional frequency analysis by considering reliability issues in a more general context of hydrosystems engineering and management. Chapter 4 elaborates the reliability analysis methods considering load-resistance interaction under the static and time-dependent conditions. First-order and second-order reliability methods, with the emphasis given to the former, are derived. For many hydrosystem infrastructures, it is sometimes practical to treat the system as a whole and analyze its performance over time without considering detailed load-resistance interaction. Chapter 5 is devoted to time-to-failure analysis that is particularly useful for dealing with systems that are repairable. Chapter 6 provides a detailed treatment of using Monte Carlo simulation and its variations applicable to reliability analysis. The subject, in most books, is covered in the context of univariate problems in which stochastic variables are treated as independent and uncorrelated. In reality, the great majority of the hydrosystem infrastructural engineering problems involve multiple stochastic variables, which are correlated. Treatment of such problems is emphasized. Chapter 7 focuses on the evaluation of system reliability by integrating load-resistance reliability analysis methods or time-to-failure analysis, along with system configuration, for assessing system reliability. Different methods for system reliability analysis are presented and demonstrated through examples. Chapter 8 presents the framework that integrates uncertainties, risk, reliability, and economics for an optimal design of hydrosystem infrastructure. A brief description of system optimization is also given.

The intended uses and audiences for the book are: (1) as a textbook for an intermediate course at the undergraduate senior level or graduate level in water resources engineering on the risk and reliability related subjects; (2) as a textbook for an advanced course in risk and reliability analysis of hydrosystem engineering; and (3) as a reference book for researchers and practicing engineers dealing risk and reliability issues in hydrosystems engineering, planning, management, and decision making.

The expected background for the readers of this book is a minimum of 12 credits of mathematics including calculus, matrix algebra, probability, and statistics; a one-semester course in elementary fluid mechanics; and a one-semester course in elementary water resources covering basic principles in hydrology and hydraulics. Additional knowledge on engineering economics, water-quality models, and optimization would be desirable.

Two possible one-semester courses could be taught from this book depending on the background of the students and the type of course designed by the instructor. Instructors can also refer to the accompanying book *Hydrosystems Engineering Uncertainty Analysis* for other relevant materials to compliment this book. The possible course outlines are presented below.

Outline 1. (For students who have taken a one-semester probability and statistics course). The objective of this outline aims at achieving higher level of capability to perform reliability analysis. The optimal risk-based design

concept can be introduced without having to formally cover subjects on optimization techniques. The subject materials could include Chapter 1, Chapter 2 (2.7), Chapter 3, Chapter 4 (4.1–4.4), Chapter 5 (5.1–5.3), Chapter 6 (6.1–6.4, 6.6), Chapter 7 (7.1–7.3), and Chapter 8 (8.1–8.4).

Outline 2. (For water resource engineers or students who have a good understanding in basic statistics, probability, and operations research.) The aim of this outline is for readers to achieve higher level and deeper appreciation of the applications of reliability assessment techniques in hydrosystems engineering. The topics might include Chapters 1, 4, 5, 6, 7, and 8.

The uncertainty and reliability issues in hydrosystem engineering problems have been attracting a lot of attention of engineers and researchers. A tremendous amount of progress has been made in the area. This book, and the accompanying book *Hydrosystems Engineering Uncertainty Analysis*, merely represent our humble offer to the hydrosystem engineering community. We hope that readers will find this book useful and enjoyable. Due to our limited knowledge and exposure in the exciting area of stochastic hydraulics, we are unable to incorporate many brilliant works in this book. It is our sincere wish that this effort will bring out much greater works from others to improve and enhance our contribution to society and mankind.

Acknowledgments

Through my academic career, I have spent most of research efforts on problems relating to probabilistic hydrosystem engineering. I am truly thankful to my advisor, Larry W. Mays, who first introduced me to this fascinating area when I was a Ph.D. student. Over the years, both Larry and the late Ben C. Yen have been my unflagging supporters and mentors. In the process of putting together the book, the use of materials from some of my former students (Drs. Wade Hathhorn, Yixing Bao, and Bing Zhao) brought many fond memories back about the time we spent together burning midnight oil, cutting fire wood, and fishing. Many of my more recent students (Chen Xingyuan, Lu Zhihua, Wang Ying, Eddy Lau, and Wu Shiang-Jen) have contributed their kind assistance in preparing figures and tables, reading manuscripts and offering their criticisms from a student's perspective. I am also grateful to Ms. Queenie Tso for skillful typing of numerous equations and painstakingly performing necessary corrections in the book. Especially, I would like to express sincere gratitude to my dear friend Ms. Joanne Lam for her prayers and encouragements during the course of writing this book.

Although writing this book has been a very rewarding experience, it nevertheless has occupied many hours and attention that I should have spent with my family. I am grateful to my wife Be-Ling and daughters (Fen, Wen, Fei, and Ting) for their understanding and support without which the completion of this book would not have been possible. By the time the final manuscript was submitted, I felt an overwhelming sense of sadness and loss since I wished Prof. Yen would have lived to see the completion of this book. I want to thank Ruth Yen for her encouragement to continue with the work. Also I am much obliged to Steve Melching for his willingness to work with me on the book. Looking back, I see how kind God has been to me. He blesses me by surrounding me with so many people who do not hold back their support, kindness, and love. I praise the Lord that through His mercy and grace the book is completed.

Last, but not the least, I am thankful to McGraw-Hill for supporting the publication of the book, to Mr. Larry Hager for his advice in preparing the book, and, in particular, to Samik Roy Choudhury (Sam) and his team at International Typesetting and Composition for editorial and production efforts.

Yeou-Koung Tung

I would like to take this opportunity to most sincerely thank my coauthors. The late Prof. Ben C. Yen was my Ph.D. advisor, mentor, and friend, and the greatest influence on my life after my parents and my Lord Jesus Christ. Professor Yen led me down the path of the study of uncertainty and reliability in hydrosystems engineering in my Ph.D. work and we worked together on many related projects throughout my professional life. Professor Y. K. Tung invited me to get involved in this book after Prof. Yen's untimely death, initially to be a second pair of eyes to ensure that the concepts were clear, concise, and correct. Eventually my small contribution grew enough that Y. K. honored me with a coauthorship.

I also would like to thank my wife, Qiong, and my children, Christine and Brian, for their patience while I hide in the basement on evenings and weekends working on this book. I also thank my former students, Satvinder Singh, Sharath Anmangandla, Chun Yoon, and Gemma Manache, whose work on uncertainty analysis gave me additional insight that is part of my contribution to this book.

Charles S. Melching

Reliability in Hydrosystems Engineering

1.1 Reliability Engineering

Occasionally, failures of engineering systems catch public attention and raise concern over the safety and performance of the systems. The cause of the malfunction and failure could be natural phenomena, human error, or deficiency in design and manufacture. Reliability engineering is a field developed in recent decades to deal with such safety and performance issues.

Based on their setup, engineering systems can be classified loosely into two types, namely, manufactured systems and infrastructural systems. *Manufactured systems* are those equipment and assemblies, such as pumping stations, cars, computers, airplanes, bulldozers, and tractors, that are designed, fabricated, operated, and moved around totally by humans. *Infrastructural systems* are the structures or facilities, such as bridges, buildings, dams, roads, levees, sewers, pipelines, power plants, and coastal and offshore structures, that are built on, attached to, or associated with the ground or earth. Most civil, environmental, and agricultural engineering systems belong to infrastructural systems, whereas the great majority of electronic, mechanical, industrial, and aeronautical/aerospace engineering systems are manufactured systems.

The major causes of failure for these two types of systems are different. Failure of infrastructures usually is caused by natural processes, such as geophysical extremes of earthquakes, tornadoes, hurricanes or typhoons, heavy rain or snow, and floods, that are beyond human control. Failure of such infrastructural systems seldom happens, but if a failure occurs, the consequences often are disastrous. Replacement after failure, if feasible, usually involves so many changes and improvements that it is essentially a different, new system.

On the other hand, the major causes of failure for manufactured systems are wear and tear, deterioration, and improper operation, which could be dealt with by human abilities but may not be economically desirable. Their failures

usually do not result in extended major calamity. If failed, they can be repaired or replaced without affecting their service environment. Their reliability analyses are usually for production, quality control, or for maintenance service and warranty planning. Thus failures of manufactured systems often are classified into repairable and nonrepairable types. Conversely, failures of infrastructural systems can be classified as structural failures and functional failures, as will be explained in Sec. 1.5.

The approaches and purposes of reliability analysis for these two types of systems are related but different. As described in Sec. 1.3, reliability analysis for manufactured systems has a history of more than 70 years and is relatively more developed than reliability analysis for civil engineering infrastructural systems. Many books and papers have been published on reliability engineering for manufactured systems. One can refer to Ireson and Coombs (1988), Kececioglu (1991), Ushakov (1994), Pecht (1995), Birolini (1999), and Modarres et al. (1999) for extensive lists of the literature. Conversely, this book deals mainly with reliability issues for hydrosystem engineering infrastructures. Nonetheless, it should be noted that many of the basic theories and methods are applicable to both systems.

1.2   Reliability of Hydrosystem Engineering Infrastructure

The performance of a hydrosystem engineering infrastructure, function of an engineering project, or completion of an operation all involve a number of contributing components, and most of them, if not all, are subject to various types of *uncertainty* (Fig. 1.1). Detailed elaboration of uncertainties in hydrosystem engineering and their analysis are given in Tung and Yen (2005). Reliability and risk, on the other hand, generally are associated with the system as a whole. Thus methods to account for the component uncertainties and to combine them are required to yield the system reliability. Such methods usually involve the use of a logic tree, which is discussed in Chap. 5. A typical logic tree for culvert design is shown in Fig. 1.2 as an example.

The reliability of an engineering system may be considered casually, such as through the use of a subjectively decided factor of safety (see Sec. 1.6). Today, reliability also may be handled in a more comprehensive and systematic manner through the aid of probability theory. Factors that contribute to the slow development and application of analyses of uncertainty and reliability in hydrosystem engineering infrastructure design and analysis include the following:

1. Those who understand the engineering processes well often are not trained adequately and are uncomfortable with probability. Contrarily, those who are good in probability theory and statistics seldom have sufficient knowledge of the details of the engineering process involved.
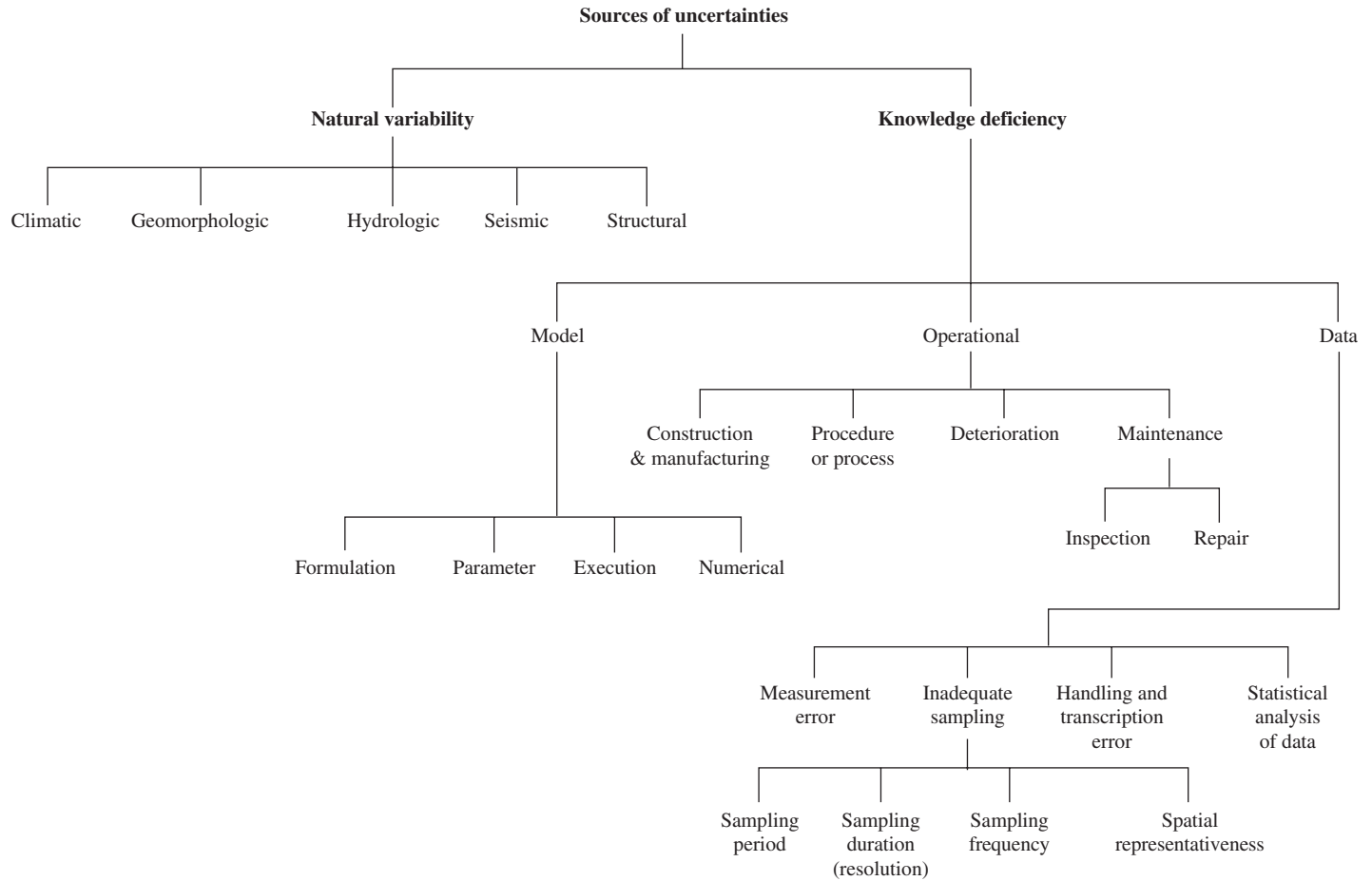
Figure 1.1 Sources of uncertainty. (*After Tung and Yen, 2005.*)
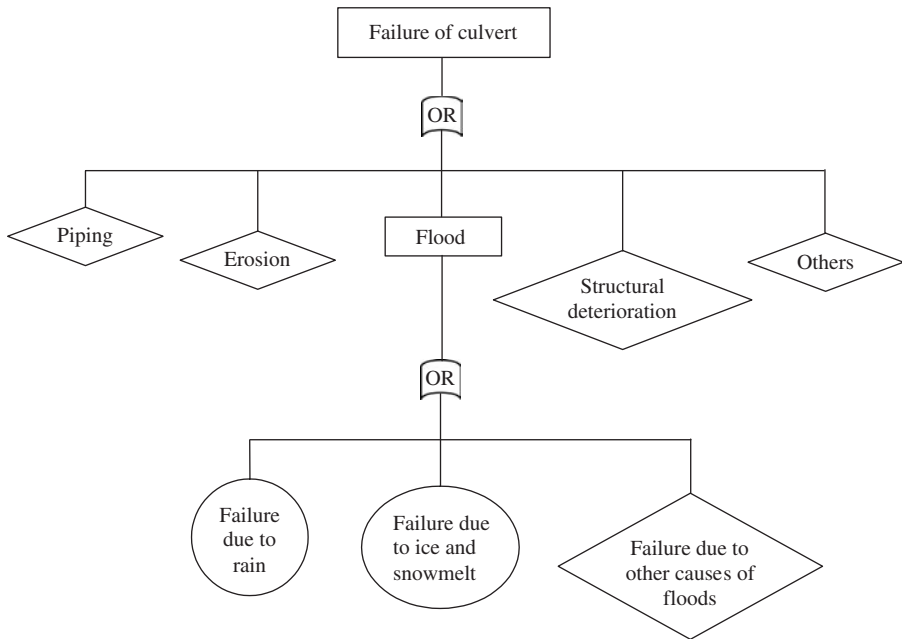
Figure 1.2 Fault tree for culvert design.

2. Many factors contribute to the reliability of an engineering system. Only recently have advances in techniques and computers rendered the combination and integration of these contributions feasible to evaluate the system reliability. Nevertheless, some of the factors are still beyond the firm grasp of engineers and statisticians. Furthermore, these factors usually require the work of experts in different disciplines, whereas interdisciplinary communication and cooperation often are a problem.

3. Engineers have a tendency to focus on components affecting their problem most while ignoring other contributing elements. For instance, hydrologists as a group perhaps have contributed more than any other discipline to frequency analysis and also have made major contributions to related probability distributions. Yet their devotion and accomplishment are a blessing as well as a curse, in that they hinder the vision to see beyond to a broader view of uncertainty and reliability analyses. As noted by Cornell (1972):

> It is important to engineering applications that we avoid the tendency to model only those probabilistic aspects that we think we know how to analyze. It is far better to have an approximate model of the whole problem than an exact model of only a portion of it.

Only more recently, uncertainties other than natural randomness of floods/rainfalls are considered in reliability-based design of flood mitigation schemes (U.S. National Research Council, 2000).

4. Inconsistent definitions of risk and risk analysis cause considerable confusion and doubt about the subject. For example, in flood protection engineering, hydraulic engineers tend to accept the definition used by structural, aerospace, and electronic engineers that risk analysis is the analysis of the probability of failure to achieve the intended objectives. Hydrologists often consider risk in terms of the *return period*, which is considered as the reciprocal of the annual exceedance probability of the hydrologic events (i.e., flood, storm, or drought). Water resources planners and decision makers mostly adopt the definition used in economics and the health science fields, regarding risk analysis as the analysis of risk costs, assessment of the economic and social consequence of a failure, and risk management. For example, the United Nations Department of Humanitarian Affairs (1992) defines risk as

> The expected losses (of lives, persons injured, property damaged and economic activity disrupted) due to a particular hazard for a given area and reference period. Based on mathematical calculations, risk is the product of hazard and vulnerability.

Further, *hazard* is defined as "a threatening event or the probability of occurrence of a potentially damaging phenomenon within a given time period and area." Hence, in the United Nations terminology, hazard is what engineers define as risk. The problem of confusion probably would be minimized if the experts in these subdisciplines worked separately, each responsible for his or her own specialty. However, the trend of the past decades, expecting jack-of-all-trades water resources engineers to be experts in all these subdisciplines, bears significant undesirable consequences, a small one of which is the confusion concerning the definition of risk.

Practically all hydrosystem engineering infrastructures placed in a natural environment are subject to various external stresses and loads. The *resistance, strength, capacity,* or *supply* of the system is its ability to accomplish the intended mission satisfactorily without failure when subjected to demands or external stresses. *Loads, stresses,* and *demands* tend to cause failure of the system. Failure occurs when the demand exceeds the supply or the load exceeds the resistance. Owing to the existence of uncertainties, the capacity of an infrastructural system and the imposed loads more often than not are random and subject to some degrees of uncertainty. Hence the design and operation of engineering systems are always subject to uncertainties and potential failures.

Nevertheless, engineers always face the dilemma of decision making or design with imperfect information. It is the engineer's responsibility to obtain a solution with limited information, guided by experience and judgment, considering the uncertainties and probable ranges of variability of the pertinent factors, as well as economic, social, and environmental implications, and assessing a reasonable level of safety.

1.3   Brief History of Engineering Reliability Analysis

Development of engineering reliability analysis started with the desire for product quality control in manufacturing engineering three-quarters of a century ago (Shewart, 1931). World War II considerably accelerated its advancement. During the war, over 60 percent of airborne equipment shipped to the Far East arrived damaged. About half the spares and equipment in storage became unserviceable before use. Mean service time before requiring repair or replacement for bomber electronics was less than 20 hours. The cost of repair and maintenance exceeded 10 times the original cost of procurement. About two-thirds of radio vacuum tubes in communications devices failed. In response to the high failure rates and damage to military airborne and electronic equipment, the U.S. Joint Army-Navy Committees on Parts Standards and on Vacuum Tube Development were established in June 1943 to improve military equipment reliability. However, when the Korean War began, about 70 percent of Navy electronic equipment did not function properly. In 1950, the U.S. Department of Defense (DOD) established an Ad Hoc Group on Reliability that was upgraded in November 1952 as the Advisory Group on the Reliability of Electronic Equipment (AGREE) to monitor and promote military-related reliability evaluation and analysis.

Meanwhile, the civilian-side activities on reliability engineering also became active in aeronautical engineering (Tye, 1944) and in communications. In 1949–1953, Bell Laboratories and Vitro Laboratories investigated the reliability of communications electronic parts. Carhart (1953) conducted an early state-of-the-art study of reliability engineering. He divided the reliability problems into five groups, namely, electronics, vacuum tubes, other components, system personnel, and organization. He listed seven factors that determined the worth of manufactured systems: (1) performance capacity, (2) reliability, (3) accuracy, (4) vulnerability, (5) operability, (6) maintainability, and (7) procurability. In 1953, RCA established the first civilian-organized industrial reliability program.

Contributions to reliability engineering through development of missiles began with a DOD project to General Dynamics in 1954. Bell Aircraft Corporation issued the first industrial reliability handbook (LeVan, 1957). In the following decades, reliability engineering played important roles in aerospace and aircraft engineering.

Henney (1956) edited the first commercial reliability book. Chorafas (1960) published a textbook combining statistics with reliability engineering. More comprehensive textbooks on reliability related to manufacturing engineering started to appear in the early 1960s (Bazovsky, 1961; Calabro, 1962). The first reliability engineering course was offered in 1963 by Kececioglu at the University of Arizona. In 1955, the Institute of Radio Engineers [IRE, now the Institute of Electrical and Electronic Engineers (IEEE)] initiated the Reliability and Quality Control Society, and in 1978, IEEE established its Reliability Society.

The American Institute of Aeronautics and Astronautics (AIAA), the Society of Automotive Engineers (SAE), and the American Society of Mechanical Engineers (ASME) initiated the Annual Reliability and Maintainability Conferences in 1962. It became the Annual Symposium on Reliability in 1966 and Annual Reliability and Maintainability Symposium in 1972, the year that the Society of Reliability Engineers was founded at Buffalo, New York.

Beyond manufacture-related reliability engineering, on the infrastructural side, Freudenthal (1947, 1956) was among the first to develop reliability analysis for structural engineering. Public attention on the safety of nuclear power plants and earthquake hazards has provoked significant development on reliability engineering for infrastructures, leading to publication of a series of comprehensive textbooks on the subject (Benjamin and Cornell, 1970; Ang and Tang, 1975, 1984; Yao, 1985; Madsen et al., 1986; Marek et al., 1995; Harr, 1996; Ayyub and McCuen, 1997; Kottagoda and Rosso, 1997; Melchers, 1999; Haldar and Mahadevan, 2000).

1.4   Concept of Reliability Engineering

The basic idea of *reliability engineering* is to determine the failure probability of an engineering system, from which the safety of the system can be assessed or a rational decision can be made on the design, operation, or forecasting of the system, as depicted in Fig. 1.3. For example, Fig. 1.4 schematically illustrates using reliability analysis for risk-based least-cost design of engineering infrastructures.

An infrastructure is a functioning system formed from a combination of a number of components. From the perspective of reliability analysis, infrastructure systems can be classified in several ways. First, they can be grouped according to the sequential layout of the components (Fig. 1.5). A *series system* is a system of components connected in sequence along a single path, i.e., in series. Failure of any one of the components leads to failure of the entire system. A *parallel system* is one with its components connected side by side, i.e., in parallel paths. Many engineering systems have built-in redundancy such that they function as a parallel system. Failure occurs when none of the parallel alternative paths function. Second, from the view point of the time consistency of the statistical characteristics of the systems, they can be classified as a time-invariant statistically stationary system (or *static system*) and a time-varying statistically nonstationary system (or *dynamic system*).

Infrastructures may follow different paths to failure. The ideal and simplest type is the case that the resistance and loading of the system are statistically independent of time, or a stationary system. Most of the existing reliability analysis methods have been developed for such a case.

A more complicated but realistic case is that for which the statistical characteristics of the loading or resistance or both are changing with time, e.g., floods from a watershed under urbanization, rainfall under the effect of global

Types of reliability engineering problems

Safety   Design   Operation   Forecasting   **Simulation (model)**   **Quality control**   **Measurement and sampling**

Infrastructure   Hazards   Infrastructure   Equipment   Product   Real time   Advanced   Measurement accuracy   Sampling frequency

Sample size   Network design

Man-made   Natural   Maintenance and repair   Inspection   Control procedure

Floods, earthquakes, high wind

Hydraulic structures   Bridges   Buildings   Others

Dams   Sewers   Pumps   Etc.

Levees   Canals   Control structures

Transportation

Roads   Airports   Navigation system   Etc.

Figure 1.3 Types of reliability engineering problems.
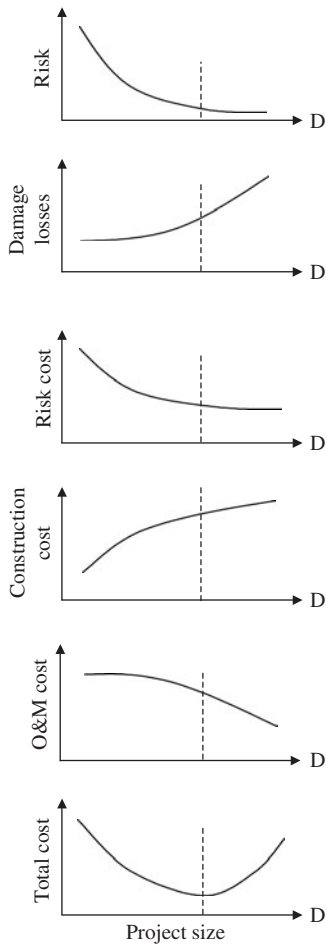
Figure 1.4 Risk-based least-cost design of infrastructural systems. (*After Yen and Tung, 1993.*)

warming, sewer or water supply pipes with deposition, and fatigue or elastic behavior of steel structure members. This case can further be subdivided into the subcases of (1) the changing process is irreversible and accumulative and (2) the changing process is reversible, e.g., repairable. For some infrastructures, the statistical characteristics of the system change with space or in time (or both), e.g., a reach of highway or levee along different terrains. There are other subsets of these time-varying or space-varying dynamic failure cases. One is the subcase that a component of the system already has malfunctioned, but failure has not occurred because the loading has not yet reached the level of such failure, or there is a redundant component to take the load, but the strength of the system is weakened. Another subcase is changing the tolerance of failure, such as changing acceptable standards by regulations.
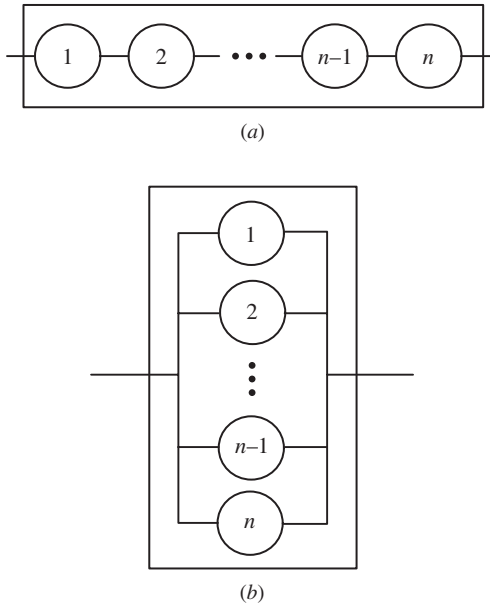
Figure 1.5 Typical configurations of infrastructural systems: (*a*) series system; (*b*) parallel system.

1.5   Definitions of Reliability and Risk

In view of the lack of generally accepted rigorous definitions for risk and reliability, it will be helpful to define these two terms in a manner amenable to mathematical formulation for their quantitative evaluation for engineering systems. The unabridged *Webster's Third New World International Dictionary* gives the following four definitions of *risk*:

1. "the possibility of loss, injury, disadvantage, or destruction, . . . ;

2. someone or something that creates or suggests a hazard or adverse chance: a dangerous element or factor;

3. a: (i) the chance of loss or the perils to the subject matter of insurance covered by a contract, (ii) the degree of probability of such loss;
   b: amount at risk;
   c: a person or thing judged as a (specified) hazard to an insurer;
   d: . . . (insure . . .);

4. the product of the amount that may be lost and the probability of losing it [*United Nations definitition*]"

The unabridged *Random House Dictionary* lists the following definitions of *risk*:

1. "exposure to the chance of injury or loss;

2. insurance: a) the hazard or chance of loss; b) the degree of probability of such loss; c) the amount that the insurance company may lose; d) a person or

thing with reference to the hazard involved in insuring him, her, or it; e) the type of loss, such as life, fire, marine disaster, or earthquake, against which an insurance policy is drawn,

3. at risk . . . ;

4. take or run a risk . . . ."

The *Oxford English Dictionary* defines *risk* as

1. "a) hazard, danger; exposure to mischance or peril; b) to run a or the risk; c) a venturous course; d) at risk or high risk: in danger, subject to hazard; e) a person who is considered a liability or danger; one who is exposed to hazard;

2. the chance or hazard of commercial loss . . . . Also, . . . the chance that is accepted in economic enterprise and considered the source of (an entrepreneur's) profit."

With reference to the first definition of the first two (American) dictionaries, *risk* is defined herein as the probability of failure to achieve the intended goal. *Reliability* is defined mathematically as the complement of the risk. In some disciplines, often the nonengineering ones, the word *risk* refers not just to the probability of failure but also to the consequence of that failure, such as the cost associated with the failure (United Nations definition). Nevertheless, to avoid possible confusion, the mathematical analysis of risk and reliability is termed herein *reliability analysis*.

Failure of an engineering system can be defined as a situation in which the load L (external forces or demands) on the system exceeds the resistance R (strength, capacity, or supply) of the system. The *reliability p_s* of an engineering system is defined as the probability of nonfailure in which the resistance of the system exceeds the load; that is,

$$p_s = P(L \leq R) \tag{1.1}$$

in which $P(\cdot)$ denotes probability. Conversely, the risk is the probability of failure when the load exceeds the resistance. Thus the failure probability (risk) p_f can be expressed mathematically as

$$p_f = P(L > R) = 1 - p_s \tag{1.2}$$

Failure of infrastructures can be classified broadly into two types (Yen and Ang, 1971; Yen et al., 1986): structural failure and functional (performance) failure. *Structural failure* involves damage or change of the structure or facility, therefore hindering its ability to function as desired. On the other hand, *performance failure* does not necessarily involve structural damage. However, the performance limit of the structure is exceeded, and undesirable consequences occur. Generally, the two types of failure are related. Some structures, such as dams, levees, and pavement to support loads, are designed on the concept of

structural failure, whereas others, such as sewers, water supply systems, and traffic networks, are designed on the basis of performance failure.

In conventional infrastructural engineering reliability analysis, the only uncertainty considered is that owing to the inherent randomness of geophysical events, such floods, rainstorms, earthquakes, etc. For instance, in hydrosystem engineering designs, uncertainties associated with the resistance of the hydraulic flow-carrying capacity are largely ignored. Under such circumstances, the preceding mathematical definitions of reliability and failure probability then are reduced to

$$p_s = P(L \leq r_*) \qquad \text{and} \qquad p_f = P(L > r_*) \qquad (1.3)$$

in which the resistance $R = r_*$ is the designated value of resistance, a deterministic quantity. By considering inherent randomness of annual maximum floods, the annual failure probability p_f for a hydraulic structure designed with a capacity to accommodate a T-year flood, i.e., $r_* = l_T$, is $1/T$.

Figure 1.6 shows the effect of hydraulic uncertainty on the overall failure probability under the assumption that both random load and resistance are independent log-normal random variables. The figure can be produced easily from the basic properties of log-normal random variables (see Sec. 2.6.2). Figure 1.6 clearly shows that by considering only inherent randomness of hydrologic load [the bottom curve corresponding to the coefficient of variation (COV), $COV(R) = 0$], the annual failure probability is significantly underestimated as the uncertainty of resistance $COV(R)$ increases. As shown in Fig. 1.1, the inherent natural randomness of hydrologic processes is only one of the many uncertainties in hydrosystems engineering design. This figure clearly demonstrates the deficiency of the conventional frequency-analysis approach in reliability assessment of hydrosystems.
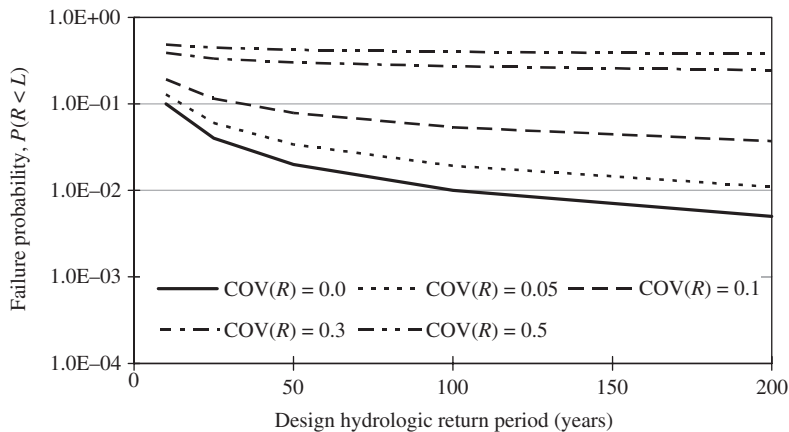
Figure 1.6 Effect of resistance uncertainty on failure probability under $COV(L) = 0.1$.

1.6   Measures of Reliability

In engineering design and analysis, loads usually arise from natural events, such as floods, storms, or earthquakes, that occur randomly in time and in space. The conventional practice for measuring the reliability of a hydrosystems engineering infrastructure is the return period or recurrence interval. The *return period* is defined as the *long-term* average (or expected) time between two successive failure-causing events. In time-to-failure analysis (Chap. 5), an equivalent term is the *mean time to failure*. Simplistically, the return period is equal to the reciprocal of the probability of the occurrence of the event in any one time interval. For many hydrosystems engineering applications, the time interval chosen is 1 year so that the probability associated with the return period is the average annual failure probability. Frequency analysis using the annual maximum flood or rainfall series is a typical example of this kind. Hence the determination of return period depends on the time period chosen (Borgman, 1963). The main theoretical disadvantage of using return period is that reliability is measured only in terms of expected time of occurrence of loads without considering their interactions with the resistance (Melchers, 1999).

In fact, the conventional interpretation of return period can be generalized as the average time period or mean time of the system failure when all uncertainties affecting load and resistance are considered. In other words, the return period can be calculated as the reciprocal of the failure probability computed by Eq. (1.2). Based on this generalized notion of return period, the equivalent return period corresponding to the conventional return period under different levels of resistance uncertainty is shown in Fig. 1.7. As can be seen, the equivalent return period becomes shorter than the conventional return period, as anticipated, when resistance uncertainty increases. For example, with COV(R) = 5 percent, a hydrosystem designed with a 100-year return
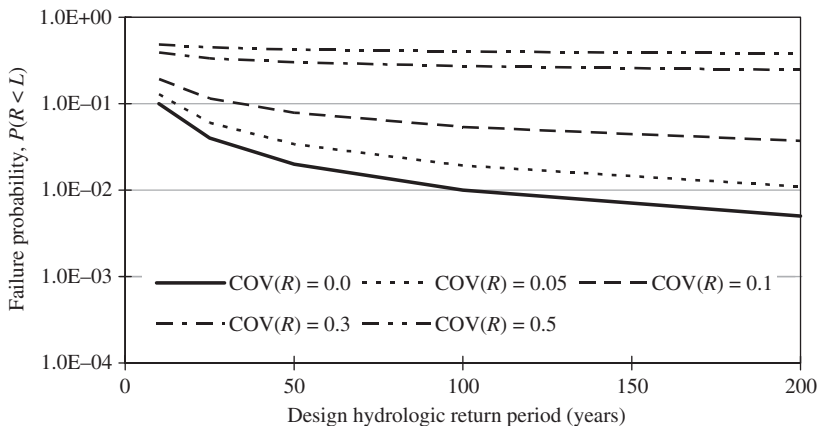
Figure 1.7 Equivalent return period versus design return period under COV(L) = 0.1.

TABLE 1.1   Different Types of Safety Factors

Type of safety factor	Definition
Preassigned	Assigned number
Central	μ_R/μ_L, where μ_R and μ_L are the true mean values of resistance and load
Mean	$\bar{R}/\bar{L}$, where $\bar{R}$ and $\bar{L}$ are the mean values of resistance and load estimated from the available data
Characteristic	R_o/L_o, where R_o and L_o are the specified resistance and load
Partial	$1/\gamma = N_L/N_R$, where $p_f = P(L > \gamma R) = P(N_L L > N_R R)$

SOURCE: After Yen, 1979.

period under the conventional approach actually has about a 50-year return period.

Two other types of reliability measures that consider the relative magnitudes of resistance and anticipated load (called *design load*) are used frequently in engineering practice. One is the *safety margin* (SM), defined as the difference between the resistance and the anticipated load, that is,

$$SM = R - L \tag{1.4}$$

The other is called the *safety factor* (SF), a ratio of resistance to load defined as

$$SF = R/L \tag{1.5}$$

Several types of safety factors are summarized in Table 1.1, and their applications to engineering systems are discussed by Yen (1979).

Preassigned safety factor.   This is an arbitrarily chosen safety factor that is used conventionally without probabilistic consideration. The value chosen largely depends on the designer's subjective judgment with regard to the amount of uncertainty involved in his or her determination of design load and the level of safety desired.

Central safety factor.   Owing to the fact that both resistance and load could be subject to uncertainty, the safety factor defined by Eq. (1.5), in fact, is a quantity subject to uncertainty as well. The central safety factor μ_{SF} is defined as

$$\mu_{SF} = \mu_R/\mu_L \tag{1.6}$$

in which μ_R and μ_L are the true mean values of resistance and load, respectively. In practice, values of μ_R and μ_L cannot be obtained precisely from the limited data. Therefore, μ_{SF} is only of theoretical interest.

Mean safety factor.   If the estimated means of R and L on the basis of data are $\bar{R}$ and $\bar{L}$, respectively, the mean safety factor ($\bar{SF}$) is defined as

$$\bar{SF} = \bar{R}/\bar{L} \tag{1.7}$$

Characteristic safety factor.   Often in a project the significant design values of the parameters are not the mean values but specified values (or range of values). For example, the load used in a spillway design is not the mean value of all the floods nor the mean value of the selected floods of an annual maximum series. It may be simply a specified flood of a given magnitude (e.g., a flood with a 100-year return period). Therefore, the characteristic safety factor (SF_c) can be defined as

$$SF_c = R_o/L_o \tag{1.8}$$

in which R_o and L_o are the specified resistance and load, respectively. If R_o and L_o both are assigned without a probabilistic analysis, Eq. (1.8) is identical to Eq. (1.5). If R_o and L_o are taken to be the mean values of resistance and load, Eq. (1.8) would become like Eq. (1.6) or Eq. (1.7). In general, R_o and L_o can be determined through a probabilistic analysis. For example, Tang and Yen (1972) use the estimated mean of resistance and the specified load, that is,

$$SF_c = \bar{R}/L_o \tag{1.9}$$

to develop a risk–safety factor relationship in storm sewer design. Tung and Mays (1981) used the 100-year flood from the frequency analysis for L_o in developing risk–safety factor curves for a levee system.

Partial safety factor.   The preceding safety factors apply to the total load and resistance of the system. It is possible, however, that different components in the system may be subject to different degrees of uncertainty. A smaller value of the safety factor can be assigned to those elements or components associated with less uncertainty than those with more uncertainty. In Table 1.1, N_R and N_L are the separate safety factors assigned to the resistance and load, respectively.

Theoretically, any one of the safety factors can be applied for its quantitative evaluation. However, the central safety factor is only of theoretical importance because in practice the exact distributions and values of the coefficient of variation are not known but estimated. Among the other four definitions, which one is preferred would depend on the nature of the problem. Clearly, these safety factors can be modified and refined. They are not mutually exclusive and can be made complementary. An in-depth comparative investigation of these factors in view of infrastructural system engineering applications would be desirable.

1.7 Overall View of Reliability Analysis Methods

There are two basic probabilistic approaches to evaluate the reliability of an infrastructural system. The most direct approach is a statistical analysis of data of past failure records for similar systems. The other approach is through reliability analysis, which considers and combines the contribution of each factor potentially influencing failure. The former is a *lumped-system approach* requiring

no knowledge about the internal physical behavior of the facility or structure and its load and resistance. For example, dam failure data show that the overall average failure probability for dams of all types over 15 m in height is around 10^{-3} per dam per year (U.S. National Research Council, 1983; Cheng, 1993). This statistical approach may fit well with manufactured systems for which planned repeated tests can be made and the performance of many identical prototypes can be observed. For infrastructural systems in most cases, this direct approach is impractical because (1) infrastructures are usually unique and site-specific, (2) the sample size is too small to be statistically reliable, especially for low-probability/high-consequence events, (3) the sample may not be representative of the structure or of the population, and (4) the physical conditions of a dam may be nonstationary, i.e., varying with respect to time. The average risk of dam failure mentioned earlier does not differentiate concrete dams from earth-fill dams, arch dams from gravity dams, large dams from small dams, and old dams from new dams. If one wished to know the likelihood of failure of a particular 10-year-old double-curvature-arch concrete high dam, most likely one will find only very few failure data of similar dams, insufficient for any meaningful statistical analysis. Since no dams are identical and conditions of dams change with time, in many circumstances it may be more desirable to use the second approach by conducting a reliability analysis.

There are two major steps in reliability analysis: (1) to identify and analyze the uncertainties of each contributing factor and (2) to combine the uncertainties of the stochastic factors to determine the overall reliability of the structure. The second step, in turn, also may proceed in two different ways: (1) directly combining the uncertainties of all factors and (2) separately combining the uncertainties of the factors belonging to different components or subsystems to evaluate first the respective subsystem reliability and then combining the reliabilities of the different components or subsystems to yield the overall reliability of the structure. The first way applies to very simple structures, whereas the second way is more suitable to complicated systems. For example, to evaluate the reliability of a dam, the hydrologic, hydraulic, geotechnical, structural, and other disciplinary reliabilities could be evaluated separately first and then combined to yield the overall dam reliability. Or the component reliabilities could be evaluated first according to the different failure modes and then combined. Analysis tools described in Chap. 5, such as fault tree and event tree, are useful to divide the system into component evaluation and combination.

References

Ang, A. H.-S., and Tang, W. H. (1975). *Probability Concepts in Engineering Planning and Design*, Vol. I: *Basic Principles*, John Wiley and Sons, New York.

Ang, A. H.-S. and Tang, W. H. (1984). *Probability Concepts in Engineering Planning and Design: Decision, Risk and Reliability*, Vol. II: *Decision, Risk, and Reliability*, John Wiley and Sons, New York.

Ayyub, B. M., and McCuen, R. (1997). *Probability, Statistics and Reliability for Engineers*, CRC Press, Boca Raton, FL.

Bazovsky, I. (1961). *Reliability Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ.

Benjamin, J. R., and Cornell, C. A. (1970). *Probability, Statistics, and Decisions for Civil Engineers*, McGraw-Hill, New York.

Birolini, A. (1999). *Reliability Engineering: Theory and Practice*, 3d ed., Springer-Verlag, Berlin.

Borgman, L. E. (1963). Risk criteria, *Journal of the Waterways and Harbors Division*, ASCE, 89(WW3):1–35.

Calabro, S. R. (1962). *Reliability Principles and Practices*, McGraw-Hill, New York.

Carhart, R. R. (1953). A survey of the current status of the electronic reliability problem, Research Memo RM-1131, Rand Corporation.

Cheng, S. T. (1993). Statistics on dam failures, in *Reliability and Uncertainty Analysis in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, ASCE, New York, pp. 97–106.

Chorafas, D. N. (1960). *Statistical Processes and Reliability Engineering*, Van Nostrand Reinhold, New York.

Cornell, C. A. (1972). First-order analysis of model and parameter uncertainty, in *Proceedings*, International Symposium on Uncertainties in Hydrologic and Water Resources Systems, Vol. 2, Tucson, AZ, pp. 1245–1272.

Freudenthal, A. M. (1947). The safety of structures, *Transactions of the ASCE*, 112:125–159.

Freudenthal, A. M. (1956). Safety and probability of structural failure, *Transactions of the ASCE*, 121:1337–1375.

Haldar, A., and Mahadevan, S. (2000). *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley and Sons, New York.

Harr, M. E. (1996). *Reliability-Based Design in Civil Engineering*, Dover Publications, New York.

Henney, K. (1956). *Relibility Factors for Ground Electronic Equipment*, McGraw-Hill, New York.

Ireson, W. G., and Coombs, C. F., eds. (1988). *Handbook of Reliability Engineering and Management*, McGraw-Hill, New York.

Kottagoda, N. T., and Rosso, R. (1997). *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York.

Kececioglu, D. (1991). *Reliability Engineering Handbook*, Prentice-Hall, Englewood Cliffs, NJ.

LeVan, W. I. (1957). *Reliability Check List of Reliability Program Practices*, *Reliability Handbook 7-58-2954-9*, Space Flight Division, Bell Aircraft Corporation, Buffalo, NY.

Madsen, H. O., Krenk, S., and Lind, N. C. (1986). *Methods of Structural Safety*, Prentice-Hall, Englewood Cliffs, NJ.

Marek, P., Gustar, M., and Anagnos, T. (1995). *Simulation-Based Reliability Assessment for Structural Engineers*, CRC Press, Boca Raton, FL.

Melchers, R. E. (1999). *Structural Reliability: Analysis and Prediction*, 2nd. ed., John Wiley and Sons, New York.

Modarres, M., Kaminskiy, M., and Krivtsov, V. (1999). *Reliability Engineering and Risk Analysis*, Marcel Dekker, New York.

Pecht, M. (1995). *Product Reliability, Maintainability, and Supportability Handbook*, CRC Press, Inc., Boca Raton, FL.

Shewart, W. A. (1931). *Economic Control of Quality of Manufactured Products*, Van Nostrand Co., New York.

Tang, W. H., and Yen, B. C. (1972). Hydrologic and hydraulic design under uncertainties, *Proceedings*, International Symposium on Uncertainties in Hydrologic and Water Resources Systems, Vol. 2, Tucson, AZ, pp. 868–882.

Tung, Y. K. and Mays, L. W. (1981). Risk and reliability model for levee design, *Water Resources Research*, 17(4):833–842.

Tung, Y. K., and Yen, B. C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, McGraw-Hill, New York.

Tye, W. (1944). Factor of safety—or of habit? *Journal of the Royal Aeronautical Society*, 58(407): 487.

United Nations Department of Humanitarian Affairs (1992). *Glossary: Internationally Agreed Glossary of Basic Terms Related to Disaster Management*, United Nations, Geneva, Switzerland.

Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*, John Wiley and Sons, New York.

U. S. National Research Council, Committee on Safety of Existing Dams (1983). *Safety of Existing Dams: Evaluation and Improvement*, National Academy Press, Washington.

U. S. National Research Council (2000). *Risk Analysis and Uncertainty in Flood Damage Reduction Studies*, National Academy Press, Washington.

Yao, J. T.-P. (1985). *Safety and Reliability of Existing Structures*, Pitman Advanced Publication Program, London.

Yen, B. C. (1979). Safety factor in hydrologic and hydraulic engineering design, in *Reliability in Water Resources Management*, ed. by E. A. McBean, K. W. Hipel, and T. E. Unny, Water Resources Publications, Littleton, CO, pp. 389–407.

Yen, B. C., and Ang, A. H. S. (1971). Risk analysis in design of hydraulic projects, in *Stochastic Hydraulics*, Proceedings of First International Symposium on Stochastic Hydraulics, University of Pittsburgh, ed. by C. L. Chiu, Pittsburgh, PA, pp. 694–701.

Yen, B. C., and Tung, Y. K. (1993). Some recent progress in reliability analysis for hydraulic design, in *Reliability and Uncertainty Analysis in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, ASCE, New York, pp. 35–79.

Yen, B. C., Cheng, S. T., and Melching, C. S. (1986). First-order reliability analysis, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 1–36.

Fundamentals of Probability and Statistics for Reliability Analysis*

Assessment of the reliability of a hydrosystems infrastructural system or its components involves the use of probability and statistics. This chapter reviews and summarizes some fundamental principles and theories essential to reliability analysis.

2.1 Terminology

In probability theory, an *experiment* represents the process of making observations of random phenomena. The outcome of an observation from a random phenomenon cannot be predicted with absolute accuracy. The entirety of all possible outcomes of an experiment constitutes the *sample space*. An *event* is any subset of outcomes contained in the sample space, and hence an event could be an empty (or null) set, a subset of the sample space, or the sample space itself. Appropriate operators for events are *union, intersection*, and *complement*. The occurrence of events A and B is denoted as $A \cup B$ (the union of A and B), whereas the joint occurrence of events A and B is denoted as $A \cap B$ or simply (A, B) (the intersection of A and B). Throughout the book, the complement of event A is denoted as A'. When two events A and B contain no common elements, then the two events are *mutually exclusive* or *disjoint events*, which is expressed as $(A, B) = \emptyset$, where $\emptyset$ denotes the null set. Venn diagrams illustrating the union and intersection of two events are shown in Fig. 2.1. When the occurrence of event A depends on that of event B, then they are *conditional events*,

---

*Most of this chapter, except Secs. 2.5 and 2.7, is adopted from Tung and Yen (2005).
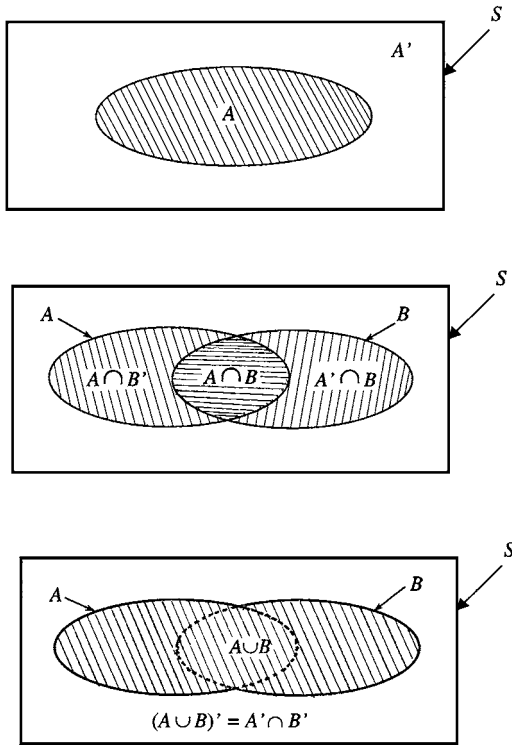
Figure 2.1   Venn diagrams for basic set operations.

which is denoted by $A\,|\,B$. Some useful set operation rules are

1. Commutative rule: $A \cup B = B \cup A$; $A \cap B = B \cap A$.
2. Associative rule: $(A \cup B) \cup C = A \cup (B \cup C)$; $(A \cap B) \cap C = A \cap (B \cap C)$.
3. Distributive rule: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. de Morgan's rule: $(A \cup B)' = A' \cap B'$; $(A \cap B)' = A' \cup B'$.

   Probability is a numeric measure of the likelihood of the occurrence of an event. Therefore, probability is a real-valued number that can be manipulated by ordinary algebraic operators, such as $+$, $-$, $\times$, and $/$. The probability of the occurrence of an event A can be assessed in two ways. In the case where an experiment can be repeated, the probability of having event A occurring can be estimated as the ratio of the number of replications in which event A occurs n_A versus the total number of replications n, that is, n_A/n. This ratio is called the *relative frequency* of occurrence of event A in the sequence of n replications. In principle, as the number of replications gets larger, the value of the relative

frequency becomes more stable, and the true probability of event A occurring could be obtained as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n} \tag{2.1}$$

The probabilities so obtained are called *objective* or *posterior probabilities* because they depend completely on observations of the occurrence of the event.

In some situations, the physical performance of an experiment is prohibited or impractical. The probability of the occurrence of an event can only be estimated subjectively on the basis of experience and judgment. Such probabilities are called *subjective* or *prior probabilities*.

2.2 Fundamental Rules of Probability Computations

2.2.1 Basic axioms of probability

The three basic axioms of probability computation are (1) *nonnegativity*: $P(A) \geq 0$, (2) *totality*: $P(S) = 1$, with S being the sample space, and (3) *additivity*: For two mutually exclusive events A and B, $P(A \cup B) = P(A) + P(B)$.

As indicated from axioms (1) and (2), the value of probability of an event occurring must lie between 0 and 1. Axiom (3) can be generalized to consider K mutually exclusive events as

$$P(A_1 \cup A_2 \cup \cdots \cup A_K) = P\left(\bigcup_{k=1}^{K} A_k\right) = \sum_{k=1}^{K} P(A_k) \tag{2.2}$$

An *impossible event* is an empty set, and the corresponding probability is zero, that is, $P(\emptyset) = 0$. Therefore, two mutually exclusive events A and B have zero probability of joint occurrence, that is, $P(A, B) = P(\emptyset) = 0$. Although the probability of an impossible event is zero, the reverse may not necessarily be true. For example, the probability of observing a flow rate of exactly 2000 m^3/s is zero, yet having a discharge of 2000 m^3/s is not an impossible event.

Relaxing the requirement of mutual exclusiveness in axiom (3), the probability of the union of two events can be evaluated as

$$P(A \cup B) = P(A) + P(B) - P(A, B) \tag{2.3}$$

which can be further generalized as

$$P\left(\bigcup_{k=1}^{K} A_k\right) = \sum_{k=1}^{K} P(A_k) - \sum \sum_{i < j} P(A_i, A_j)$$
$$+ \sum \sum \sum_{i < j < k} P(A_i, A_j, A_k) - \cdots + (-1)^K P(A_1, A_2, \ldots, A_K)$$

$$\tag{2.4}$$

If all are mutually exclusive, all but the first summation term on the right-hand side of Eq. (2.3) vanish, and it reduces to Eq. (2.2).

Example 2.1   There are two tributaries in a watershed. From past experience, the probability that water in tributary 1 will overflow during a major storm event is 0.5, whereas the probability that tributary 2 will overflow is 0.4. Furthermore, the probability that both tributaries will overflow is 0.3. What is the probability that at least one tributary will overflow during a major storm event?

Solution   Define E_i = event that tributary i overflows for $i = 1, 2$. From the problem statements, the following probabilities are known: $P(E_1) = 0.5$, $P(E_2) = 0.4$, and $P(E_1, E_2) = 0.3$.

The probability having at least one tributary overflowing is the probability of event E_1 or E_2 occurring, that is, $P(E_1 \cup E_2)$. Since the overflow of one tributary does not preclude the overflow of the other tributary, E_1 and E_2 are not mutually exclusive. Therefore, the probability that at least one tributary will overflow during a major storm event can be computed, according to Eq. (2.3), as

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1, E_2) = 0.5 + 0.4 - 0.3 = 0.6$$

2.2.2   Statistical independence

If two events are *statistically independent* of each other, the occurrence of one event has no influence on the occurrence of the other. Therefore, events A and B are independent if and only if $P(A, B) = P(A)P(B)$. The probability of joint occurrence of K independent events can be generalized as

$$P\left(\bigcap_{k=1}^{K} A_k\right) = P(A_1) \times P(A_2) \times \cdots \times P(A_K) = \prod_{k=1}^{K} P(A_k) \qquad (2.5)$$

It should be noted that the mutual exclusiveness of two events does not, in general, imply independence, and vice versa, unless one of the events is an impossible event. If the two events A and B are independent, then A, A', B, and B' all are independent, but not necessarily mutually exclusive, events.

Example 2.2   Referring to Example 2.1, the probabilities that tributaries 1 and 2 overflow during a major storm event are 0.5 and 0.4, respectively. For simplicity, assume that the occurrences of overflowing in the two tributaries are independent of each other. Determine the probability of at least one tributary overflowing in a major storm event.

Solution   Use the same definitions for events E_1 and E_2. The problem is to determine $P(E_1 \cup E_2)$ by

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1, E_2)$$

Note that in this example the probability of joint occurrences of both tributaries overflowing, that is, $P(E_1, E_2)$, is not given directly by the problem statement, as in Example 2.1. However, it can be determined from knowing that the occurrences of

overflows in the tributaries are independent events, according to Eq. (2.5), as

$$P(E_1, E_2) = P(E_1)P(E_2) = (0.5)(0.4) = 0.2$$

Then the probability that at least one tributary would overflow during a major storm event is

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1, E_2) = 0.5 + 0.4 - 0.2 = 0.7$$

2.2.3 Conditional probability

The *conditional probability* is the probability that a conditional event would occur. The conditional probability $P(A \mid B)$ can be computed as

$$P(A \mid B) = \frac{P(A, B)}{P(B)} \tag{2.6}$$

in which $P(A \mid B)$ is the occurrence probability of event A given that event B has occurred. It represents a reevaluation of the occurrence probability of event A in the light of the information that event B has occurred. Intuitively, A and B are two independent events if and only if $P(A \mid B) = P(A)$. In many cases it is convenient to compute the joint probability $P(A, B)$ by

$$P(A, B) = P(B)P(A \mid B) \qquad \text{or} \qquad P(A, B) = P(A)P(B \mid A)$$

The probability of the joint occurrence of K dependent events can be generalized as

$$P\left(\bigcap_{k=1}^{K} A_k\right) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_2, A_1) \times \cdots \times P(A_K \mid A_{K-1}, \ldots, A_2, A_1) \tag{2.7}$$

Example 2.3 Referring to Example 2.2, the probabilities that tributaries 1 and 2 would overflow during a major storm event are 0.5 and 0.4, respectively. After examining closely the assumption about the independence of overflow events in the two tributaries, its validity is questionable. Through an analysis of historical overflow events, it is found that the probability of tributary 2 overflowing is 0.6 if tributary 1 overflows. Determine the probability that at least one tributary would overflow in a major storm event.

Solution Let E_1 and E_2 be the events that tributary 1 and 2 overflow, respectively. From the problem statement, the following probabilities can be identified:

$$P(E_1) = 0.5 \qquad P(E_2) = 0.4 \qquad P(E_2 \mid E_1) = 0.6$$

in which $P(E_2 \mid E_1)$ is the conditional probability representing the likelihood that tributary 2 would overflow given that tributary 1 has overflowed. The probability of at least one tributary overflowing during a major storm event can be computed by

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1, E_2)$$

in which the probability of joint occurrence of both tributaries overflowing, that is, $P(E_1, E_2)$, can be obtained from the given conditional probability, according to Eq. (2.7), as

$$P(E_1, E_2) = P(E_2 \mid E_1)P(E_1) = (0.6)(0.5) = 0.3$$

The probability that at least one tributary would overflow during a major storm event can be obtained as

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1, E_2) = 0.5 + 0.4 - 0.3 = 0.6$$

2.2.4 Total probability theorem and Bayes' theorem

The probability of the occurrence of an event E, in general, cannot be determined directly or easily. However, the event E may occur along with other attribute events A_k. Referring to Fig. 2.2, event E could occur jointly with K mutually exclusive ($A_j \cap A_k = \emptyset$ for $j \neq k$) and *collectively exhaustive* ($A_1 \cup A_2 \cup \cdots \cup A_K) = S$ attributes A_k, $k = 1, 2, \ldots, K$. Then the probability of the occurrence of an event E, regardless of the attributes, can be computed as

$$P(E) = \sum_{k=1}^{K} P(E, A_k) = \sum_{k=1}^{K} P(E \mid A_k)P(A_k) \tag{2.8}$$

Equation (2.8) is called the *total probability theorem*.

Example 2.4 Referring to Fig. 2.3, two upstream storm sewer branches (I_1 and I_2) merge to a sewer main (I_3). Assume that the flow-carrying capacities of the two upstream sewer branches I_1 and I_2 are equal. However, hydrologic characteristics of the contributing drainage basins corresponding to I_1 and I_2 are somewhat different. Therefore, during a major storm event, the probabilities that sewers I_1 and I_2 will exceed their capacities (surcharge) are 0.5 and 0.4, respectively. For simplicity, assume that the occurrences of surcharge events in the two upstream sewer branches are independent of each other. If the flow capacity of the downstream sewer main I_3 is
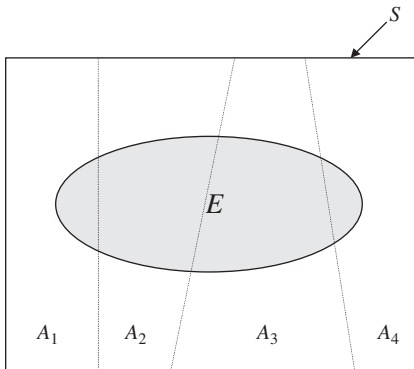
Figure 2.2 Schematic diagram of total probability theorem.
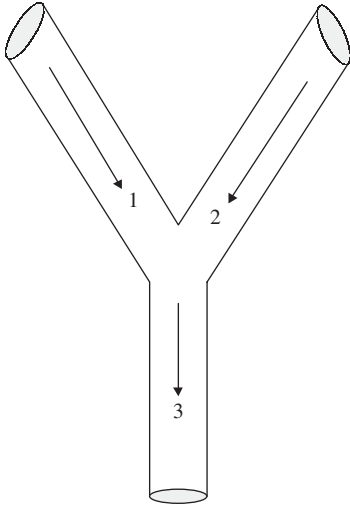
Figure 2.3 A system with three sewer sections.

the same as its two upstream branches, what is the probability that the flow capacity of the sewer main I_3 will be exceeded? Assume that when both upstream sewers are carrying less than their full capacities, the probability of downstream sewer main I_3 exceeding its capacity is 0.2.

Solution Let E_1, E_2, and E_3, respectively, be events that sewer I_1, I_2, and I_3 exceed their respective flow capacity. From the problem statements, the following probabilities can be identified: $P(E_1) = 0.50$, $P(E_2) = 0.40$, and $P(E_3 \mid E_1', E_2') = 0.2$.

To determine $P(E_3)$, one first considers the basic events occurring in the two upstream sewer branches that would result in surcharge in the downstream sewer main E_3. There are four possible attribute events that can be defined from the flow conditions of the two upstream sewers leading to surcharge in the downstream sewer main. They are $A_1 = (E_1, E_2)$, $A_2 = (E_1', E_2)$, $A_3 = (E_1, E_2')$, and $A_4 = (E_1', E_2')$. Furthermore, the four events A_1, A_2, A_3, and A_4 are mutually exclusive.

Since the four attribute events A_1, A_2, A_3, and A_4 contribute to the occurrence of event E_3, the probability of the occurrence of E_3 can be calculated, according to Eq. (2.8), as

$$P(E_3) = P(E_3, A_1) + P(E_3, A_2) + P(E_3, A_3) + P(E_3, A_4)$$
$$= P(E_3 \mid A_1)P(A_1) + P(E_3 \mid A_2)P(A_2) + P(E_3 \mid A_3)P(A_3) + P(E_3 \mid A_4)P(A_4)$$

To solve this equation, each of the probability terms on the right-hand side must be identified. First, the probability of the occurrence of A_1, A_2, A_3, and A_4 can be determined as the following:

$$P(A_1) = P(E_1, E_2) = P(E_1) \times P(E_2) = (0.5)(0.4) = 0.2$$

The reason that $P(E_1, E_2) = P(E_1) \times P(E_2)$ is due to the independence of events E_1 and E_2. Since E_1 and E_2 are independent events, then E_1, E_1', E_2, and E_2' are also

independent events. Therefore,

$$P(A_2) = P(E'_1, E_2) = P(E'_1) \times P(E_2) = (1 - 0.5)(0.4) = 0.2$$
$$P(A_3) = P(E_1, E'_2) = P(E_1) \times P(E'_2) = (0.5)(1 - 0.4) = 0.3$$
$$P(A_4) = P(E'_1, E'_2) = P(E'_1) \times P(E'_2) = (1 - 0.5)(1 - 0.4) = 0.3$$

The next step is to determine the values of the conditional probabilities, that is, $P(E_3 \mid A_1)$, $P(E_3 \mid A_2)$, $P(E_3 \mid A_3)$, and $P(E_3 \mid A_4)$. The value of $P(E_3 \mid A_4) = P(E_3 \mid E'_1, E'_2) = 0.2$ is given by the problem statement. On the other hand, the values of the remaining three conditional probabilities can be determined from an understanding of the physical process. Note that from the problem statement the downstream sewer main has the same conveyance capacity as the two upstream sewers. Hence any upstream sewer exceeding its flow-carrying capacity would result in surcharge in the downstream sewer main. Thus the remaining three conditional probabilities can be easily determined as

$$P(E_3 \mid A_1) = P(E_3 \mid E_1, E_2) = 1.0$$
$$P(E_3 \mid A_2) = P(E_3 \mid E'_1, E_2) = 1.0$$
$$P(E_3 \mid A_3) = P(E_3 \mid E_1, E'_2) = 1.0$$

Putting all relevant information into the total probability formula given earlier, the probability that the downstream sewer main I_3 would be surcharged in a major storm is

$$P(E_3) = P(E_3 \mid A_1)P(A_1) + P(E_3 \mid A_2)P(A_2) + P(E_3 \mid A_3)P(A_3) + P(E_3 \mid A_4)P(A_4)$$
$$= (1.0)(0.2) + (1.0)(0.2) + (1.0)(0.3) + (0.2)(0.3)$$
$$= 0.76$$

The total probability theorem describes the occurrence of an event E that may be affected by a number of attribute events A_k, $k = 1, 2, \ldots, K$. In some situations, one knows $P(E \mid A_k)$ and would like to determine the probability that a particular event A_k contributes to the occurrence of event E. In other words, one likes to find $P(A_k \mid E)$. Based on the definition of the conditional probability (Eq. 2.6) and the total probability theorem (Eq. 2.8), $P(A_k \mid E)$ can be computed as

$$P(A_k \mid E) = \frac{P(A_k, E)}{P(E)} = \frac{P(E \mid A_k)P(A_k)}{\sum_{k'=1}^{K} P(E \mid A_{k'})P(A_{k'})} \qquad \text{for } k = 1, 2, \ldots, K \quad (2.9)$$

Equation (2.9) is called *Bayes' theorem*, and $P(A_k)$ is the *prior probability*, representing the initial belief of the likelihood of occurrence of attribute event A_k. $P(E \mid A_k)$ is the *likelihood function*, and $P(A_k \mid E)$ is the *posterior probability*, representing the new evaluation of A_k being responsible in the light of the occurrence of event E. Hence Bayes' theorem can be used to update and revise the calculated probability as more information becomes available.

Example 2.5   Referring to Example 2.4, if surcharge is observed in the downstream storm sewer main I_3, what is the probability that the incident is caused by simultaneous surcharge of both upstream sewer branches?

Solution   From Example 2.4, A_1 represents the event that both upstream storm sewer branches exceed their flow-carrying capacities. The problem is to find the conditional probability of A_1, given that event E_3 has occurred, that is, $P(A_1 \mid E_3)$. This conditional probability can be expressed as

$$P(A_1 \mid E_3) = \frac{P(A_1, E_3)}{P(E_3)} = \frac{P(E_3 \mid A_1)P(A_1)}{P(E_3)}$$

From Example 2.4, the numerator and denominator of the preceding conditional probability can be computed as

$$P(A_1 \mid E_3) = \frac{P(E_3 \mid A_1)P(A_1)}{P(E_3)} = \frac{(1.0)(0.2)}{0.76} = 0.263$$

The original assessment of the probability is 20 percent that both upstream sewer branches would exceed their flow-carrying capacities. After an observation of downstream surcharge from a new storm event, the probability of surcharge occurring in both upstream sewers is revised to 26.3 percent.

2.3   Random Variables and their Distributions

In analyzing the statistical features of infrastructural system responses, many events of interest can be defined by the related random variables. A *random variable* is a real-value function defined on the sample space. In other words, a random variable can be viewed as a mapping from the sample space to the real line, as shown in Fig. 2.4. The standard convention is to denote a random variable by an upper-case letter, whereas a lower-case letter is used to represent the realization of the corresponding random variable. For example, one may use Q to represent flow magnitude, a random variable, whereas q is used to represent the values that Q takes. A random variable can be discrete or continuous. Examples of discrete random variables encountered in hydrosystems infrastructural designs are the number of storm events occurring in a specified time period, the number of overtopping events per year for a levee system, and so on. On the other hand, examples of continuous random variables are flow rate, rainfall intensity, water-surface elevation, roughness factor, and pollution concentration, among others.

2.3.1   Cumulative distribution function and probability density function

The *cumulative distribution function* (CDF), or simply *distribution function* (DF), of a random variable X is defined as

$$F_x(x) = P(X \le x) \tag{2.10}$$
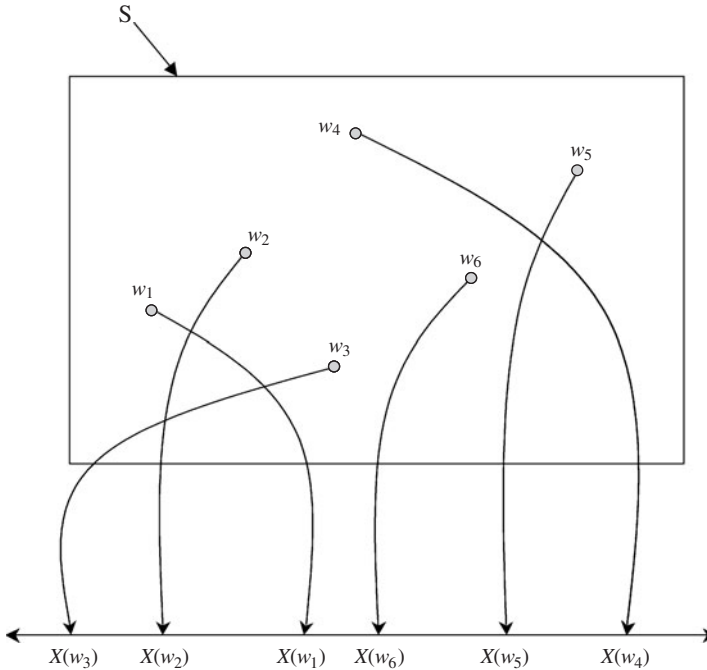
Figure 2.4 A random variable $X(w)$ as mapped from the sample space to the real line.

The CDF $F_x(x)$ is the nonexceedance probability, which is a nondecreasing function of the argument x, that is, $F_x(a) \leq F_x(b)$, for $a < b$. As the argument x approaches the lower bounds of the random variable X, the value of $F_x(x)$ approaches zero, that is, $\lim_{x \to -\infty} F_x(x) = 0$; on the other hand, the value of $F_x(x)$ approaches unity as its argument approaches the upper bound of X, that is, $\lim_{x \to \infty} F_x(x) = 1$. With $a < b$, $P(a < X \leq b) = F_x(b) - F_x(a)$.

For a discrete random variable X, the *probability mass function* (PMF), is defined as

$$p_x(x) = P(X = x) \tag{2.11}$$

The PMF of any discrete random variable, according to axioms (1) and (2) in Sec. 2.1, must satisfy two conditions: (1) $p_x(x_k) \geq 0$, for all x_k's, and (2) $\Sigma_{\text{all } k} p_x(x_k) = 1$. The PMF of a discrete random variable and its associated CDF are sketched schematically in Fig. 2.5. As can be seen, the CDF of a discrete random variable is a staircase function.

For a continuous random variable, the *probability density function* (PDF) $f_x(x)$ is defined as

$$f_x(x) = \frac{dF_x(x)}{dx} \tag{2.12}$$
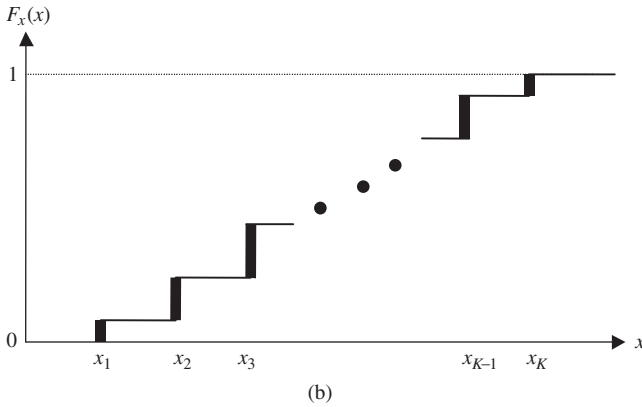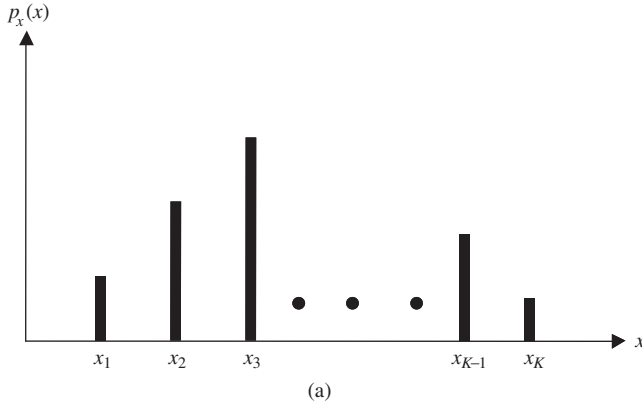
Figure 2.5 (*a*) Probability mass function (PMF) and (*b*) cumulative distribution function (CDF) of a discrete random variable.

The PDF of a continuous random variable $f_x(x)$ is the slope of its corresponding CDF. Graphic representations of a PDF and a CDF are shown in Fig. 2.6. Similar to the discrete case, any PDF of a continuous random variable must satisfy two conditions: (1) $f_x(x) \geq 0$ and (2) $\int f_x(x)\,dx = 1$. Given the PDF of a random variable X, its CDF can be obtained as

$$F_x(x) = \int_{-\infty}^{x} f_x(u)\,du \tag{2.13}$$

in which u is the dummy variable. It should be noted that $f_x(\cdot)$ is not a probability; it only has meaning when it is integrated between two points. The probability of a continuous random variable taking on a particular value is zero, whereas this may not be the case for discrete random variables.
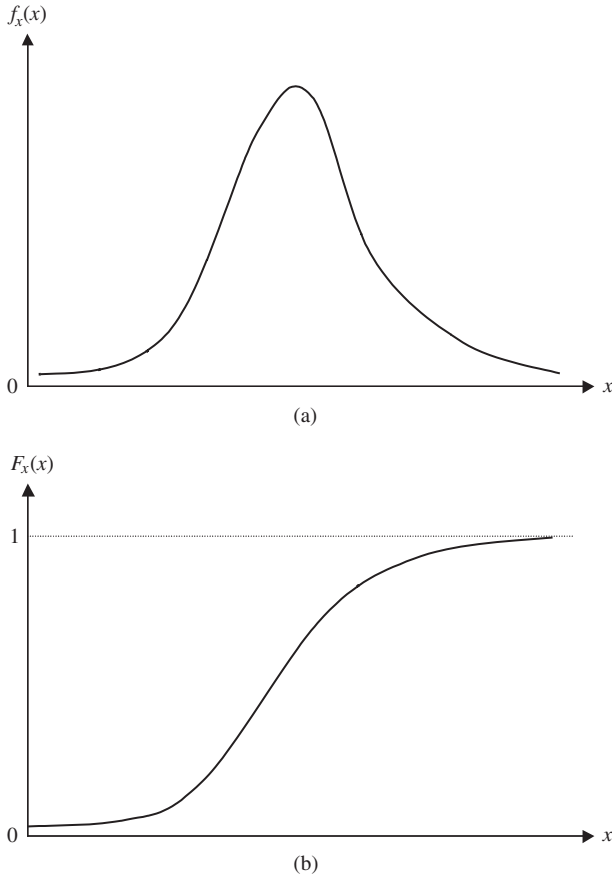
Figure 2.6 (*a*) Probability density function (PDF) and (*b*) cumulative distribution function (CDF) of a continuous random variable.

Example 2.6 The time to failure T of a pump in a water distribution system is a continuous random variable having the PDF of

$$f_t(t) = \exp(-t/1250)/\beta \qquad \text{for } t \geq 0, \ \beta > 0$$

in which t is the elapsed time (in hours) before the pump fails, and β is the parameter of the distribution function. Determine the constant β and the probability that the operating life of the pump is longer than 200 h.

Solution The shape of the PDF is shown in Fig. 2.7. If the function $f_t(t)$ is to serve as a PDF, it has to satisfy two conditions: (1) $f_t(t) \geq 0$, for all t, and (2) the area under $f_t(t)$ must equal unity. The compliance of the condition (1) can be proved easily. The value of the constant β can be determined through condition (2) as

$$1 = \int_0^\infty f_t(t)\,dt = \int_0^\infty \frac{e^{-t/1250}}{\beta}\,dt = \left[\frac{-1250e^{-t/1250}}{\beta} \right]_0^\infty = \frac{1250}{\beta}$$
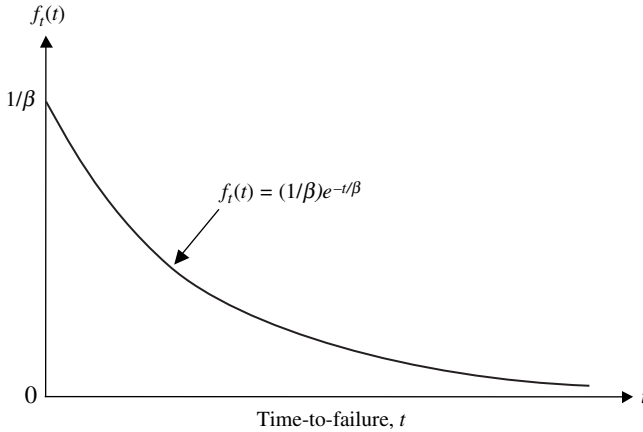
Figure 2.7 Exponential failure density curve.

Therefore, the constant $\beta = 1250$ h/failure. This particular PDF is called the *exponential distribution* (see Sec. 2.6.3). To determine the probability that the operational life of the pump would exceed 200 h, one calculates $P(T \geq 200)$:

$$P(T \geq 200) = \int_{200}^{\infty} \frac{e^{-t/1250}}{1250} \, dt = \left[-e^{-t/1250} \right]_{200}^{\infty} = e^{-200/1250} = 0.852$$

2.3.2 Joint, conditional, and marginal distributions

The *joint distribution* and *conditional distribution*, analogous to the concepts of joint probability and conditional probability, are used for problems involving multiple random variables. For example, flood peak and flood volume often are considered simultaneously in the design and operation of a flood-control reservoir. In such cases, one would need to develop a *joint PDF* of flood peak and flood volume. For illustration purposes, the discussions are limited to problems involving two random variables.

The *joint PMF* and *joint CDF* of two discrete random variables X and Y are defined, respectively, as

$$p_{x,y}(x, y) = P(X = x, Y = y) \tag{2.14a}$$

$$F_{x,y}(u, v) = P(X \leq u, Y \leq v) = \sum_{x \leq u} \sum_{y \leq v} p_{x,y}(x, y) \tag{2.14b}$$

Schematic diagrams of the joint PMF and joint CDF of two discrete random variables are shown in Fig. 2.8.
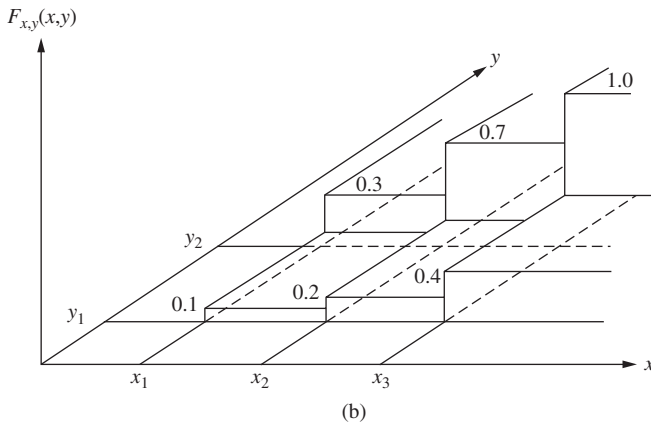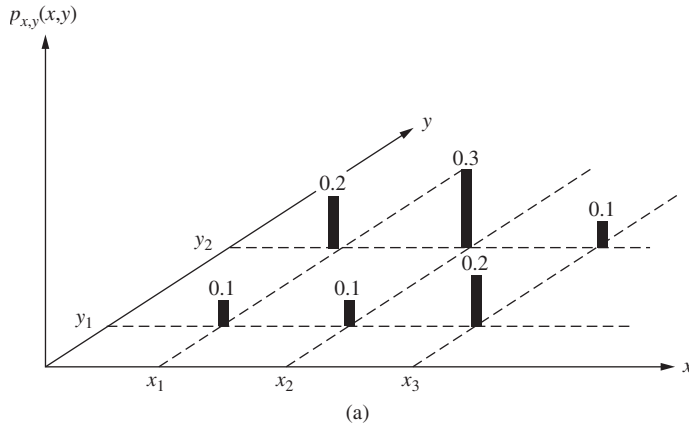
Figure 2.8 (*a*) Joint probability mass function (PMF) and (*b*) cumulative distribution function (CDF) of two discrete random variables.

The joint PDF of two continuous random variables X and Y, denoted as $f_{x,y}(x, y)$, is related to its corresponding joint CDF as

$$f_{x,y}(x, y) = \frac{\partial^2 [F_{x,y}(x, y)]}{\partial x \, \partial y} \tag{2.15a}$$

$$F_{x,y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{x,y}(u, v) \, du \, dv \tag{2.15b}$$

Similar to the univariate case, $F_{x,y}(-\infty, -\infty) = 0$ and $F_{x,y}(\infty, \infty) = 1$. Two random variables X and Y are statistically independent if and only if $f_{x,y}(x, y) = f_x(x) \times f_y(y)$ and $F_{x,y}(x, y) = F_x(x) \times F_y(y)$. Hence a problem involving multiple independent random variables is, in effect, a univariate problem in which each individual random variable can be treated separately.

If one is interested in the distribution of one random variable regardless of all others, the *marginal distribution* can be used. Given the joint PDF $f_{x,y}(x, y)$, the *marginal PDF* of a random variable X can be obtained as

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x, y)\, dy \tag{2.16}$$

For continuous random variables, the *conditional PDF* for $X \mid Y$, similar to the conditional probability shown in Eq. (2.6), can be defined as

$$f_x(x \mid y) = \frac{f_{x,y}(x, y)}{f_y(y)} \tag{2.17}$$

in which $f_y(y)$ is the marginal PDF of random variable Y. The conditional PMF for two discrete random variables similarly can be defined as

$$p_x(x \mid y) = \frac{p_{x,y}(x, y)}{p_y(y)} \tag{2.18}$$

Figure 2.9 shows the joint and marginal PDFs of two continuous random variables X and Y. It can be shown easily that when the two random variables are statistically independent, $f_x(x \mid y) = f_x(x)$.

Equation (2.17) alternatively can be written as

$$f_{x,y}(x, y) = f_x(x \mid y) \times f_y(y) \tag{2.19}$$

which indicates that a joint PDF between two correlated random variables can be formulated by multiplying a conditional PDF and a suitable marginal PDF.
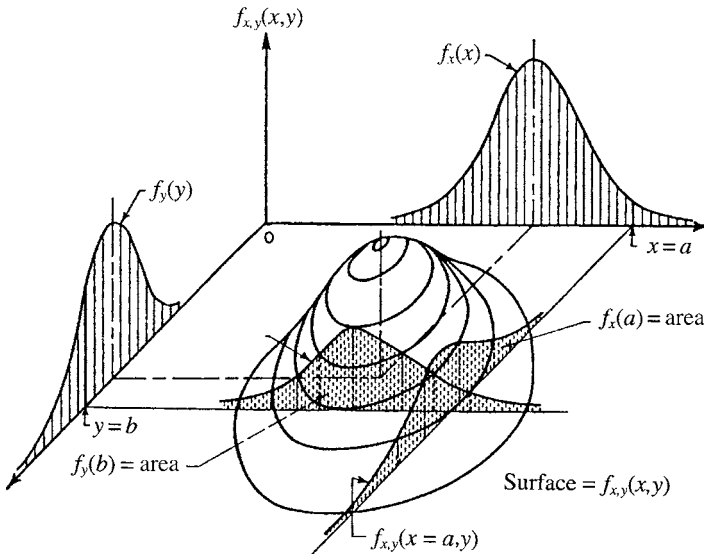
Figure 2.9 Joint and marginal probability density function (PDFs) of two continuous random variables. (*After Ang and Tang, 1975.*)

Note that the marginal distributions can be obtained from the joint distribution function, but not vice versa.

Example 2.7 Suppose that X and Y are two random variables that can only take values in the intervals $0 \leq x \leq 2$ and $0 \leq y \leq 2$. Suppose that the joint CDF of X and Y for these intervals has the form of $F_{x,y}(x, y) = cxy(x^2 + y^2)$. Find (a) the joint PDF of X and Y, (b) the marginal PDF of X, (c) the conditional PDF $f_y(y \mid x = 1)$, and (d) $P(Y \leq 1 \mid x = 1)$.

Solution First, one has to find the constant c so that the function $F_{x,y}(x, y)$ is a legitimate CDF. It requires that the value of $F_{x,y}(x, y) = 1$ when both arguments are at their respective upper bounds. That is,

$$F_{x,y}(x = 2, \ y = 2) = 1 = c(2)(2)(2^2 + 2^2)$$

Therefore, $c = 1/32$. The resulting joint CDF is shown in Fig. 2.10a.

(a)

(b)
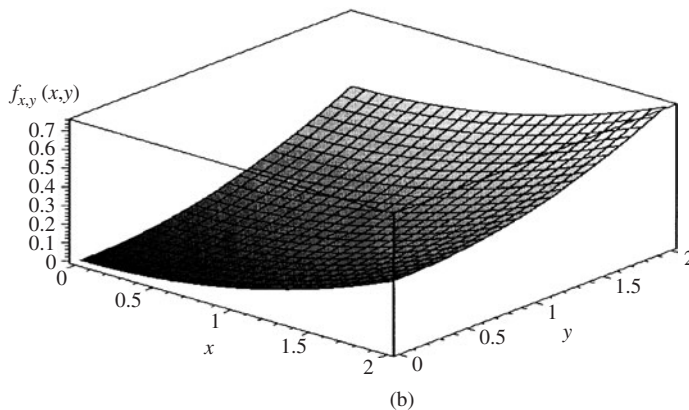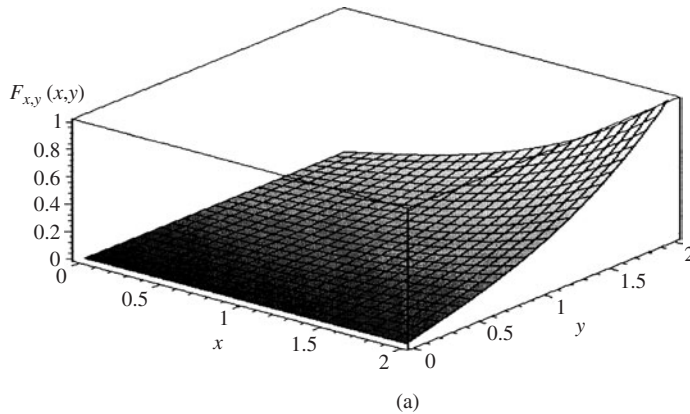
Figure 2.10 (a) Joint cumulative distribution function (CDF) and (b) probability density function (PDF) for Example 2.7.

(a) To derive the joint PDF, Eq. (2.15a) is applied, that is,

$$f_{x,y}(x, y) = \frac{\partial}{\partial x} \left\{ \frac{\partial}{\partial y} \left[\frac{xy(x^2 + y^2)}{32} \right] \right\} = \frac{\partial}{\partial x} \left(\frac{x^3 + 3xy^2}{32} \right) = \frac{3(x^2 + y^2)}{32} \quad \text{for } 0 \leq x, \, y \leq 2$$

A plot of joint PDF is shown in Fig. 2.10b.

(b) To find the marginal distribution of X, Eq. (2.16) can be used:

$$f_x(x) = \int_0^2 \frac{3(x^2 + y^2)}{32} \, dx = \frac{4 + 3x^2}{16} \qquad \text{for } 0 \leq x \leq 2$$

(c) The conditional distribution $f_y(y \mid x)$ can be obtained by following Eq. (2.17) as

$$f_y(y \mid x = 1) = \frac{f_{x,y}(x = 1, y)}{f_x(x = 1)} = \frac{\frac{3[(1)^2 + y^2]}{32}}{\frac{4 + 3(1)^2}{16}} = \frac{3(1 + y^2)}{14} \qquad \text{for } 0 \leq y \leq 2$$

(d) The conditional probability $P(Y \leq 1 \mid X = 1)$ can be computed as

$$P(Y \leq 1 \mid X = 1) = \int_0^1 f_y(y \mid x = 1) \, dy = \int_0^1 \frac{3(1 + y^2)}{14} \, dy = \frac{2}{7}$$

2.4 Statistical Properties of Random Variables

In statistics, the term *population* is synonymous with the sample space, which describes the complete assemblage of all the values representative of a particular random process. A *sample* is any subset of the population. Furthermore, *parameters* in a statistical model are quantities that are descriptive of the population. In this book, Greek letters are used to denote statistical parameters. *Sample statistics*, or simply *statistics*, are quantities calculated on the basis of sample observations.

2.4.1 Statistical moments of random variables

In practical statistical applications, descriptors commonly used to show the statistical properties of a random variable are those indicative of (1) the central tendency, (2) the dispersion, and (3) the asymmetry of a distribution. The frequently used descriptors in these three categories are related to the *statistical moments* of a random variable. Currently, two types of statistical moments are used in hydrosystems engineering applications: product-moments and L-moments. The former is a conventional one with a long history of practice, whereas the latter has been receiving great attention recently from water resources engineers in analyzing hydrologic data (Stedinger et al., 1993; Rao and Hamed 2000). To be consistent with the current general practice and usage, the terms *moments* and *statistical moments* in this book refer to the conventional product-moments unless otherwise specified.

Product-moments.   The r th-order *product-moment* of a random variable X about any reference point $X = x_o$ is defined, for the continuous case, as

$$E[(X - x_o)^r] = \int_{-\infty}^{\infty} (x - x_o)^r f_x(x)\, dx = \int_{-\infty}^{\infty} (x - x_o)^r \, dF_x(x) \qquad (2.20a)$$

whereas for the discrete case,

$$E[(X - x_o)^r] = \sum_{k=1}^{K} (x_k - x_o)^r p_x(x_k) \qquad (2.20b)$$

where $E[\cdot]$ is a *statistical expectation operator*. In practice, the first three moments ($r = 1, 2, 3$) are used to describe the central tendency, variability, and asymmetry of the distribution of a random variable. Without losing generality, the following discussions consider continuous random variables. For discrete random variables, the integral sign is replaced by the summation sign. Here it is convenient to point out that when the PDF in Eq. (2.20a) is replaced by a conditional PDF, as described in Sec. 2.3, the moments obtained are called the *conditional moments*.

Since the expectation operator $E[\cdot]$ is for determining the average value of the random terms in the brackets, the sample estimator for the product-moments for $\mu'_r = E(X_r)$, based on n available data $(x_1, x_2, \ldots, x_n)$, can be written as

$$\widehat{\mu'_r} = \sum_{i=1}^{n} w_i(n)\, x_i^r \qquad (2.21)$$

where $w_i(n)$ is a weighting factor for sample observation x_i, which depends on sample size n. Most commonly, $w_i(n) = 1/n$, for all $i = 1, 2, \ldots, n$. The last column of Table 2.1 lists the formulas applied in practice for computing some commonly used statistical moments.

Two types of product-moments are used commonly: *moments about the origin*, where $x_o = 0$, and *central moments*, where $x_o = \mu_x$, with $\mu_x = E[X]$. The r th-order central moment is denoted as $\mu_r = E[(X - \mu_x)^r]$, whereas the r th-order moment about the origin is denoted as $\mu'_r = E(X^r)$. It can be shown easily, through the *binomial expansion*, that the central moments $\mu_r = E[(X - \mu_x)^r]$ can be obtained from the moments about the origin as

$$\mu_r = \sum_{i=0}^{r} (-1)^i C_{r,i}\, \mu_x^i\, \mu'_{r-i} \qquad (2.22)$$

where $C_{r,i} = \binom{r}{i} = \frac{r!}{i!(r-i)!}$ is a *binomial coefficient*, with ! representing factorial, that is, $r! = r \times (r - 1) \times (r - 2) \times \cdots \times 2 \times 1$. Conversely, the moments about the origin can be obtained from the central moments in a similar fashion as

$$\mu'_r = \sum_{i=0}^{r} C_{r,i}\, \mu_x^i\, \mu_{r-i} \qquad (2.23)$$

TABLE 2.1 Product-Moments of Random Variables

Moment	Measure of	Definition	Continuous variable	Discrete variable	Sample estimator
First	Central location	Mean, expected value $E(X) = \mu_x$	$\mu_x = \int_{-\infty}^{\infty} x f_x(x)\, dx$	$\mu_x = \sum_{\text{all } x's} x_k p(x_k)$	$\bar{x} = \sum x_i / n$
Second	Dispersion	Variance, $\text{Var}(X) = \mu_2 = \sigma_x^2$	$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x)\, dx$	$\sigma_x^2 = \sum_{\text{all } x's} (x_k - \mu_x)^2 P_x(x_k)$	$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
		Standard deviation, σ_x	$\sigma_x = \sqrt{\text{Var}(X)}$	$\sigma_x = \sqrt{\text{Var}(X)}$	$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$
		Coefficient of variation, Ω_x	$\Omega_x = \sigma_x / \mu_x$	$\Omega_x = \sigma_x / \mu_x$	$C_v = s / \bar{x}$
Third	Asymmetry	Skewness	$\mu_3 = \int_{-\infty}^{\infty} (x - \mu_x)^3 f_x(x)\, dx$	$\mu_3 = \sum_{\text{all } x's} (x_k - \mu_x)^3 p_x(x_k)$	$m_3 = \frac{n}{(n-1)(n-2)} \sum (x_i - \bar{x})^3$
		Skewness coefficient, γ_x	$\gamma_x = \mu_3 / \sigma_x^3$	$\gamma_x = \mu_3 / \sigma_x^3$	$g = m_3 / s^3$
Fourth	Peakedness	Kurtosis, κ_x	$\mu_4 = \int_{-\infty}^{\infty} (x - \mu_x)^4 f_x(x)\, dx$	$\mu_4 = \sum_{\text{all } x's} (x_k - \mu_x)^4 p_x(x_k)$	$m_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum (x_i - \bar{x})^4$
		Excess coefficient, ε_x	$\kappa_x = \mu_4 / \sigma_x^4$	$\kappa_x = \mu_4 / \sigma_x^4$	$k = m_4 / s^4$
			$\varepsilon_x = \kappa_x - 3$	$\varepsilon_x = \kappa_x - 3$	

Equation (2.22) enables one to compute central moments from moments about the origin, whereas Eq. (2.23) does the opposite. Derivations for the expressions of the first four central moments and the moments about the origin are left as exercises (Problems 2.10 and 2.11).

The main disadvantages of the product-moments are (1) that estimation from sample observations is sensitive to the presence of extraordinary values (called *outliers*) and (2) that the accuracy of sample product-moments deteriorates rapidly with an increase in the order of the moments. An alternative type of moments, called *L-moments*, can be used to circumvent these disadvantages.

Example 2.8 (after Tung and Yen, 2005) Referring to Example 2.6, determine the first two moments about the origin for the time to failure of the pump. Then calculate the first two central moments.

Solution From Example 2.6, the random variable T is the time to failure having an exponential PDF as

$$f_t(t) = \left(\frac{1}{\beta}\right) \exp(-t/1250) \qquad \text{for } t \geq 0, \ \beta > 0$$

in which t is the elapsed time (in hours) before the pump fails, and $\beta = 1250$ h/failure.

The moments about the origin, according to Eq. (2.20a), are

$$E(T^r) = \mu'_r = \int_0^\infty t^r \left(\frac{e^{-t/\beta}}{\beta}\right) dt$$

Using integration by parts, the results of this integration are

for $r = 1$, $\qquad\qquad\qquad \mu'_1 = E(T) = \mu_t = \beta = 1250$ h

for $r = 2$, $\qquad\qquad\qquad \mu'_2 = E(T^2) = 2\beta^2 = 3{,}125{,}000$ h^2

Based on the moments about the origin, the central moments can be determined, according to Eq. (2.22) or Problem (2.10), as

for $r = 1$, $\qquad \mu_1 = E(T - \mu_t) = 0$

for $r = 2$, $\qquad \mu_2 = E[(T - \mu_t)^2] = \mu'_2 - \mu^2 = 2\beta^2 - \beta^2 = \beta^2 = 1{,}562{,}500$ h^2

L-moments. The rth-order *L-moments* are defined as (Hosking, 1986, 1990)

$$\lambda_r = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E(X_{r-j:r}) \qquad r = 1, 2, \ldots \tag{2.24}$$

in which $X_{j:n}$ is the jth-*order statistic* of a random sample of size n from the distribution $F_x(x)$, namely, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(j)} \leq \cdots \leq X_{(n)}$. The "L" in L-moments emphasizes that λ_r is a linear function of the expected order statistics. Therefore, sample L-moments can be made a linear combination of the ordered data values. The definition of the L-moments given in Eq. (2.24) may appear to be mathematically perplexing; the computations, however, can be simplified greatly through their relations with the *probability-weighted moments*,

which are defined as (Greenwood et al., 1979)

$$M_{r,p,q} = E\{X^r [F_x(X)]^p [1 - F_x(X)]^q\} = \int_{-\infty}^{\infty} x^r [F_x(x)]^p [1 - F_x(x)]^q \, dF_x(x)$$
(2.25)

Compared with Eq. (2.20a), one observes that the conventional product-moments are a special case of the probability-weighted moments with $p = q = 0$, that is, $M_{r,0,0} = \mu_r'$. The probability-weighted moments are particularly attractive when the closed-form expression for the CDF of the random variable is available.

To work with the random variable linearly, $M_{1,p,q}$ can be used. In particular, two types of probability-weighted moments are used commonly in practice, that is,

$$\alpha_r = M_{1,0,r} = E\{X[1 - F_x(X)]^r\} \qquad r = 0, 1, 2, \ldots \qquad (2.26a)$$

$$\beta_r = M_{1,r,0} = E\{X[F_x(X)]^r\} \qquad r = 0, 1, 2, \ldots \qquad (2.26b)$$

In terms of α_r or β_r, the rth-order L-moment λ_r can be obtained as (Hosking, 1986)

$$\lambda_{r+1} = (-1)^r \sum_{j=0}^{r} p_{r,j}^* \alpha_j = \sum_{j=0}^{r} p_{r,j}^* \beta_j \qquad r = 0, 1, \ldots \qquad (2.27)$$

in which

$$p_{r,j}^* = (-1)^{r-j} \binom{r}{j} \binom{r+i}{j} = \frac{(-1)^{r-j}(r+j)!}{(j!)^2(r-j)!}$$

For example, the first four L-moments of random variable X are

$$\lambda_1 = \beta_0 = \mu_1' = \mu_x \qquad (2.28a)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \qquad (2.28b)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \qquad (2.28c)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \qquad (2.28d)$$

To estimate sample α- and β-moments, random samples are arranged in ascending or descending order. For example, arranging n random observations in ascending order, that is, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(j)} \leq \cdots \leq X_{(n)}$, the rth-order β-moment β_r can be estimated as

$$\widehat{\beta_r} = \frac{1}{n} \sum_{i=1}^{n} X_{(i)} \widehat{F}(X_{(i)})^r \qquad (2.29)$$

where $\widehat{F}(X_{(i)})$ is an estimator for $F(X_{(i)}) = P(X \leq X_{(i)})$, for which many *plotting-position formulas* have been used in practice (Stedinger et al., 1993).

The one that is used often is the *Weibull plotting-position formula*, that is, $\widehat{F}(X_{(i)}) = i/(n+1)$.

L-moments possess several advantages over conventional product-moments. Estimators of L-moments are more robust against outliers and are less biased. They approximate asymptotic normal distributions more rapidly and closely. Although they have not been used widely in reliability applications as compared with the conventional product-moments, L-moments could have a great potential to improve reliability estimation. However, before more evidence becomes available, this book will limit its discussions to the uses of conventional product-moments.

Example 2.9 (after Tung and Yen, 2005) Referring to Example 2.8, determine the first two L-moments, that is, λ_1 and λ_2, of random time to failure T.

Solution To determine λ_1 and λ_2, one first calculates β_0 and β_1, according to Eq. (2.26b), as

$$\beta_0 = E\{T [F_t(T)]^0\} = E(T) = \mu_t = \beta$$

$$\beta_1 = E\{T [F_t(T)]^1\} = \int_0^\infty [t \, F_t(t)] f_t(t) \, dt = \int_0^\infty [t(1 - e^{-t/\beta})](e^{-t/\beta}/\beta) \, dt = \tfrac{3}{4}\beta$$

From Eq. (2.28), the first two L-moments can be computed as

$$\lambda_1 = \beta_0 = \mu_t = \beta \qquad \lambda_2 = 2\beta_1 - \beta_0 = \frac{6\beta}{4} - \beta = \frac{\beta}{2}$$

2.4.2 Mean, mode, median, and quantiles

The central tendency of a continuous random variable X is commonly represented by its *expectation*, which is the first-order moment about the origin:

$$E(X) = \mu_x = \int_{-\infty}^\infty x f_x(x) \, dx = \int_{-\infty}^\infty x \, dF_x(x) = \int_{-\infty}^\infty [1 - F_x(x)] \, dx \qquad (2.30)$$

This expectation is also known as the *mean* of a random variable. It can be seen easily that the mean of a random variable is the first-order L-moment λ_1. Geometrically, the mean or expectation of a random variable is the location of the centroid of the PDF or PMF. The second and third integrations in Eq. (2.30) indicate that the mean of a random variable is the shaded area shown in Fig. 2.11.

The following two operational properties of the expectation are useful:

1. The expectation of the sum of several random variables (regardless of their dependence) equals the sum of the expectation of the individual random
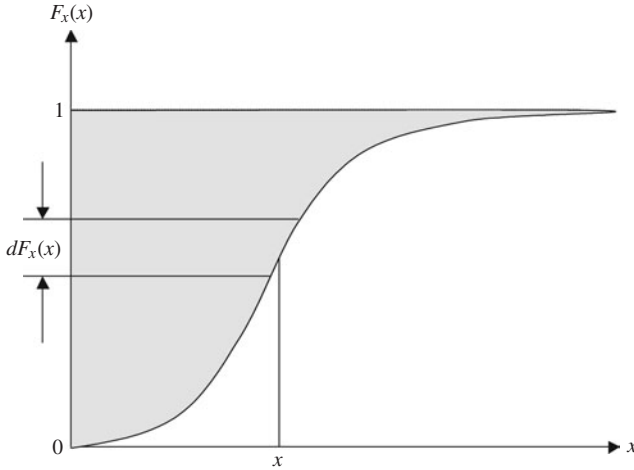
Figure 2.11 Geometric interpretation of the mean.

variable, that is,

$$E\left(\sum_{k=1}^{K} a_k X_k\right) = \sum_{k=1}^{K} a_k \mu_k \tag{2.31}$$

in which $\mu_k = E(X_k)$, for $k = 1, 2, \ldots, K$.

2. The expectation of multiplication of several independent random variables equals the product of the expectation of the individual random variables, that is,

$$E\left(\prod_{k=1}^{K} X_k\right) = \prod_{k=1}^{K} \mu_k \tag{2.32}$$

Two other types of measures of central tendency of a random variable, namely, the median and mode, are sometimes used in practice. The *median* of a random variable is the value that splits the distribution into two equal halves. Mathematically, the median x_{md} of a continuous random variable satisfies

$$F_x(x_{\text{md}}) = \int_{-\infty}^{x_{\text{md}}} f_x(x)\, dx = 0.5 \tag{2.33}$$

The median, therefore, is the 50th *quantile* (or *percentile*) of random variable X. In general, the $100p$th quantile of a random variable X is a quantity x_p that satisfies

$$P(X \le x_p) = F_x(x_p) = p \tag{2.34}$$

The *mode* is the value of a random variable at which the value of a PDF is peaked. The mode x_{mo} of a random variable X can be obtained by solving the

following equation:

$$\left[\frac{\partial f_x(x)}{\partial x}\right]_{x=x_{mo}} = 0 \qquad (2.35)$$

Referring to Fig. 2.12, a PDF could be unimodal with a single peak, bimodal with two peaks, or multimodal with multiple peaks. Generally, the mean, median, and mode of a random variable are different unless the PDF is symmetric and unimodal. Descriptors for the central tendency of a random variable are summarized in Table 2.1.

Example 2.10 (after Tung and Yen, 2005)   Refer to Example 2.8, the pump reliability problem. Find the mean, mode, median, and 10 percent quantile for the random time to failure T.

Solution   The mean of the time to failure, called the *mean time to failure* (MTTF), is the first-order moment about the origin, which is $\mu_t = 1250$ h as calculated previously in Example 2.8. From the shape of the PDF for the exponential distribution as shown in Fig. 2.7, one can immediately identify that the mode, representing the most likely time of pump failure, is at the beginning of pump operation, that is, $t_{mo} = 0$ h.

(a)

(b)
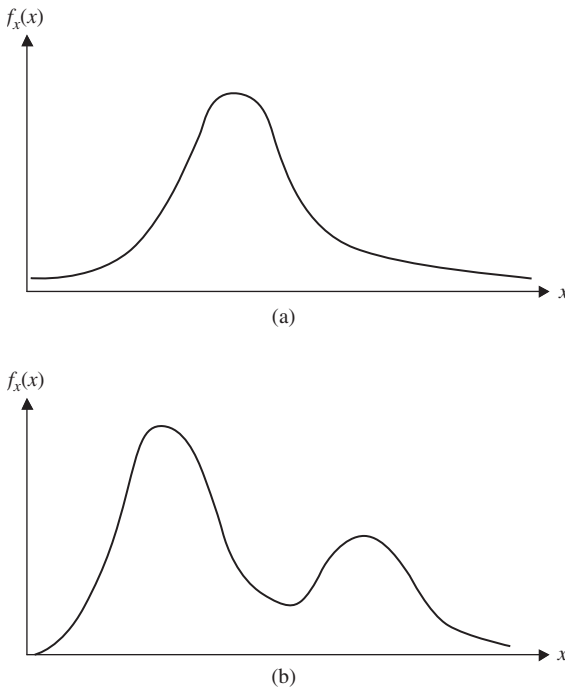
Figure 2.12   Unimodal (a) and bimodal (b) distributions.

To determine the median time to failure of the pump, one can first derive the expression for the CDF from the given exponential PDF as

$$F_t(t) = P(T \le t) = \int_0^t \frac{e^{-u/1250}}{1250} \, du = 1 - e^{-t/1250} \qquad \text{for } t \ge 0$$

in which u is a dummy variable. Then the median time to failure t_{md} can be obtained, according to Eq. (2.33), by solving

$$F_t(t_{\mathrm{md}}) = 1 - \exp(-t_{\mathrm{md}}/1250) = 0.5$$

which yields $t_{\mathrm{md}} = 866.43$ h.

Similarly, the 10 percent quantile $t_{0.1}$, namely, the elapsed time over which the pump would fail with a probability of 0.1, can be found in the same way as the median except that the value of the CDF is 0.1, that is,

$$F_t(t_{0.1}) = 1 - \exp(-t_{0.1}/1250) = 0.1$$

which yields $t_{0.1} = 131.7$ h.

2.4.3 Variance, standard deviation, and coefficient of variation

The spreading of a random variable over its range is measured by the *variance*, which is defined for the continuous case as

$$\mathrm{Var}(X) = \mu_2 = \sigma_x^2 = E\left[(X - \mu_x)^2\right] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x) \, dx \qquad (2.36)$$

The variance is the second-order central moment. The positive square root of the variance is called the *standard deviation σ_x*, which is often used as a measure of the degree of uncertainty associated with a random variable.

The standard deviation has the same units as the random variable. To compare the degree of uncertainty of two random variables with different units, a dimensionless measure $\Omega_x = \sigma_x/\mu_x$, called the *coefficient of variation*, is useful. By its definition, the coefficient of variation indicates the variation of a random variable relative to its mean. Similar to the standard deviation, the second-order L-moment λ_2 is a measure of dispersion of a random variable. The ratio of λ_2 to λ_1, that is, $\tau_2 = \lambda_2/\lambda_1$, is called the *L-coefficient of variation*.

Three important properties of the variance are

1. $\mathrm{Var}(a) = 0$ when a is a constant. $\hfill (2.37)$

2. $\mathrm{Var}(X) = E(X^2) - E^2(X) = \mu_2' - \mu_x^2$ $\hfill (2.38)$

3. The variance of the sum of several independent random variables equal the sum of variance of the individual random variables, that is,

$$\mathrm{Var}\left(\sum_{k=1}^{K} a_k X_k\right) = \sum_{k=1}^{K} a_k^2 \, \sigma_k^2 \qquad (2.39)$$

where a_k is a constant, and σ_k is the standard deviation of random variable X_k, $k = 1, 2, \ldots, K$.

Example 2.11 (modified from Mays and Tung, 1992) Consider the mass balance of a surface reservoir over a 1-month period. The end-of-month storage S can be computed as

$$S_{m+1} = S_m + P_m + I_m - E_m - r_m$$

in which the subscript m is an indicator for month, S_m is the initial storage volume in the reservoir, P_m is the precipitation amount on the reservoir surface, I_m is the surface-runoff inflow, E_m is the total monthly evaporation amount from the reservoir surface, and r_m is the controlled monthly release volume from the reservoir.

It is assumed that at the beginning of the month, the initial storage volume and total monthly release are known. The monthly total precipitation amount, surface-runoff inflow, and evaporation are uncertain and are assumed to be independent random variables. The means and standard deviations of P_m, I_m, and E_m from historical data for month m are estimated as

$$E(P_m) = 1000\,\text{m}^3, \quad E(I_m) = 8000\,\text{m}^3, \quad E(E_m) = 3000\,\text{m}^3$$

$$\sigma(P_m) = 500\,\text{m}^3, \quad \sigma(I_m) = 2000\,\text{m}^3, \quad \sigma(E_m) = 1000\,\text{m}^3$$

Determine the mean and standard deviation of the storage volume in the reservoir by the end of the month if the initial storage volume is 20,000 m³ and the designated release for the month is 10,000 m³.

Solution From Eq. (2.31), the mean of the end-of-month storage volume in the reservoir can be determined as

$$E(S_{m+1}) = S_m + E(P_m) + E(I_m) - E(E_m) - r_m$$

$$= 20,000 + 1000 + 8000 - 3000 - 10,000 = 16,000\,\text{m}^3$$

Since the random hydrologic variables are statistically independent, the variance of the end-of-month storage volume in the reservoir can be obtained, from Eq. (2.39), as

$$\text{Var}(S_{m+1}) = \text{Var}(P_m) + \text{Var}(I_m) + \text{Var}(E_m)$$

$$= [(0.5)^2 + (2)^2 + (1)^2] \times (1000\,\text{m}^3)^2 = 5.25 \times (1000\,\text{m}^3)^2$$

The standard deviation and coefficient of variation of S_{m+1} then are

$$\sigma(S_{m+1}) = \sqrt{5.25} \times 1000 = 2290\,\text{m}^3 \quad \text{and} \quad \Omega(S_{m+1}) = 2290/16{,}000 = 0.143$$

2.4.4 Skewness coefficient and kurtosis

The asymmetry of the PDF of a random variable is measured by the *skewness coefficient* γ_x, defined as

$$\gamma_x = \frac{\mu_3}{\mu_2^{1.5}} = \frac{E\left[(X - \mu_x)^3\right]}{\sigma_x^3} \tag{2.40}$$

The skewness coefficient is dimensionless and is related to the third-order central moment. The sign of the skewness coefficient indicates the degree of symmetry of the probability distribution function. If $\gamma_x = 0$, the distribution is symmetric about its mean. When $\gamma_x > 0$, the distribution has a long tail to the right, whereas $\gamma_x < 0$ indicates that the distribution has a long tail to the left. Shapes of distribution functions with different values of skewness coefficients and the relative positions of the mean, median, and mode are shown in Fig. 2.13.
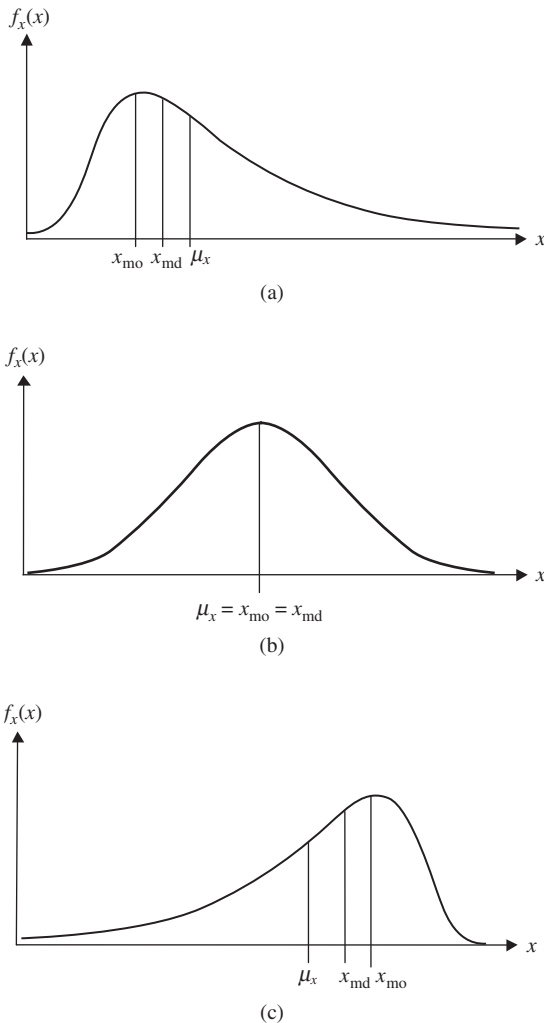
Figure 2.13 Relative locations of mean, median, and mode for (a) positively skewed, (b) symmetric and (c) negatively skewed distributions.

Similarly, the degree of asymmetry can be measured by the *L-skewness coefficient* τ_3, defined as

$$\tau_3 = \lambda_3 / \lambda_2 \tag{2.41}$$

The value of the L-skewness coefficient for all feasible distribution functions must lie within the interval of $[-1, 1]$ (Hosking, 1986).

Another indicator of the asymmetry is the *Pearson skewness coefficient*, defined as

$$\gamma_1 = \frac{\mu_x - x_{\mathrm{mo}}}{\sigma_x} \tag{2.42}$$

As can be seen, the Pearson skewness coefficient does not require computing the third-order moment. In practice, product-moments higher than the third order are used less because they are unreliable and inaccurate when estimated from a small number of samples. Equations used to compute the sample product-moments are listed in the last column of Table 2.1.

Kurtosis κ_x is a measure of the peakedness of a distribution. It is related to the fourth-order central moment of a random variable as

$$\kappa_x = \frac{\mu_4}{\mu_2^2} = \frac{E\left[(X - \mu_x)^4\right]}{\sigma_x^4} \tag{2.43}$$

with $\kappa_x > 0$. For a random variable having a normal distribution (Sec. 2.6.1), its kurtosis is equal to 3. Sometimes the *coefficient of excess*, defined as $\varepsilon_x = \kappa_x - 3$, is used. For all feasible distribution functions, the skewness coefficient and kurtosis must satisfy the following inequality relationship (Stuart and Ord, 1987)

$$\gamma_x^2 + 1 \leq \kappa_x \tag{2.44}$$

By the definition of L-moments, the *L-kurtosis* is defined as

$$\tau_4 = \lambda_4 / \lambda_2 \tag{2.45}$$

Similarly, the relationship between the L-skewness and L-kurtosis for all feasible probability distribution functions must satisfy (Hosking, 1986)

$$\frac{5\tau_3^2 - 1}{4} \leq \tau_4 < 1 \tag{2.46}$$

Royston (1992) conducted an analysis comparing the performance of sample skewness and kurtosis defined by the product-moments and L-moments. Results indicated that the L-skewness and L-kurtosis have clear advantages

over the conventional product-moments in terms of being easy to interpret, fairly robust to outliers, and less unbiased in small samples.

2.4.5 Covariance and correlation coefficient

When a problem involves two dependent random variables, the degree of linear dependence between the two can be measured by the *correlation coefficient* $\rho_{x,y}$, which is defined as

$$\text{Corr}(X, Y) = \rho_{x,y} = \text{Cov}(X, Y)/\sigma_x \sigma_y \tag{2.47}$$

where $\text{Cov}(X, Y)$ is the *covariance* between random variables X and Y, defined as

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - \mu_x \mu_y \tag{2.48}$$

Various types of correlation coefficients have been developed in statistics for measuring the degree of association between random variables. The one defined by Eq. (2.47) is called the *Pearson product-moment correlation coefficient*, or correlation coefficient for short in this and general use.

It can be shown easily that $\text{Cov}(X'_1, X'_2) = \text{Corr}(X_1, X_2)$, with X'_1 and X'_2 being the *standardized random variables*. In probability and statistics, a random variable can be standardized as

$$X' = (X - \mu_x)/\sigma_x \tag{2.49}$$

Hence a standardized random variable has zero mean and unit variance. Standardization will not affect the skewness coefficient and kurtosis of a random variable because they are dimensionless.

Figure 2.14 graphically illustrates several cases of the correlation coefficient. If the two random variables X and Y are statistically independent, then $\text{Corr}(X, Y) = \text{Cov}(X, Y) = 0$ (Fig. 2.14c). However, the reverse statement is not necessarily true, as shown in Fig. 2.14d. If the random variables involved are not statistically independent, Eq. (2.70) for computing the variance of the sum of several random variables can be generalized as

$$\text{Var}\left(\sum_{k=1}^{K} a_k X_k\right) = \sum_{k=1}^{K} a_k^2 \sigma_k^2 + 2 \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} a_k a_{k'} \text{Cov}(X_k, X_{k'}) \tag{2.50}$$

Example 2.12 (after Tung and Yen, 2005)   Perhaps the assumption of independence of P_m, I_m, and E_m in Example 2.11 may not be reasonable in reality. One examines the historical data closely and finds that correlations exist among the three hydrologic random variables. Analysis of data reveals that $\text{Corr}(P_m, I_m) = 0.8$, $\text{Corr}(P_m, E_m) = -0.4$, and $\text{Corr}(I_m, E_m) = -0.3$. Recalculate the standard deviation associated with the end-of-month storage volume.
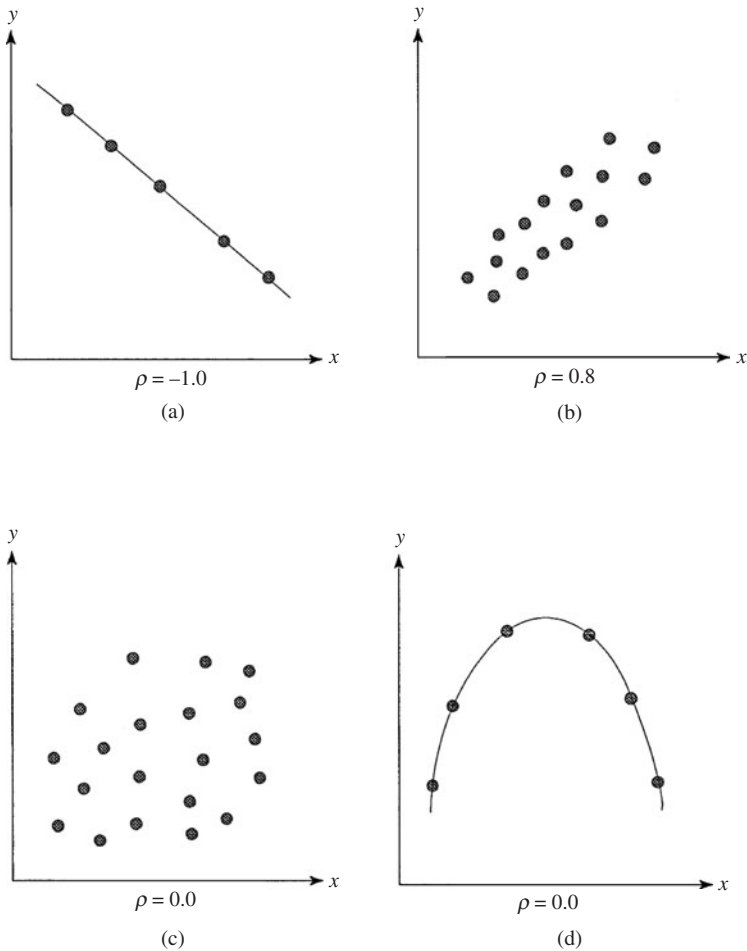
Figure 2.14 Different cases of correlation between two random variables: (*a*) perfectly linearly correlated in opposite directions; (*b*) strongly linearly correlated in a positive direction; (*c*) uncorrelated in linear fashion; (*d*) perfectly correlated in nonlinear fashion but uncorrelated linearly.

Solution By Eq. (2.50), the variance of the reservoir storage volume at the end of the month can be calculated as

$$\text{Var}(S_{m+1}) = ar(P_m) + \text{Var}(I_m) + \text{Var}(E_m) + 2\,\text{Cov}(P_m, I_m)$$

$$- 2\,\text{Cov}(P_m, E_m) - 2\,\text{Cov}(I_m, E_m)$$

$$= \text{Var}(P_m) + \text{Var}(I_m) + \text{Var}(E_m) + 2\,\text{Corr}(P_m, I_m)\sigma(P_m)\sigma(I_m)$$

$$- 2\text{Corr}(P_m, E_m)\sigma(P_m)\sigma(E_m) - 2\,\text{Corr}(I_m, E_m)\sigma(I_m)\sigma(E_m)$$

$$= (500)^2 + (2000)^2 + (1000)^2 + 2(0.8)(500)(2000)$$

$$- 2(-0.4)(500)(1000) - 2(-0.3)(2000)(1000)$$

$$= 8.45(1000 \text{ m}^3)^2$$

The corresponding standard deviation of the end-of-month storage volume is

$$\sigma(S_{m+1}) = \sqrt{8.45} \times 1000 = 2910 \text{ m}^3$$

In this case, consideration of correlation increases the standard deviation by 27 percent compared with the uncorrelated case in Example 2.11.

Example 2.13 Referring to Example 2.7, compute correlation coefficient between X and Y.

Solution Referring to Eqs. (2.47) and (2.48), computation of the correlation coefficient requires the determination of μ_x, μ_y, σ_x, and σ_y from the marginal PDFs of X and Y:

$$f_x(x) = \frac{4 + 3x^2}{16} \quad \text{for } 0 \le x \le 2 \qquad f_y(y) = \frac{4 + 3y^2}{16} \quad \text{for } 0 \le y \le 2$$

as well as $E(XY)$ from their joint PDF obtained earlier:

$$f_{x,y}(x, y) = \frac{3(x^2 + y^2)}{32} \qquad \text{for } 0 \le x, y \le 2$$

From the marginal PDFs, the first two moments of X and Y about the origin can be obtained easily as

$$\mu_x = E(X) = \int_0^2 x f_x(x)\,dx = \tfrac{5}{4} = E(Y) = \mu_y \qquad E(X^2) = \int_0^2 x^2 f_x(x)\,dx = \tfrac{28}{15} = E(Y^2)$$

Hence the variances of X and Y can be calculated as

$$\text{Var}(X) = E(X^2) - (\mu_x)^2 = 73/240 = \text{Var}(Y)$$

To calculate $\text{Cov}(X, Y)$, one could first compute $E(XY)$ from the joint PDF as

$$E(XY) = \int_0^2 \int_0^2 xy\, f_{x,y}(x, y)\,dx\,dy = \tfrac{3}{2}$$

Then the covariance of X and Y, according to Eq. (2.48), is

$$\text{Cov}(X, Y) = E(XY) - \mu_x \mu_y = -1/16$$

The correlation between X and Y can be obtained as

$$\text{Corr}(X, Y) = \rho_{x,y} = \frac{-1/16}{73/240} = -0.205$$

2.5   Discrete Univariate Probability Distributions

In the reliability analysis of hydrosystems engineering problems, several probability distributions are used frequently. Based on the nature of the random variable, probability distributions are classified into discrete and continuous types. In this section, two discrete distributions, namely, the binomial distribution and the Poisson distribution, that are used commonly in hydrosystems reliability analysis, are described. Section 2.6 describes several frequently used univariate continuous distributions. For the distributions discussed in this chapter and others not included herein, their relationships are shown in Fig. 2.15.
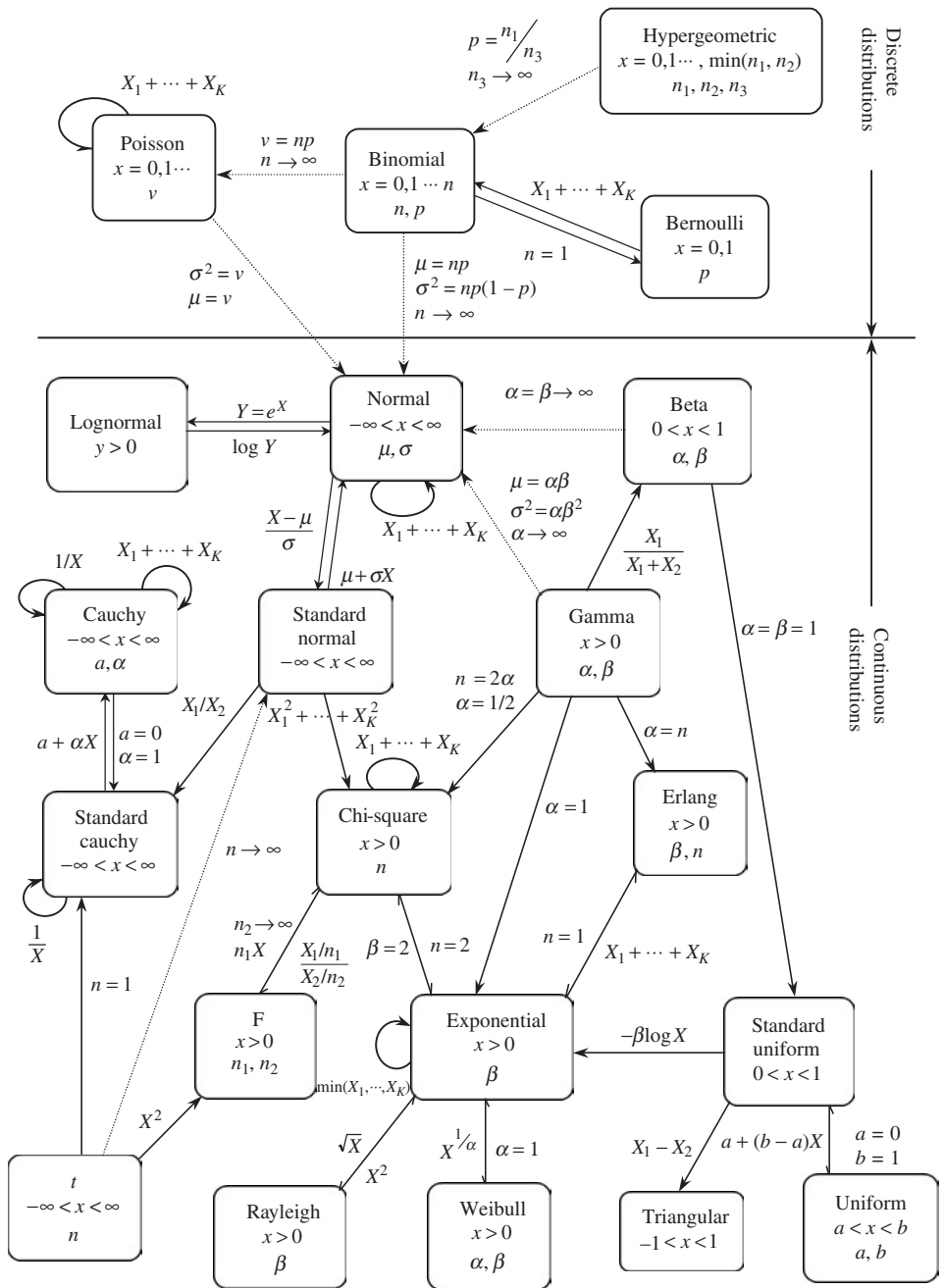
Figure 2.15   Relationships among univariate distributions. (*After Leemis, 1986.*)

Computations of probability and quantiles for the great majority of the distribution functions described in Secs. 2.5 and 2.6 are available in Microsoft Excel.

2.5.1 Binomial distribution

The *binomial distribution* is applicable to random processes with only two types of outcomes. The state of components or subsystems in many hydrosystems can be classified as either functioning or failed, which is a typical example of a binary outcome. Consider an experiment involving a total of n independent trials with each trial having two possible outcomes, say, success or failure. In each trial, if the probability of having a successful outcome is p, the probability of having x successes in n trials can be computed as

$$p_x(x) = C_{n,x} p^x q^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots, n \qquad (2.51)$$

where $C_{n,x}$ is the binomial coefficient, and $q = 1 - p$, the probability of having a failure in each trial. Computationally, it is convenient to use the following recursive formula for evaluating the binomial PMF (Drane et al., 1993):

$$p_x(x \,|\, n, p) = \left(\frac{n + 1 - x}{x} \right) \left(\frac{p}{q} \right) p_x(x - 1 \,|\, n, p) = R_B(x) p_x(x - 1 \,|\, n, p) \quad (2.52)$$

for $x = 0, 1, 2, \ldots, n$, with the initial probability $p_x(x = 0 | n, p) = q^n$. A simple recursive scheme for computing the binomial cumulative probability is given by Tietjen (1994).

A random variable X having a binomial distribution with parameters n and p has the expectation $E(X) = np$ and variance $\text{Var}(X) = npq$. Shape of the PMF of a binomial random variable depends on the values of p and q. The skewness coefficient of a binomial random variable is $(q - p)/\sqrt{npq}$. Hence the PMF is positively skewed if $p < q$, symmetric if $p = q = 0.5$, and negatively skewed if $p > q$. Plots of binomial PMFs for different values of p with a fixed n are shown in Fig. 2.16. Referring to Fig. 2.15, the sum of several independent binomial random variables, each with a common parameter p and different n_ks, is still a binomial random variable with parameters p and $\Sigma_k n_k$.

Example 2.14 A roadway-crossing structure, such as a bridge or a box or pipe culvert, is designed to pass a flood with a return period of 50 years. In other words, the annual probability that the roadway-crossing structure would be overtopped is a 1-in-50 chance or $1/50 = 0.02$. What is the probability that the structure would be overtopped over an expected service life of 100 years?

Solution In this example, the random variable X is the number of times the roadway-crossing structure will be overtopped over a 100-year period. One can treat each year as an independent trial from which the roadway structure could be overtopped or not overtopped. Since the outcome of each "trial" is binary, the binomial distribution is applicable.

(a) $n = 10\ p = 0.1$

(b) $n = 10\ p = 0.5$

(c) $n = 10\ p = 0.9$
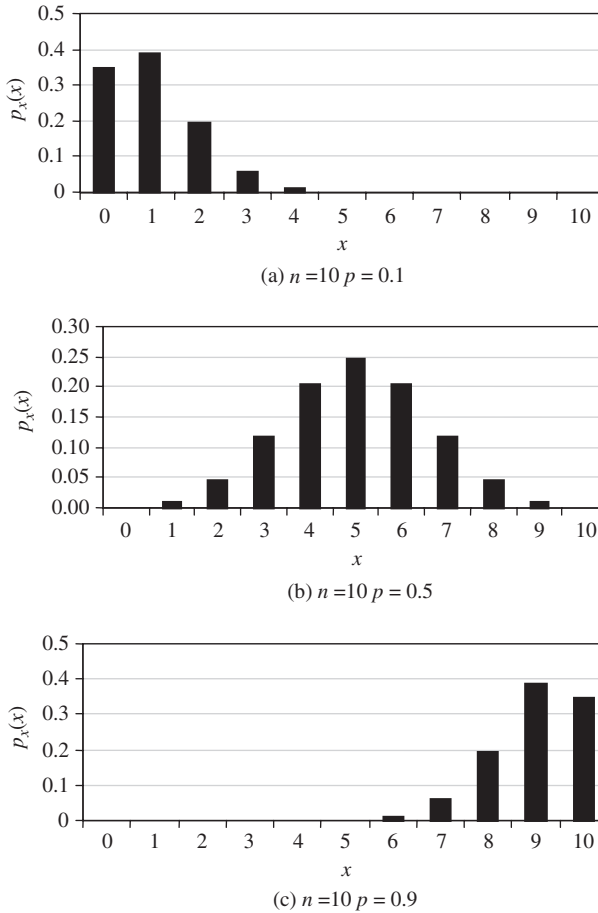
Figure 2.16 Probability mass functions of binomial random variables with different values of p.

The event of interest is the overtopping of the roadway structure. The probability of such an event occurring in each trial (namely, each year), is 0.02. A period of 100 years represents 100 trials. Hence, in the binomial distribution model, the parameters are $p = 0.02$ and $n = 100$. The probability that overtopping occurs in a period of 100 years can be calculated, according to Eq. (2.51), as

$$P(\text{overtopping occurs in an 100-year period})$$

$$= P(\text{overtopping occurs at least once in an 100-year period})$$

$$= P(X \geq 1 \mid n = 100,\ p = 0.02)$$

$$= \sum_{x=1}^{100} p_x(x) = \sum_{x=1}^{100} C_{100,x}(0.02)^x(0.98)^{100-x}$$

This equation for computing the overtopping probability requires evaluations of 100 binomial terms, which could be very cumbersome. In this case, one could solve the problem by looking at the other side of the coin, i.e., the nonoccurrence of overtopping events. In other words,

$$P(\text{overtopping occurs in a 100-year period})$$

$$= P(\text{overtopping occurs at least once in a 100-year period})$$

$$= 1 - P(\text{no overtopping occurs in a 100-year period})$$

$$= 1 - p(X = 0) = 1 - (0.98)^{100}$$

$$= 1 - 0.1326 = 0.8674$$

Calculation of the overtopping risk, as illustrated in this example, is made under an implicit assumption that the occurrence of floods is a stationary process. In other words, the flood-producing random mechanism for the watershed under consideration does not change with time. For a watershed undergoing changes in hydrologic characteristics, one should be cautious about the estimated risk.

The preceding example illustrates the basic application of the binomial distribution to reliability analysis. A commonly used alternative is the Poisson distribution described in the next section. More detailed descriptions of these two distributions in time-dependent reliability analysis of hydrosystems infrastructural engineering are given in Sec. 4.7.

2.5.2 Poisson distribution

The *Poisson distribution* has the PMF as

$$p_x(x \mid v) = \frac{e^{-v} v^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots \tag{2.53}$$

where the parameter $v > 0$ represents the mean of a Poisson random variable. Unlike the binomial random variables, Poisson random variables have no upper bound. A recursive formula for calculating the Poisson PMF is (Drane et al., 1993)

$$p_x(x \mid v) = \left(\frac{v}{x}\right) p_x(x - 1 \mid v) = R_p(x) p_x(x - 1 \mid v) \qquad \text{for } x = 1, 2, \ldots \tag{2.54}$$

with $p_x(x = 0 \mid v) = e^{-v}$ and $R_P(x) = v/x$. When $v \to \infty$ and $p \to 0$ while $np = v = $ constant, the term $R_B(x)$ in Eq. (2.52) for the binomial distribution becomes $R_P(x)$ for the Poisson distribution. Tietjen (1994) presents a simple recursive scheme for computing the Poisson cumulative probability.

For a Poisson random variable, the mean and the variance are identical to v. Plots of Poisson PMFs corresponding to different values of v are shown in Fig. 2.17. As shown in Fig. 2.15, Poisson random variables also have the same reproductive property as binomial random variables. That is, the sum of several *independent* Poisson random variables, each with a parameter v_k, is still a Poisson random variable with a parameter $v_1 + v_2 + \cdots + v_K$. The skewness
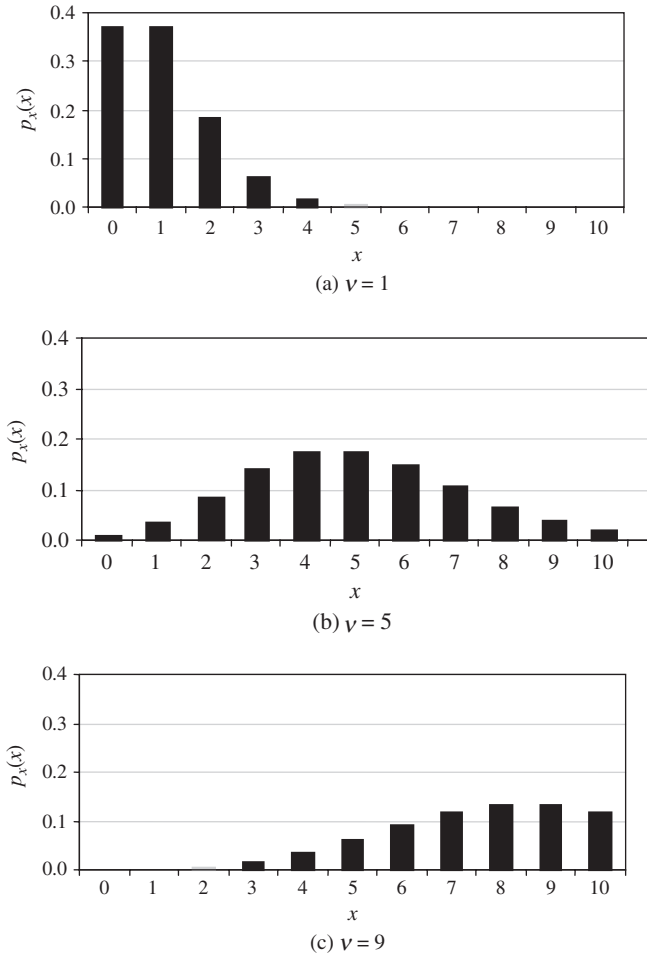
Figure 2.17   Probability mass functions of Poisson random variables with different parameter values.

coefficient of a Poisson random variable is $1/\sqrt{\nu}$, indicating that the shape of the distribution approaches symmetry as ν gets large.

The Poisson distribution has been applied widely in modeling the number of occurrences of a random event within a specified time or space interval. Equation (2.2) can be modified as

$$P_x(x \mid \lambda, t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \qquad \text{for } x = 0, 1, 2, \dots \qquad (2.55)$$

in which the parameter λ can be interpreted as the average rate of occurrence of the random event in a time interval $(0, t)$.

Example 2.15 Referring to Example 2.14, the use of binomial distribution assumes, implicitly, that the overtopping occurs, at most, once each year. The probability is zero for having more than two overtopping events annually. Relax this assumption and use the Poisson distribution to reevaluate the probability of overtopping during a 100-year period.

Solution Using the Poisson distribution, one has to determine the average number of overtopping events in a period of 100 years. For a 50-year event, the average rate of overtopping is $\lambda = 0.02$/year. Therefore, the average number of overtopping events in a period of 100 years can be obtained as $\nu = (0.02)(100) = 2$ overtoppings. The probability of overtopping in an 100-year period, using a Poisson distribution, is

$$P(\text{overtopping occurs in a 100-year period})$$

$$= P(\text{overtopping occurs at least once in a 100-year period})$$

$$= 1 - P(\text{no overtopping occurs in a 100-year period})$$

$$= 1 - p(X = 0 \mid \nu = 2) = 1 - e^{-2}$$

$$= 1 - 0.1353 = 0.8647$$

Comparing with the result from Example 2.14, use of the Poisson distribution results in a slightly smaller risk of overtopping.

To relax the restriction of equality of the mean and variance for the Poisson distribution, Consul and Jain (1973) introduced the *generalized Poisson distribution* (GPD) having two parameters θ and λ with the probability mass function as

$$p_x(x \mid \theta, \lambda) = \frac{\theta(\theta + x\lambda)^{n-1}e^{-(\theta+x\lambda)}}{x!} \qquad \text{for } x = 0, 1, 2, \ldots; \lambda \geq 0 \qquad (2.56)$$

The parameters (θ, λ) can be determined by the first two moments (Consul, 1989) as

$$E(X) = \frac{\theta}{1 - \lambda} \qquad \text{Var}(X) = \frac{\theta}{(1 - \lambda)^3} \qquad (2.57)$$

The variance of the GPD model can be greater than, equal to, or less than the mean depending on whether the second parameter λ is positive, zero, or negative. The values of the mean and variance of a GPD random variable tend to increase as θ increases. The GPD model has greater flexibility to fit various types of random counting processes, such as binomial, negative binomial, or Poisson, and many other observed data.

2.6   Some Continuous Univariate Probability Distributions

Several continuous PDFs are used frequently in reliability analysis. They include normal, lognormal, gamma, Weibull, and exponential distributions. Other distributions, such as beta and extremal distributions, also are used sometimes.

The relations among the various continuous distributions considered in this chapter and others are shown in Fig. 2.15.

2.6.1   Normal (Gaussian) distribution

The *normal distribution* is a well-known probability distribution involving two parameters: the mean and variance. A normal random variable having the mean μ_x and variance σ_x^2 is denoted herein as $X \sim N(\mu_x, \sigma_x)$ with the PDF

$$f_N\left(x \mid \mu_x, \sigma_x^2\right) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\,\exp\left[-\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2\right] \qquad \text{for } -\infty < x < \infty \quad (2.58)$$

The relationship between μ_x and σ_x and the L-moments are $\mu_x = \lambda_1$ and $\sigma_x = \sqrt{\pi}\lambda_2$.

The normal distribution is bell-shaped and symmetric with respect to the mean μ_x. Therefore, the skewness coefficient of a normal random variable is zero. Owing to the symmetry of the PDF, all odd-order central moments are zero. The kurtosis of a normal random variable is $\kappa_x = 3.0$. Referring to Fig. 2.15, a linear function of several normal random variables also is normal. That is, the linear combination of K normal random variables $W = a_1X_1 + a_2X_2 + \cdots + a_KX_K$, with $X_k \sim N(\mu_k, \sigma_k)$, for $k = 1, 2, \ldots, K$, is also a normal random variable with the mean μ_w and variance σ_w^2, respectively, as

$$\mu_w = \sum_{k=1}^{K} a_k \mu_k \qquad \sigma_w^2 = \sum_{k=1}^{K} a_k^2 \sigma_k^2 + 2\sum_{k=1}^{K-1}\sum_{k'=k+1}^{K} a_k a_{k'} \text{Cov}(X_k, X_{k'})$$

The normal distribution sometimes provides a viable alternative to approximate the probability of a nonnormal random variable. Of course, the accuracy of such an approximation depends on how closely the distribution of the nonnormal random variable resembles the normal distribution. An important theorem relating to the sum of independent random variables is the *central limit theorem*, which loosely states that the distribution of the sum of a number of independent random variables, regardless of their individual distributions, can be approximated by a normal distribution, as long as none of the variables has a dominant effect on the sum. The larger the number of random variables involved in the summation, the better is the approximation. Because many natural processes can be thought of as the summation of a large number of independent component processes, none dominating the others, the normal distribution is a reasonable approximation for these overall processes. Finally, Dowson and Wragg (1973) have shown that when only the mean and variance are specified, the maximum entropy distribution on the interval $(-\infty, +\infty)$ is the normal distribution. That is, when only the first two moments are specified, the use of the normal distribution implies more information about the nature of the underlying process specified than any other distributions.

Probability computations for normal random variables are made by first transforming the original variable to a standardized normal variable Z by

Eq. (2.49), that is,

$$Z = (X - \mu_x)/\sigma_x$$

in which Z has a mean of zero and a variance of one. Since Z is a linear function of the normal random variable X, Z is therefore normally distributed, that is, $Z \sim N(\mu_z = 0, \sigma_z = 1)$. The PDF of Z, called the *standard normal distribution*, can be obtained easily as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \qquad \text{for } -\infty < z < \infty \qquad (2.59)$$

The general expressions for the product-moments of the standard normal random variable are

$$E(Z^{2r}) = \frac{(2r)!}{2^r \times r!} \qquad \text{and} \qquad E(Z^{2r+1}) = 0 \qquad \text{for } r \geq 1 \qquad (2.60)$$

Computations of probability for $X \sim N(\mu_x, \sigma_x)$ can be made as

$$P(X \leq x) = P\left[\frac{X - \mu_x}{\sigma_x} \leq \frac{x - \mu_x}{\sigma_x}\right] = P(Z \leq z) = \Phi(z) \qquad (2.61)$$

where $z = (x - \mu_x)/\sigma_x$, and $\Phi(z)$ is the standard normal CDF defined as

$$\Phi(z) = \int_{-\infty}^{z} \phi(z)\, dz \qquad (2.62)$$

Figure 2.18 shows the shape of the PDF of the standard normal random variable.

The integral result of Eq. (2.62) is not analytically available. A table of the standard normal CDF, such as Table 2.2 or similar, can be found in many statistics textbooks (Abramowitz and Stegun, 1972; Haan, 1977; Blank, 1980;
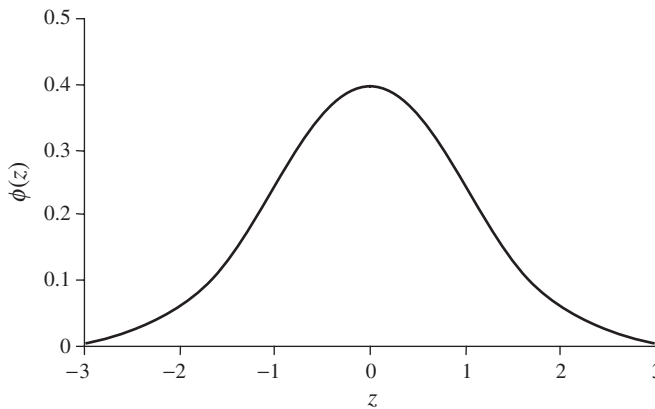
Figure 2.18 Probability density of the standard normal variable.

TABLE 2.2   Table of Standard Normal Probability, $\Phi(z) = P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

NOTE: $\Phi(-z) = 1 - \Phi(z)$, $z \geq 0$.

Devore, 1987). For numerical computation purposes, several highly accurate approximations are available for determining $\Phi(z)$. One such approximation is the polynomial approximation (Abramowitz and Stegun, 1972)

$$\Phi(z) = 1 - \phi(z)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) \qquad \text{for } z \geq 0 \qquad (2.63)$$

in which $t = 1/(1 + 0.2316419z)$, $b_1 = 0.31938153$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, and $b_5 = 1.33027443$. The maximum absolute error of the approximation is 7.5×10^{-8}, which is sufficiently accurate for most practical applications. Note that Eq. (2.63) is applicable to the non-negative-valued z. For $z < 0$, the value of standard normal CDF can be computed as $\Phi(z) = 1 - \Phi(|z|)$ by the symmetry of $\phi(z)$. Approximation equations, such as

Eq. (2.63), can be programmed easily for probability computations without needing the table of the standard normal CDF.

Equally practical is the inverse operation of finding the standard normal quantile z_p with the specified probability level p. The standard normal CDF table can be used, along with some mechanism of interpolation, to determine z_p. However, for practical algebraic computations with a computer, the following rational approximation can be used (Abramowitz and Stegun, 1972):

$$z_p = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \qquad \text{for } 0.5 < p \leq 1 \qquad (2.64)$$

in which $p = \Phi(z_p)$, $t = \sqrt{-2 \ln(1-p)}$, $c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, and $d_3 = 0.001308$. The corresponding maximum absolute error by this rational approximation is 4.5×10^{-4}. Note that Eq. (2.64) is valid for the value of $\Phi(z)$ that lies between [0.5, 1]. When $p < 0.5$, one can still use Eq. (2.64) by letting $t = \sqrt{-2 \times \ln(p)}$ and attaching a negative sign to the computed quantile value. Vedder (1995) proposed a simple approximation for computing the standard normal cumulative probabilities and standard normal quantiles.

Example 2.16 Referring to Example 2.14, determine the probability of more than five overtopping events over a 100-year period using a normal approximation.

Solution In this problem, the random variable X of interest is the number of overtopping events in a 100-year period. The exact distribution of X is binomial with parameters $n = 100$ and $p = 0.02$ or the Poisson distribution with a parameter $\nu = 2$. The exact probability of having more than five occurrences of overtopping in 100 years can be computed as

$$P(X > 5) = P(X \geq 6) = \sum_{x=6}^{100} \binom{100}{x} (0.02)^x (0.98)^{100-x}$$

$$= 1 - P(X \leq 5) = 1 - \sum_{x=6}^{5} \binom{100}{x} (0.02)^x (0.98)^{100-x}$$

$$= 1 - 0.9845 = 0.0155$$

As can be seen, there are a total of six terms to be summed up on the right-hand side. Although the computation of probability by hand is within the realm of a reasonable task, the following approximation is viable. Using a normal probability approximation, the mean and variance of X are

$$\mu_x = np = (100)(0.02) = 2.0 \qquad \sigma_x^2 = npq = (100)(0.02)(0.98) = 1.96$$

The preceding binomial probability can be approximated as

$$P(X \geq 6) \approx P(X \geq 5.5) = 1 - P(X < 5.5) = 1 - P[Z < (5.5 - 2.0)/\sqrt{1.96}]$$

$$= 1 - \Phi(2.5) = 1 - 0.9938 = 0.062$$

DeGroot (1975) showed that when $np^{1.5} > 1.07$, the error of using the normal distribution to approximate the binomial probability did not exceed 0.05. The error in the approximation gets smaller as the value of $np^{1.5}$ becomes larger. For this example, $np^{1.5} = 0.283 \leq 1.07$, and the accuracy of approximation was not satisfactory as shown.

Example 2.17 (adopted from Mays and Tung, 1992) The annual maximum flood magnitude in a river has a normal distribution with a mean of 6000 ft³/s and standard deviation of 4000 ft³/s. (a) What is the annual probability that the flood magnitude would exceed 10,000 ft³/s? (b) Determine the flood magnitude with a return period of 100 years.

Solution (a) Let Q be the random annual maximum flood magnitude. Since Q has a normal distribution with a mean $\mu_Q = 6000$ ft³/s and standard deviation $\sigma_Q = 4000$ ft³/s, the probability of the annual maximum flood magnitude exceeding 10,000 ft³/s is

$$P(Q > 10,000) = 1 - P[Z \leq (10,000 - 6000)/4000]$$
$$= 1 - \Phi(1.00) = 1 - 0.8413$$
$$= 0.1587$$

(b) A flood event with a 100-year return period represents the event the magnitude of which has, on average, an annual probability of 0.01 being exceeded. That is, $P(Q \geq q_{100}) = 0.01$, in which q_{100} is the magnitude of the 100-year flood. This part of the problem is to determine q_{100} from

$$P(Q \leq q_{100}) = 1 - P(Q \geq q_{100}) = 0.99$$

because

$$P(Q \leq q_{100}) = P\{Z \leq [(q_{100} - \mu_Q)/\sigma_Q]\}$$
$$= P[Z \leq (q_{100} - 6000)/4000]$$
$$= \Phi[q_{100} - 6000)/4000] = 0.99$$

From Table 2.2 or Eq. (2.64), one can find that $\Phi(2.33) = 0.99$. Therefore,

$$(q_{100} - 6000)/4000 = 2.33$$

which gives that the magnitude of the 100-year flood event as $q_{100} = 15,320$ ft³/s.

2.6.2 Lognormal distribution

The *lognormal distribution* is a commonly used continuous distribution for positively valued random variables. Lognormal random variables are closely related to normal random variables, by which a random variable X has a lognormal distribution if its logarithmic transform $Y = \ln(X)$ has a normal distribution with mean $\mu_{\ln x}$ and variance $\sigma_{\ln x}^2$. From the central limit theorem, if a natural process can be thought of as a multiplicative product of a large number of an independent component processes, none dominating the others, the lognormal

distribution is a reasonable approximation for these natural processes. The PDF of a lognormal random variable is

$$f_{\mathrm{LN}}\left(x \mid \mu_{\ln x}, \sigma_{\ln x}^2\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{\ln x}\,x}\,\exp\left\{-\frac{1}{2}\left[\frac{\ln(x) - \mu_{\ln x}}{\sigma_{\ln x}}\right]^2\right\} \qquad \text{for } x > 0 \quad (2.65)$$

which can be derived from the normal PDF. Statistical properties of a lognormal random variable in the original scale can be computed from those of log-transformed variables as

$$\mu_x = \lambda_1 = \exp\left(\mu_{\ln x} + \frac{\sigma_{\ln x}^2}{2}\right) \tag{2.66a}$$

$$\sigma_x^2 = \mu_x^2\left[\exp\left(\sigma_{\ln x}^2\right) - 1\right] \tag{2.66b}$$

$$\Omega_x^2 = \exp\left(\sigma_{\ln x}^2\right) - 1 \tag{2.66c}$$

$$\gamma_x = \Omega_x^3 + 3\Omega_x \tag{2.66d}$$

From Eq. (2.66d), one realizes that the shape of a lognormal PDF is always positively skewed (Fig. 2.19). Equations (2.66a) and (2.66b) can be derived easily by the moment-generating function (Tung and Yen, 2005, Sec. 4.2). Conversely, the statistical moments of $\ln(X)$ can be computed from those of X by

$$\mu_{\ln x} = \frac{1}{2}\ln\left[\frac{\mu_x^2}{1 + \Omega_x^2}\right] = \ln(\mu_x) - \frac{1}{2}\sigma_{\ln x}^2 \tag{2.67a}$$

$$\sigma_{\ln x}^2 = \ln\left(1 + \Omega_x^2\right) \tag{2.67b}$$

It is interesting to note from Eq. (2.67b) that the variance of a log-transformed variable is dimensionless.

In terms of the L-moments, the second-order L-moment for a two- and three-parameter lognormal distribution is (Stedinger et al., 1993)

$$\lambda_2 = \exp\left(\mu_{\ln x} + \frac{\sigma_{\ln x}^2}{2}\right)\mathrm{erf}\left(\frac{\sigma_{\ln x}}{2}\right) = \exp\left(\mu_{\ln x} + \frac{\sigma_{\ln x}^2}{2}\right)\left[2\Phi\left(\frac{\sigma_{\ln x}}{\sqrt{2}}\right) - 1\right]$$

$$\tag{2.68}$$

in which $\mathrm{erf}(\cdot)$ is an error function the definitional relationship of which, with $\Phi(z)$ is

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,du = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2/2}\,dz = 2\Phi(\sqrt{2}x) - 1 \tag{2.69}$$

Hence the L-coefficient of variation is $\tau_2 = 2\Phi(\sigma_{\ln x}/\sqrt{2}) - 1$. The relationship between the third- and fourth-order L-moment ratios can be approximated by the following polynomial function with accuracy within 5×10^{-4} for $|\tau_2| < 0.9$ (Hosking, 1991):

$$\tau_4 = 0.12282 + 0.77518\tau_3^2 + 0.12279\tau_3^4 - 0.13638\tau_3^6 + 0.11386\tau_3^8 \tag{2.70}$$
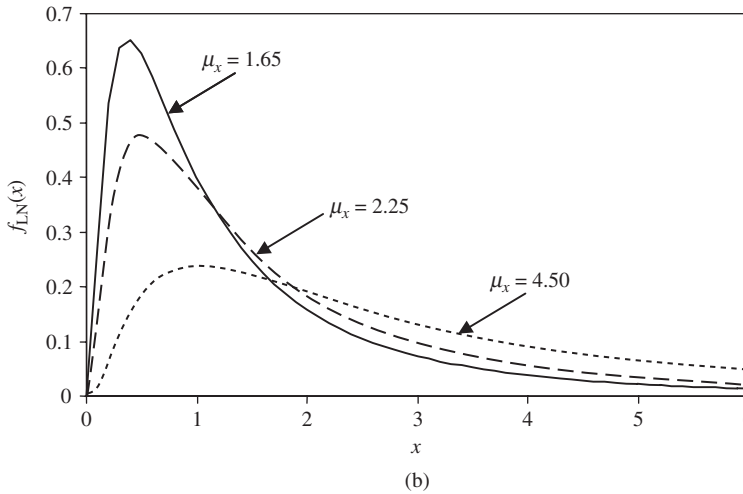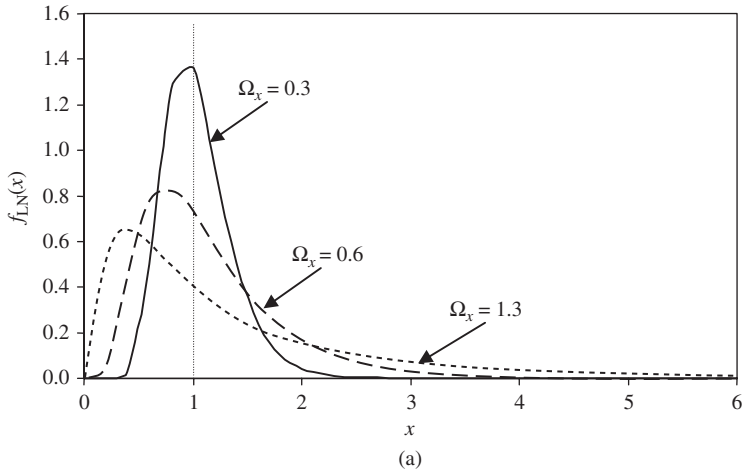
Figure 2.19 Shapes of lognormal probability density functions: (*a*) $\mu_x = 1.0$; (*b*) $\Omega_x = 1.30$.

Since the sum of normal random variables is normally distributed, the product of lognormal random variables also is lognormally distributed (see Fig. 2.15). This useful reproductive property of lognormal random variables can be stated as if $X_1, X_2, \ldots, X_K$ are independent lognormal random variables, then $W = b_0 \Pi_{k=1}^{K} X_k^{b_k}$ has a lognormal distribution with mean and variance as

$$\mu_{\ln w} = \ln(b_0) + \sum_{k=1}^{K} b_k \mu_{\ln x_k} \qquad \sigma_{\ln w}^2 = \sum_{k=1}^{K} b_k^2 \sigma_{\ln x_k}^2$$

In the case that two lognormal random variables are correlated with a correlation coefficient $\rho_{x,y}$ in the original scale, then the covariance terms in the

log-transformed space must be included in calculating $\sigma_{\ln w}^2$. Given $\rho_{x,y}$, the correlation coefficient in the log-transformed space can be computed as

$$\text{Corr}(\ln X, \ln Y) = \rho_{\ln x, \ln y} = \frac{\ln(1 + \rho_{x,y}\Omega_x\Omega_y)}{\sqrt{\ln\left(1 + \Omega_x^2\right) \times \ln\left(1 + \Omega_y^2\right)}} \qquad (2.71)$$

Derivation of Eq. (2.71) can be found in Tung and Yen (2005).

Example 2.18 Re-solve Example 2.17 by assuming that the annual maximum flood magnitude in the river follows a lognormal distribution.

Solution (a) Since Q has a lognormal distribution, $\ln(Q)$ is normally distributed with a mean and variance that can be computed from Eqs. (2.67a) and (2.67b), respectively, as

$$\Omega_Q = 4000/6000 = 0.667$$
$$\sigma_{\ln Q}^2 = \ln(1 + 0.667^2) = 0.368$$
$$\mu_{\ln Q} = \ln(6000) - 0.368/2 = 8.515$$

The probability of the annual maximum flood magnitude exceeding 10,000 ft^3/s is

$$P(Q > 10,000) = P[\ln Q > \ln(10,000)]$$
$$= 1 - P[Z \leq (9.210 - 8.515)/\sqrt{0.368}]$$
$$= 1 - \Phi(1.146) = 1 - 0.8741 = 0.1259$$

(b) A 100-year flood q_{100} represents the event the magnitude of which corresponds to $P(Q \geq q_{100}) = 0.01$, which can be determined from

$$P(Q \leq q_{100}) = 1 - P(Q \geq q_{100}) = 0.99$$

because
$$P(Q \leq q_{100}) = P[\ln Q \leq \ln(q_{100})]$$
$$= P\{Z \leq [\ln(q_{100}) - \mu_{\ln Q}]/\sigma_{\ln Q}\}$$
$$= P\{Z \leq [\ln(q_{100}) - 8.515]/\sqrt{0.368}\}$$
$$= \Phi\{[\ln(q_{100}) - 8.515]/\sqrt{0.368}\} = 0.99$$

From Table 2.2 or Eq. (2.64), one can find that $\Phi(2.33) = 0.99$. Therefore,

$$[\ln(q_{100}) - 8.515]/\sqrt{0.368} = 2.33$$

which yields $\ln(q_{100}) = 9.9284$. The magnitude of the 100-year flood event then is $q_{100} = \exp(9.9284) = 20,500$ ft^3/s.

2.6.3 Gamma distribution and variations

The *gamma distribution* is a versatile continuous distribution associated with a positive-valued random variable. The *two-parameter gamma distribution* has

a PDF defined as

$$f_G(x \mid \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} (x/\beta)^{\alpha-1} e^{x/\beta} \qquad \text{for } x > 0 \tag{2.72}$$

in which $\beta > 0$ and $\alpha > 0$ are the parameters and $\Gamma(\bullet)$ is a *gamma function* defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt \tag{2.73}$$

The mean, variance, and skewness coefficient of a gamma random variable having a PDF as Eq. (2.72) are

$$\mu_x = \lambda_1 = \alpha\beta \qquad \sigma_x^2 = \alpha\beta^2 \qquad \gamma_x = 2/\sqrt{\alpha} \tag{2.74}$$

In terms of L-moments, the second-order L-moment is

$$\lambda_2 = \frac{\beta \Gamma(\alpha + 0.5)}{\sqrt{\pi} \Gamma(\alpha)} \tag{2.75}$$

and the relationship between the third- and fourth-order L-moment ratios can be approximated as (Hosking, 1991)

$$\tau_4 = 0.1224 + 0.30115\tau_3^2 + 0.95812\tau_3^4 - 0.57488\tau_3^6 + 0.19383\tau_3^8 \tag{2.76}$$

In the case that the lower bound of a gamma random variable is a positive quantity, the preceding two-parameter gamma PDF can be modified to a *three-parameter gamma PDF* as

$$f_G(x \mid \xi, \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} \left[\frac{x - \xi}{\beta} \right]^{\alpha-1} e^{-(x-\xi)/\beta} \qquad \text{for } x > \xi \tag{2.77}$$

where ξ is the lower bound. The two-parameter gamma distribution can be reduced to a simpler form by letting $Y = X/\beta$, and the resulting *one-parameter gamma PDF* (called the *standard gamma distribution*) is

$$f_G(y \mid \alpha) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^y \qquad \text{for } y > 0 \tag{2.78}$$

Tables of the cumulative probability of the standard gamma distribution can be found in Dudewicz (1976). Shapes of some gamma distributions are shown in Fig. 2.20 to illustrate its versatility. If α is a positive integer in Eq. (2.78), the distribution is called an *Erlang distribution*.

When $\alpha = 1$, the two-parameter gamma distribution reduces to an *exponential distribution* with the PDF

$$f_{EXP}(x \mid \beta) = e^{-x/\beta}/\beta \qquad \text{for } x > 0 \tag{2.79}$$

An exponential random variable with a PDF as Eq. (2.79) has the mean and standard deviation equal to β (see Example 2.8). Therefore, the coefficient of
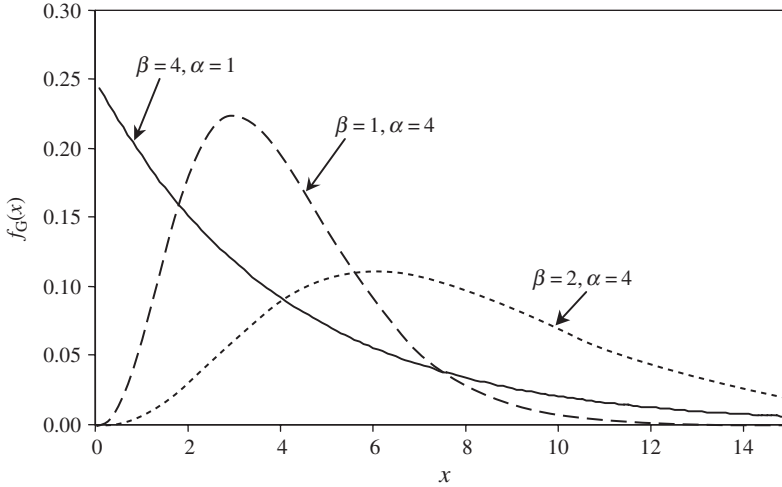
Figure 2.20 Shapes of gamma probability density functions.

variation of an exponential random variable is equal to unity. The exponential distribution is used commonly for describing the life span of various electronic and mechanical components. It plays an important role in reliability mathematics using time-to-failure analysis (see Chap. 5).

Two variations of the gamma distribution are used frequently in hydrologic frequency analysis, namely, the Pearson and log-Pearson type 3 distributions. In particular, the log-Pearson type 3 distribution is recommended for use by the U.S. Water Resources Council (1982) as the standard distribution for flood frequency analysis. A *Pearson type 3* random variable has the PDF

$$f_{P3}(x \mid \xi, \alpha, \beta) = \frac{1}{|\beta|\Gamma(\alpha)} \left(\frac{x - \xi}{\beta} \right)^{\alpha-1} e^{-(x-\xi)/\beta} \tag{2.80}$$

with $\alpha > 0$, $x \geq \xi$ when $\beta > 0$ and with $\alpha > 0$, $x \leq \xi$ when $\beta < 0$. When $\beta > 0$, the Pearson type 3 distribution is identical to the three-parameter gamma distribution. However, the Pearson type 3 distribution has the flexibility to model negatively skewed random variables corresponding to $\beta < 0$. Therefore, the skewness coefficient of the Pearson type 3 distribution can be computed, from modifying Eq. (2.74), as $\text{sign}(\beta)2/\sqrt{\alpha}$.

Similar to the normal and lognormal relationships, the PDF of a *log-Pearson type 3* random variable is

$$f_{LP3}(x \mid \xi, \alpha, \beta) = \frac{1}{x|\beta|\Gamma(\alpha)} \left[\frac{\ln(x) - \xi}{\beta} \right]^{\alpha-1} e^{-[\ln(x)-\xi]/\beta} \tag{2.81}$$

with $\alpha > 0$, $x \geq e^{\xi}$ when $\beta > 0$ and with $\alpha > 0$, $x \leq e^{\xi}$ when $\beta < 0$. Numerous studies can be found in the literature about Pearson type 3 and log-Pearson

type 3 distributions. Kite (1977), Stedinger et al. (1993), and Rao and Hamed (2000) provide good summaries of these two distributions.

Evaluation of the probability of gamma random variables involves computations of the gamma function, which can be made by using the following recursive formula:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \qquad (2.82)$$

When the argument α is an integer number, then $\Gamma(\alpha) = (\alpha - 1)! = (\alpha - 1)(\alpha - 2) \cdots 1$. However, when α is a real number, the recursive relation would lead to $\Gamma(\alpha')$ as the smallest term, with $1 < \alpha' < 2$. The value of $\Gamma(\alpha')$ can be determined by a table of the gamma function or by numerical integration on Eq. (2.73). Alternatively, the following approximation could be applied to accurately estimate the value of $\Gamma(\alpha')$ (Abramowitz and Stegun, 1972):

$$\Gamma(\alpha') = \Gamma(x + 1) = 1 + \sum_{i=1}^{5} a_i x^i \qquad \text{for } 0 < x < 1 \qquad (2.83)$$

in which $a_1 = -0.577191652$, $a_2 = 0.988205891$, $a_3 = -0.897056937$, $a_4 = 0.4245549$, and $a_5 = -0.1010678$. The maximum absolute error associated with Eq. (2.83) is 5×10^{-5}.

2.6.4 Extreme-value distributions

Hydrosystems engineering reliability analysis often focuses on the statistical characteristics of extreme events. For example, the design of flood-control structures may be concerned with the distribution of the largest events over the recorded period. On the other hand, the establishment of a drought-management plan or water-quality management scheme might be interested in the statistical properties of minimum flow over a specified period. Statistics of extremes are concerned with the statistical characteristics of $X_{\max,n} = \max\{X_1, X_2, \ldots, X_n\}$ and/or $X_{\min,n} = \min\{X_1, X_2, \ldots, X_n\}$ in which $X_1, X_2, \ldots, X_n$ are observations of random processes. In fact, the exact distributions of extremes are functions of the underlying (or parent) distribution that generates the random observations $X_1, X_2, \ldots, X_n$ and the number of observations. Of practical interest are the asymptotic distributions of extremes. *Asymptotic distribution* means that the resulting distribution is the limiting form of $F_{\max,n}(y)$ or $F_{\min,n}(y)$ as the number of observations n approaches infinity. The asymptotic distributions of extremes turn out to be independent of the sample size n and the underlying distribution for random observations. That is,

$$\lim_{n \to \infty} F_{\max,n}(y) = F_{\max}(y) \qquad \lim_{n \to \infty} F_{\min,n}(y) = F_{\min}(y)$$

Furthermore, these asymptotic distributions of the extremes largely depend on the tail behavior of the parent distribution in either direction toward the extremes. The center portion of the parent distribution has little significance for defining the asymptotic distributions of extremes. The work on statistics of extremes was pioneered by Fisher and Tippett (1928) and later was extended

by Gnedenko (1943). Gumbel (1958), who dealt with various useful applications of $X_{\max,n}$ and $X_{\min,n}$ and other related issues.

Three types of asymptotic distributions of extremes are derived based on the different characteristics of the underlying distribution (Haan, 1977):

Type I. Parent distributions are unbounded in the direction of extremes, and all statistical moments exist. Examples of this type of parent distribution are normal (for both largest and smallest extremes), lognormal, and gamma distributions (for the largest extreme).

Type II. Parent distributions are unbounded in the direction of extremes, but all moments do not exist. One such distribution is the Cauchy distribution (Sec. 2.6.5). Thus the type II extremal distribution has few applications in practical engineering analysis.

Type III. Parent distributions are bounded in the direction of the desired extreme. Examples of this type of underlying distribution are the beta distribution (for both largest and smallest extremes) and the lognormal and gamma distributions (for the smallest extreme).

Owing to the fact that $X_{\min,n} = -\max\{-X_1, -X_2, \ldots, -X_n\}$, the asymptotic distribution functions of $X_{\max,n}$ and $X_{\min,n}$ satisfy the following relation (Leadbetter et al., 1983):

$$F_{\min}(y) = 1 - F_{\max}(-y) \tag{2.84}$$

Consequently, the asymptotic distribution of $X_{\min}$ can be obtained directly from that of $X_{\max}$. Three types of asymptotic distributions of the extremes are listed in Table 2.3.

Extreme-value type I distribution. This is sometimes referred to as the *Gumbel distribution, Fisher-Tippett distribution*, and *double exponential distribution*. The CDF and PDF of the extreme-value type I (EV1) distribution have, respectively, the following forms:

$$F_{\text{EV1}}(x \mid \xi, \beta) = \exp\left\{-\exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right\} \qquad \text{for maxima}$$
$$= 1 - \exp\left\{-\exp\left[+\left(\frac{x-\xi}{\beta}\right)\right]\right\} \qquad \text{for minima} \tag{2.85a}$$

TABLE 2.3 Three Types of Asymptotic Cumulative Distribution Functions (CDFs) of Extremes

Type	Maxima	Range	Minima	Range
I	$\exp(-e^{-y})$	$-\infty < y < \infty$	$1 - \exp(-e^{y})$	$-\infty < y < \infty$
II	$\exp(-y^{\alpha})$	$\alpha < 0, y > 0$	$1 - \exp[-(-y)^{\alpha}]$	$\alpha < 0, y < 0$
III	$\exp[-(-y)^{\alpha}]$	$\alpha > 0, y < 0$	$1 - \exp(-y^{\alpha})$	$\alpha > 0, y > 0$

NOTE : $y = (x - \xi)/\beta$.
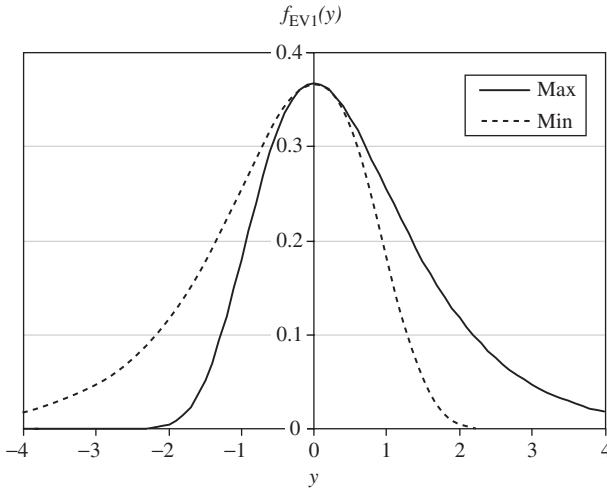
Figure 2.21 Probability density function of extreme-value type I random variables.

$$f_{EV1}(x \mid \xi, \beta) = \frac{1}{\beta} \exp\left\{-\left(\frac{x-\xi}{\beta}\right) - \exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right\} \qquad \text{for maxima}$$

$$= \frac{1}{\beta} \exp\left\{+\left(\frac{x-\xi}{\beta}\right) - \exp\left[+\left(\frac{x-\xi}{\beta}\right)\right]\right\} \qquad \text{for minima}$$

(2.85b)

for $-\infty < x, \xi < \infty$, and $\beta \geq 0$. The shapes of the EV1 distribution are shown in Fig. 2.21, in which transformed random variable $Y = (X - \xi)/\beta$ is used. As can be seen, the PDF associated with the largest extreme is a mirror image of the smallest extreme with respect to the vertical line passing through the common mode, which happens to be the parameter ξ. The first three product-moments of an EV1 random variable are

$$\mu_x = \lambda_1 = \xi + 0.5772\beta \qquad \text{for the largest extreme}$$

$$= \xi - 0.5772\beta \qquad \text{for the smallest extreme} \qquad (2.86a)$$

$$\sigma_x^2 = 1.645\beta^2 \qquad \text{for both types} \qquad (2.86b)$$

$$\gamma_x = 1.13955 \qquad \text{for the largest extreme}$$

$$= -1.13955 \qquad \text{for the smallest extreme} \qquad (2.86c)$$

The second- to fourth-order L-moments of the EV1 distribution for maxima are

$$\lambda_2 = \beta \ln(2) \qquad \tau_3 = 0.1699 \qquad \tau_4 = 0.1504 \qquad (2.87)$$

Using the transformed variable $Y = (X - \xi)/\beta$, the CDFs of the EV1 for the maxima and minima are shown in Table 2.3. Shen and Bryson (1979) showed

that if a random variable had an EV1 distribution, the following relationship is satisfied when ξ is small:

$$x_{T_1} \approx \left[\frac{\ln(T_1)}{\ln(T_2)}\right] x_{T_2} \tag{2.88}$$

where x_T is the quantile corresponding to the exceedance probability of $1/T$.

Example 2.19   Repeat Example 2.17 by assuming that the annual maximum flood follows the EV1 distribution.

Solution   Based on the values of a mean of 6000 ft^3/s and standard deviation of 4000 ft^3/s, the values of distributional parameters ξ and β can be determined as follows. For maxima, β is computed from Eq. (2.86b) as

$$\beta = \frac{\sigma_Q}{\sqrt{1.645}} = \frac{4000}{1.2826} = 3118.72 \text{ ft}^3/\text{s}$$

and from Eq. (2.86a), one has

$$\xi = \mu_Q - 0.577\beta = 6000 - 0.577(3118.72) = 4200.50 \text{ ft}^3/\text{s}$$

(a) The probability of exceeding 10,000 ft^3/s, according to Eq. (2.85a), is

$$P(Q > 10,000) = 1 - F_{\text{EV1}}(10,000)$$

$$= 1 - \exp\left[-\exp\left(-\frac{10,000 - 4200.50}{3118.72}\right)\right]$$

$$= 1 - \exp[-\exp(-1.860)]$$

$$= 1 - 0.8558 = 0.1442$$

(b) On the other hand, the magnitude of the 100-year flood event can be calculated as

$$y_{100} = \frac{q_{100} - \xi}{\beta} = -\ln[-\ln(1 - 0.01)] = 4.60$$

Hence $q_{100} = 4200.50 + 4.60(3118.7) = 18,550 \text{ ft}^3/\text{s}$.

Extreme-value type III distribution.   For the extreme-value type III (EV3) distribution, the corresponding parent distributions are bounded in the direction of the desired extreme (see Table 2.3). For many hydrologic and hydraulic random variables, the lower bound is zero, and the upper bound is infinity. For this reason, the EV3 distribution for the maxima has limited applications. On the other hand, the EV3 distribution of the minima is used widely for modeling the smallest extremes, such as drought or low-flow condition. The EV3 distribution for the minima is also known as the *Weibull distribution*, having a PDF defined as

$$f_W(x \mid \xi, \alpha, \beta) = \frac{\alpha}{\beta}\left(\frac{x - \xi}{\beta}\right)^{\alpha - 1} \exp\left[-\left(\frac{x - \xi}{\beta}\right)^{\alpha}\right] \qquad \text{for } x \geq \xi \text{ and } \alpha, \beta > 0 \tag{2.89}$$

When $\xi = 0$ and $\alpha = 1$, the Weibull distribution reduces to the exponential distribution. Figure 2.22 shows that the versatility of the Weibull distribution function depends on the parameter values. The CDF of Weibull random variables can be derived as

$$F_W(x \mid \xi, \alpha, \beta) = 1 - \exp\left[-\left(\frac{x - \xi}{\beta}\right)^{\alpha}\right] \tag{2.90}$$

The mean and variance of a Weibull random variable can be derived as

$$\mu_x = \lambda_1 = \xi + \beta\Gamma\left(1 + \frac{1}{\beta}\right) \tag{2.91a}$$

$$\sigma_x^2 = \beta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right] \tag{2.91b}$$

and the second-order L-moment is

$$\lambda_2 = \beta(1 - 2^{-1/\alpha})\,\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{2.92}$$

Generalized extreme-value distribution. The *generalized extreme-value (GEV) distribution* provides an expression that encompasses all three types of extreme-value distributions. The CDF of a random variable corresponding to the maximum with a GEV distribution is

$$F_{\text{GEV}}(x \mid \xi, \alpha, \beta) = \exp\left\{-\left[1 - \frac{\alpha(x - \xi)}{\beta}\right]^{1/\alpha}\right\} \qquad \text{for } \alpha \neq 0 \tag{2.93}$$
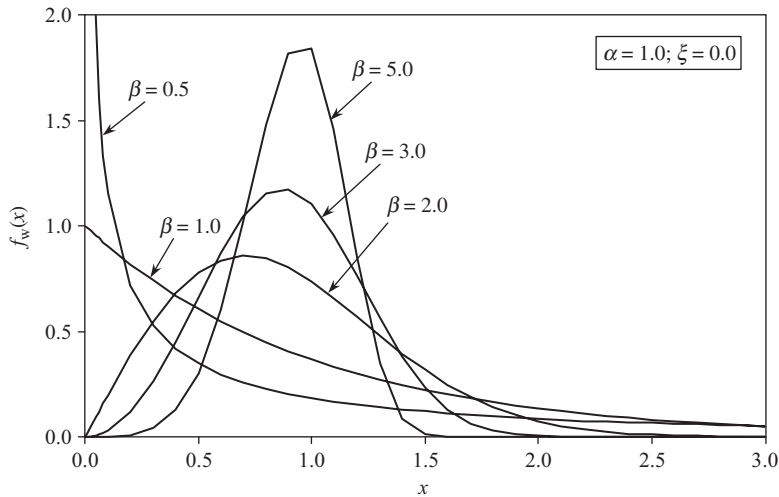
Figure 2.22 Probability density functions of a Weibull random variable.

When $\alpha = 0$, Eq. (2.93) reduces to Eq. (2.85a) for the Gumbel distribution. For $\alpha < 0$, it corresponds to the EV2 distribution having a lower bound $x > \xi + \beta/\alpha$, whereas, on the other hand, for $\alpha > 0$, it corresponds to the EV3 distribution having an upper bound $x < \xi + \beta/\alpha$. For $|\alpha| < 0.3$, the shape of the GEV distribution is similar to the Gumbel distribution, except that the right-hand tail is thicker for $\alpha < 0$ and thinner for $\alpha > 0$ (Stedinger et al., 1993).

The first three moments of the GEV distribution, respectively, are

$$\mu_x = \lambda_1 = \xi + \left(\frac{\beta}{\alpha}\right)[1 - \Gamma(1 + \alpha)] \tag{2.94a}$$

$$\sigma_x^2 = \left(\frac{\beta}{\alpha}\right)^2 [\Gamma(1 + 2\alpha) - \Gamma^2(1 + \alpha)] \tag{2.94b}$$

$$\gamma_x = \text{sign}(\alpha)\frac{-\Gamma(1 + 3\alpha) + 3\Gamma(1 + 2\alpha)\Gamma(1 + \alpha) - 2\Gamma^3(1 + \alpha)}{[\Gamma(1 + 2\alpha) - \Gamma^2(1 + \alpha)]^{1.5}} \tag{2.94c}$$

where $\text{sign}(\alpha)$ is $+1$ or -1 depending on the sign of α. From Eqs. (2.94b) and (2.94c) one realizes that the variance of the GEV distribution exists when $\alpha > -0.5$, and the skewness coefficient exists when $\alpha > -0.33$. The GEV distribution recently has been used frequently in modeling the random mechanism of hydrologic extremes, such as precipitation and floods.

The relationships between the L-moments and GEV model parameters are

$$\lambda_2 = \frac{\beta}{\alpha}(1 - 2^{-\alpha})\Gamma(1 + \alpha) \tag{2.95a}$$

$$\tau_3 = \frac{2(1 - 3^{-\alpha})}{(1 - 2^{-\alpha})} - 3 \tag{2.95b}$$

$$\tau_4 = \frac{1 - 5(4^{-\alpha}) + 10(3^{-\alpha}) - 6(2^{-\alpha})}{1 - 2^{-\alpha}} \tag{2.95c}$$

2.6.5 Beta distributions

The *beta distribution* is used for describing random variables having both lower and upper bounds. Random variables in hydrosystems that are bounded on both limits include reservoir storage and groundwater table for unconfined aquifers. The *nonstandard beta* PDF is

$$f_{\text{NB}}(x \mid a, b, \alpha, \beta) = \frac{1}{B(\alpha, \beta)(b - a)^{\alpha + \beta - 1}}(x - a)^{\alpha - 1}(b - x)^{\beta - 1} \qquad \text{for } a \le x \le b \tag{2.96}$$

in which a and b are the lower and upper bounds of the beta random variable, respectively; $\alpha > 0$, $\beta > 0$; and $B(\alpha, \beta)$ is a *beta function* defined as

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \tag{2.97}$$

Using the new variable $Y = (X - a)/(b - a)$, the nonstandard beta PDF can be reduced to the *standard beta PDF* as

$$f_B(y \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1 - y)^{\beta-1} \qquad \text{for } 0 < y < 1 \qquad (2.98)$$

The beta distribution is also a very versatile distribution that can have many shapes, as shown in Fig. 2.23. The mean and variance of the standard beta random variable Y, respectively, are

$$\mu_y = \frac{\alpha}{\alpha + \beta} \qquad \sigma_y^2 = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \qquad (2.99)$$

When $\alpha = \beta = 1$, the beta distribution reduces to a *uniform distribution* as

$$f_U(x) = \frac{1}{b - a} \qquad \text{for } a \leq x \leq b \qquad (2.100)$$

2.6.6 Distributions related to normal random variables

The normal distribution has been playing an important role in the development of statistical theories. This subsection briefly describes two distributions related to the functions of normal random variables.
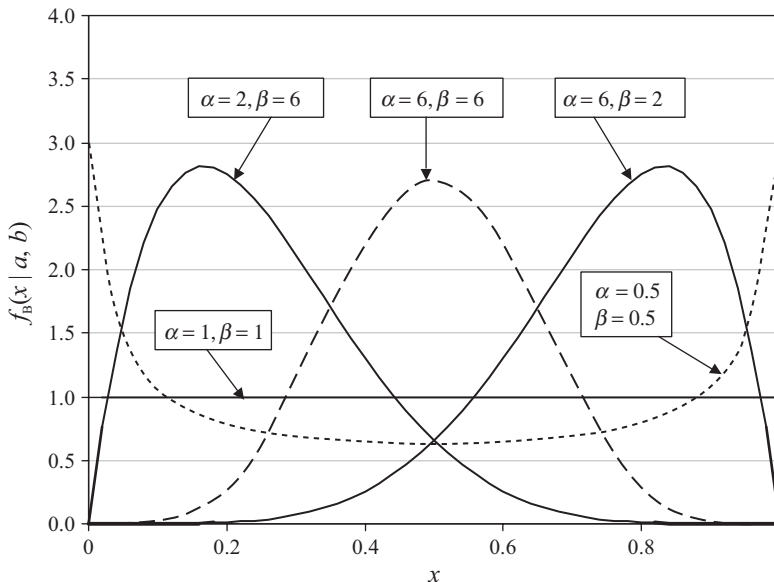
Figure 2.23 Shapes of standard beta probability density functions. (*After Johnson and Kotz, 1972.*)

χ^2 (chi-square) distribution.   The sum of the squares of K independent standard normal random variables results in a χ^2 (chi-square) random variable with K degrees of freedom, denoted as χ_K^2. In other words,

$$\sum_{k=1}^{K} Z_k^2 \sim \chi_K^2 \tag{2.101}$$

in which the Z_ks are independent standard normal random variables. The PDF of a χ^2 random variable with K degrees of freedom is

$$f_{\chi^2}(x \mid K) = \frac{1}{2^{K/2}\Gamma(K/2)}x^{(K/2-1)}e^{-x/2} \qquad \text{for } x > 0 \tag{2.102}$$

Comparing Eq. (2.102) with Eq. (2.72), one realizes that the χ^2 distribution is a special case of the two-parameter gamma distribution with $\alpha = K/2$ and $\beta = 2$. The mean, variance, and skewness coefficient of a χ_K^2 random variable, respectively, are

$$\mu_x = K \qquad \sigma_x^2 = 2K \qquad \gamma_x = 2/\sqrt{K/2}$$

Thus, as the value of K increases, the χ^2 distribution approaches a symmetric distribution. Figure 2.24 shows a few χ^2 distributions with various degrees of freedom. If $X_1, X_2, \ldots, X_K$ are independent normal random variables with the common mean μ_x and variance σ_x^2, the χ^2 distribution is related to the sample of normal random variables as follows:

1. The sum of K squared standardized normal variables $Z_k = (X_k - \overline{X})/\sigma_x$, $k = 1, 2, \ldots, K$, has a χ^2 distribution with $(K - 1)$ degrees of freedom.

2. The quantity $(K - 1)S^2/\sigma_x^2$ has a χ^2 distribution with $(K - 1)$ degrees of freedom in which S^2 is the unbiased sample variance computed according to Table 2.1.
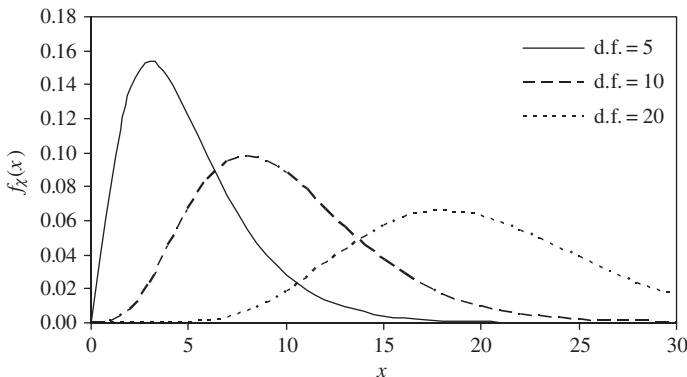
Figure 2.24 Shapes of chi-square probability density functions where d.f. refers to the degrees of freedom.

t-distribution. A random variable having a t-distribution results from the ratio of the standard normal random variable to the square root of the χ^2 random variable divided by its degrees of freedom, that is,

$$T_K = \frac{Z}{\sqrt{\chi_K^2/K}} \tag{2.103}$$

in which T_K is a t-distributed random variable with K degrees of freedom. The PDF of T_K can be expressed as

$$f_{\mathrm{T}}(x\mid K) = \frac{\Gamma[(K+1)/2]}{\sqrt{\pi K}\,\Gamma(K/2)}\left(1 + \frac{x^2}{K}\right)^{-(K+1)/2} \qquad \text{for } -\infty < x < \infty \tag{2.104}$$

A t-distribution is symmetric with respect to the mean $\mu_x = 0$ when $K \geq 1$. Its shape is similar to the standard normal distribution, except that the tails of the PDF are thicker than $\phi(z)$. However, as $K \to \infty$, the PDF of a t-distributed random variable approaches the standard normal distribution. Figure 2.25 shows some PDFs for t-random variables of different degrees of freedom. It should be noted that when $K = 1$, the t-distribution reduces to the *Cauchy distribution*, for which all product-moments do not exist. The mean and variance of a t-distributed random variable with K degrees of freedom are

$$\mu_x = 0 \qquad \sigma_x^2 = K/(K-2) \qquad \text{for } K \geq 3$$

When the population variance of normal random variables is known, the sample mean $\overline{X}$ of K normal random samples from $N(\mu_x, \sigma_x^2)$ has a normal distribution with mean μ_x and variance σ_x^2/K. However, when the population variance is unknown but is estimated by S^2 according to Table 2.1, then the quantity $\sqrt{K}(\overline{X} - \mu_x)/S$, which is the standardized sample mean using the sample variance, has a t-distribution with $(K-1)$ degrees of freedom.
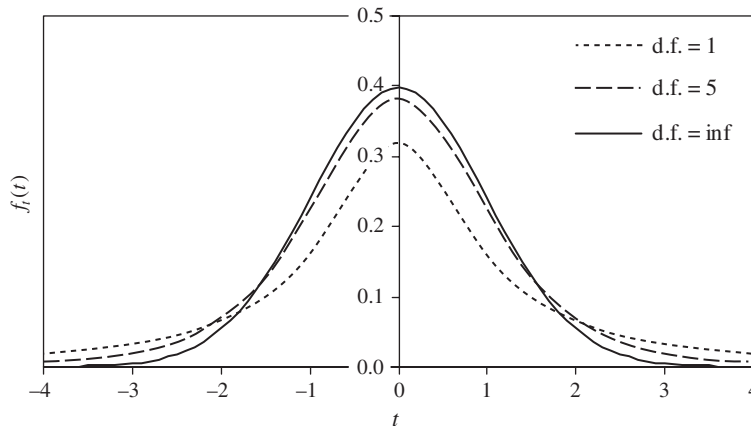
Figure 2.25 Shapes of t-distributions where d.f. refers to degrees of freedom.

2.7   Multivariate Probability Distributions

Multivariate probability distributions are extensions of univariate probability distributions that jointly account for more than one random variable. *Bivariate* and *trivariate* distributions are special cases where two and three random variables, respectively, are involved. The fundamental basis of multivariate probability distributions is described in Sec. 2.3.2. In general, the availability of multivariate distribution models is significantly less than that for univariate cases. Owing to their frequent use in multivariate modeling and reliability analysis, two multivariate distributions, namely, multivariate normal and multivariate lognormal, are presented in this section. Treatments of some multivariate nonnormal random variables are described in Secs. 4.5 and 7.5. For other types of multivariate distributions, readers are referred to Johnson and Kotz (1976) and Johnson (1987).

Several ways can be used to construct a multivariate distribution (Johnson and Kotz, 1976; Hutchinson and Lai, 1990). Based on the joint distribution discussed in Sec. 2.2.2, the straightforward way of deriving a joint PDF involving K multivariate random variables is to extend Eq. (2.19) as

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = f_1(x_1) \times f_2(x_2 \,|\, x_1) \times \cdots \times f_K(x_1, x_2, \ldots, x_{K-1}) \qquad (2.105)$$

in which $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)^t$ is a vector containing variates of K random variables with the superscript t indicating the transpose of a matrix or vector. Applying Eq. (2.105) requires knowledge of the conditional PDFs of the random variables, which may not be easily obtainable.

One simple way of constructing a joint PDF of two random variables is by mixing. Morgenstern (1956) suggested that the joint CDF of two random variables could be formulated, according to their respective marginal CDFs, as

$$F_{1,2}(x_1, x_2) = F_1(x_1) F_2(x_2)\{1 + \theta[1 - F_1(x_1)][1 - F_2(x_2)]\} \qquad \text{for } -1 \le \theta \le 1$$
$$(2.106)$$

in which $F_k(x_k)$ is the marginal CDF of the random variable X_k, and θ is a weighting constant. When the two random variables are independent, the weighting constant $\theta = 0$. Furthermore, the sign of θ indicates the positiveness or negativeness of the correlation between the two random variables. This equation was later extended by Farlie (1960) to

$$F_{1,2}(x_1, x_2) = F_1(x_1) F_2(x_2)[1 + \theta f_1(x_1) f_2(x_2)] \qquad \text{for } -1 \le \theta \le 1 \qquad (2.107)$$

in which $f_k(x_k)$ is the marginal PDF of the random variable X_k. Once the joint CDF is obtained, the joint PDF can be derived according to Eq. (2.15a).

Constructing a bivariate PDF by the mixing technique is simple because it only requires knowledge about the marginal distributions of the involved random variables. However, it should be pointed out that the joint distribution obtained from Eq. (2.106) or Eq. (2.107) does not necessarily cover the entire range of the correlation coefficient $[-1, \ 1]$ for the two random variables

under consideration. This is illustrated in Example 2.20. Liu and Der Kiureghian (1986) derived the range of the valid correlation coefficient value for the bivariate distribution by mixing, according to Eq. (2.106), from various combinations of marginal PDFs, and the results are shown in Table 2.4.

Nataf (1962), Mardia (1970a, 1970b), and Vale and Maurelli (1983) proposed other ways to construct a bivariate distribution for any pair of random variables. This was done by finding the transforms $Z_k = t(X_k)$, for $k = 1, 2$, such that Z_1 and Z_2 are standard normal random variables. Then a bivariate normal distribution is ascribed to Z_1 and Z_2. One such transformation is $z_k = \Phi^{-1}[F_k(x_k)]$, for $k = 1, 2$. A detailed description of such a normal transformation is given in Sec. 4.5.3.

Example 2.20   Consider two correlated random variables X and Y, each of which has a marginal PDF of an exponential distribution type as

$$f_x(x) = e^{-x} \quad \text{for } x \geq 0 \qquad f_y(y) = e^{-y} \quad \text{for } y \geq 0$$

To derive a joint distribution for X and Y, one could apply the Morgenstern formula. The marginal CDFs of X and Y can be obtained easily as

$$F_x(x) = 1 - e^{-x} \quad \text{for } x \geq 0 \qquad F_y(y) = 1 - e^{-y} \quad \text{for } y \geq 0$$

According to Eq. (2.106), the joint CDF of X and Y can be expressed as

$$F_{x,y}(x, y) = (1 - e^{-x})(1 - e^{-y})(1 + \theta e^{-x-y}) \qquad \text{for } x, y \geq 0$$

Then the joint PDF of X and Y can be obtained, according to Eq. (2.7a), as

$$f_{x,y}(x, y) = e^{-x-y}[1 + \theta(2e^{-x} - 1)(2e^{-y} - 1)] \qquad \text{for } x, y \geq 0$$

TABLE 2.4   Valid Range of Correlation Coefficients for the Bivariate Distribution Using the Morgenstern Formula

Marginal distribution	N	U	SE	SR	T1L	T1S	LN	GM	T2L	T3S
N	0.318									
U	0.326	0.333								
SE	0.282	0.289	0.25							
SR	0.316	0.324	0.28	0.314						
T1L	0.305	0.312	0.27	0.303	0.292					
T1S	0.305	0.312	0.27	0.303	0.292	0.292				
LN	<0.318	<0.326	<0.282	<0.316	<0.305	<0.305	<0.318			
GM	<0.318	<0.326	<0.282	<0.316	<0.305	<0.305	<0.318	<0.381		
T2L	<0.305	<0.312	<0.270	<0.303	<0.292	<0.292	<0.305	<0.305	<0.292	
T3S	<0.305	<0.312	<0.270	<0.303	<0.292	<0.292	<0.305	<0.305	<0.292	<0.292

NOTE : N = normal; U = uniform; SE = shifted exponential; SR = shifted Rayleigh; T1L = type I largest value; T1S = type I smallest value; LN = lognormal; GM = gamma; T2L = type II largest value; T3S = type III smallest value.
SOURCE : After Lin and Der Kiureghian (1986).

To compute the correlation coefficient between X and Y, one first computes the covariance of X and Y as $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$, in which $E(XY)$ is computed by

$$E(XY) = \int_0^\infty \int_0^\infty xy f_{x,y}(x, y)\, dx\, dy = 1 + \frac{\theta}{4}$$

Referring to Eq. (2.79), since the exponential random variables X and Y currently considered are the special cases of $\beta = 1$, therefore, $\mu_x = \mu_y = 1$ and $\sigma_x = \sigma_y = 1$. Consequently, the covariance of X and Y is $\theta/4$, and the corresponding correlation coefficient is $\theta/4$. Note that the weighing constant θ lies between $[-1, 1]$. The preceding bivariate exponential distribution obtained from the Morgenstern formula could only be valid for X and Y having a correlation coefficient in the range $[-1/4, 1/4]$.

2.7.1 Multivariate normal distributions

A *bivariate normal distribution* has a PDF defined as

$$f_{x_1, x_2}(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho_{12}^2}} \exp\left[\frac{-Q}{2(1 - \rho_{12}^2)}\right] \tag{2.108}$$

for $-\infty < x_1, x_2 < \infty$, in which

$$Q = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho_{12}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)$$

where μ and σ are, respectively, the mean and standard deviation, the subscripts 1 and 2 indicate the random variables X_1 and X_2, respectively, and ρ_{12} is the correlation coefficient of the two random variables. Plots of the bivariate normal PDF in a three-dimensional form are shown in Fig. 2.26. The contour curves of the bivariate normal PDF of different correlation coefficients are shown in Fig. 2.27.

The marginal PDF of X_k can be derived, according to Eq. (2.8), as

$$f_k(x_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_k - \mu_k}{\sigma_k}\right)^2\right] \qquad \text{for } -\infty < x_k < \infty$$

for $k = 1$ and 2. As can be seen, the two random variables having a bivariate normal PDF are, individually, normal random variables. It should be pointed out that given two normal marginal PDFs, one can construct a bivariate PDF that is not in the form of a bivariate normal as defined by Eq. (2.108).

According to Eq. (2.17), the *conditional normal PDF* of $X_1 \mid x_2$ can be obtained as

$$f_{x_1 \mid x_2}(x_1 \mid x_2) = \frac{1}{\sigma_1 \sqrt{2\pi\left(1 - \rho_{12}^2\right)}} \exp\left\{-\frac{1}{2}\left[\frac{(x_1 - \mu_1) - \rho_{12}(\sigma_1/\sigma_2)(x_2 - \mu_2)}{\sigma_1 \sqrt{1 - \rho_{12}^2}}\right]^2\right\}$$

$$\tag{2.109}$$
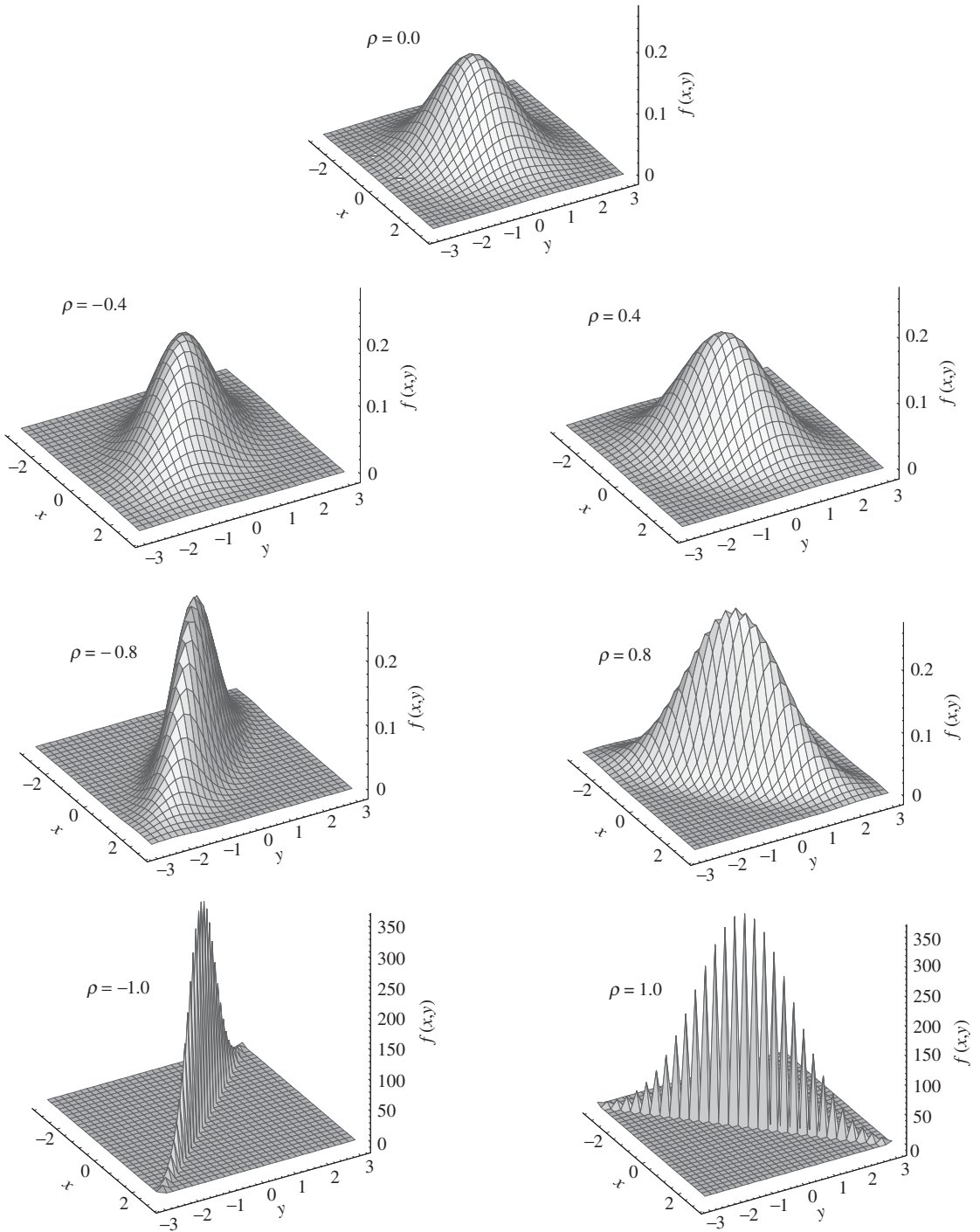
Figure 2.26 Three-dimensional plots of bivariate standard normal probability density functions. (*After Johnson and Kotz, 1976.*)
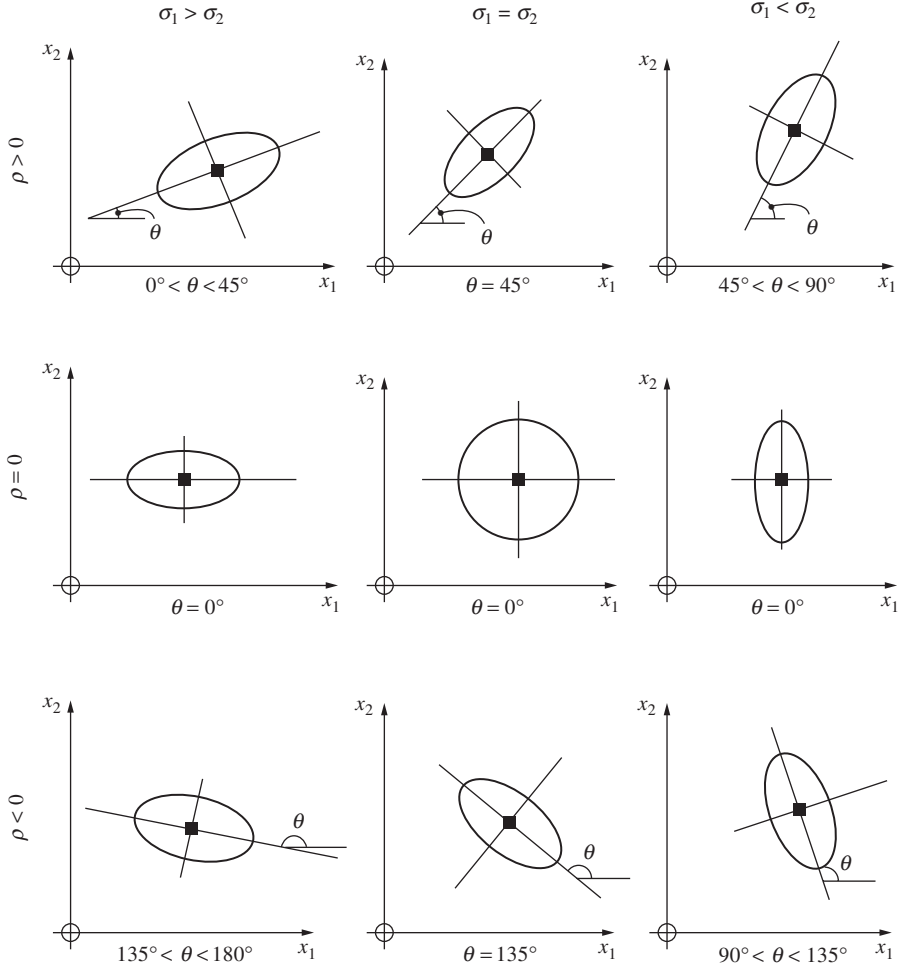
Figure 2.27 Contour of equal density of bivariate standard normal probability density functions. (*After Johnson and Kotz, 1976.*)

for $-\infty < x_1 < \infty$. Based on Eq. (2.109), the *conditional expectation and variance of the normal random variable* $X_1 \,|\, x_2$ can be obtained as

$$E(X_1 \,|\, x_2) = \mu_1 + \rho_{12}(\sigma_1/\sigma_2)(x_2 - \mu_2) \tag{2.110}$$

$$\mathrm{Var}(X_1 \,|\, x_2) = \sigma_1^2\big(1 - \rho_{12}^2\big) \tag{2.111}$$

Expressions of the conditional PDF, expectation, and variance for $X_2 \,|\, x_1$ can be obtained immediately by exchanging the subscripts in Eqs. (2.109) through (2.111).

For the general case involving K correlated normal random variables, the *multivariate normal PDF* is

$$f_x(\boldsymbol{x}) = \frac{|\boldsymbol{C}_x^{-1}|^{1/2}}{(2\pi)^{K/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^t \boldsymbol{C}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right] \qquad \text{for } -\infty < \boldsymbol{x} < \infty \quad (2.112)$$

in which $\boldsymbol{\mu}_x = (\mu_1, \mu_2, \ldots, \mu_K)^t$, a $K \times 1$ column vector of the mean values of the variables, with the superscript t indicating the transpose of a matrix or vector, and $\boldsymbol{C}_x$ is a $K \times K$ covariance matrix:

$$\text{Cov}(\boldsymbol{X}) = \boldsymbol{C}_x = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \cdot & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \cdot & \sigma_{2K} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{K1} & \sigma_{K2} & \cdot & \cdot & \cdot & \sigma_{KK} \end{bmatrix}$$

This covariance matrix is symmetric, that is, $\sigma_{jk} = \sigma_{kj}$, for $j \neq k$, where $\sigma_{jk} = \text{Cov}(X_j, X_k)$. In matrix notation, the covariance matrix for a vector of random variables can be expressed as

$$\boldsymbol{C}_x = E\left[(\boldsymbol{X}-\boldsymbol{\mu}_x)(\boldsymbol{X}-\boldsymbol{\mu}_x)^t\right] \qquad (2.113)$$

In terms of standard normal random variables, $Z_k = (X_k - \mu_k)/\sigma_k$, the *standardized multivariate normal PDF*, can be expressed as

$$\phi(\boldsymbol{z}) = \frac{|\boldsymbol{R}_x^{-1}|^{1/2}}{(2\pi)^{K/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}^t \boldsymbol{R}_x^{-1}\boldsymbol{z}\right) \qquad \text{for } -\infty < \boldsymbol{z} < \infty \qquad (2.114)$$

in which $\boldsymbol{R}_x = \boldsymbol{C}_z = E(\boldsymbol{ZZ}^t)$ is a $K \times K$ correlation matrix:

$$\text{Corr}(\boldsymbol{X}) = \text{Cov}(\boldsymbol{Z}) = \boldsymbol{R}_x = \begin{bmatrix} 1 & \rho_{12} & \cdot & \cdot & \cdot & \rho_{1K} \\ \rho_{21} & 1 & \cdot & \cdot & \cdot & \rho_{2K} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{K1} & \rho_{K2} & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

with $\rho_{jk} = \text{Cov}(Z_j, Z_k)$ being the correlation coefficient between each pair of normal random variables X_j and X_k. For bivariate standard normal variables,

the following relationships of cross-product moments are useful (Hutchinson and Lai, 1990):

$$E\left[Z_1^{2m} Z_2^{2n}\right] = \frac{(2m)!(2n)!}{2^{m+n}} \sum_{j=0}^{\min(m,n)} \frac{(2\rho_{12})^{2j}}{(m-j)!(n-j)!\,(2j)!}$$

$$E\left[Z_1^{2m+1} Z_2^{2n+1}\right] = \frac{(2m+1)!(2n+1)!}{2^{m+n}} \rho_{12} \sum_{j=0}^{\min(m,n)} \frac{(2\rho_{12})^{2j}}{(m-j)!(n-j)!(2j+1)!}$$

$$E\left[Z_1^{2m+1} Z_2^{2n}\right] = E\left[Z_1^{2m} Z_2^{2n+1}\right] = 0 \tag{2.115}$$

for m and n being positive integer numbers.

2.7.2 Computation of multivariate normal probability

Evaluation of the probability of multivariate normal random variables involves multidimensional integration as

$$\Phi(\boldsymbol{z} \mid \boldsymbol{R}_x) = P(Z_1 \le z_1, Z_2 \le z_2, \ldots, Z_K \le z_K \mid \boldsymbol{R}_x) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \cdots \int_{-\infty}^{z_K} \phi(\boldsymbol{z} \mid \boldsymbol{R}_x)\, d\boldsymbol{z} \tag{2.116}$$

Accurate evaluation for $\Phi(\boldsymbol{z} \mid \boldsymbol{R}_x)$ is generally difficult and often is resolved by approximations.

Bivariate normal probability.   For a bivariate case, Fortran computer programs for computing the lower left volume under the density surface, that is, $\Phi(a, b \mid \rho) = P(Z_1 \le a, Z_2 \le b \mid \rho)$, have been developed by Donnelly (1973) and Baughman (1988). The double integral for evaluating the bivariate normal probability can be reduced to a single integral as shown in Eq. (2.111). Several approximations have been derived (Johnson and Kotz, 1976). Derivation of bounds for the bivariate normal probability is presented in Sec. 2.7.3. For a bivariate normal probability, exact solutions have been obtained in the form of figures such as Fig. 2.28 for computing the upper-right volume under the bivariate standard normal density surface, that is, $L(a, b \mid \rho) = P(Z_1 \ge a, Z_2 \ge b \mid \rho)$, in which $L(a, b \mid \rho)$ can be expressed in terms of $L(a, 0 \mid \rho)$ as

$$L(a, b \mid \rho) = L\left(a, 0 \,\middle|\, \frac{(\rho a - b)(\operatorname{sign} a)}{\sqrt{a^2 - 2\rho ab + b^2}}\right) + L\left(b, 0 \,\middle|\, \frac{(\rho b - a)(\operatorname{sign} b)}{\sqrt{a^2 - 2\rho ab + b^2}}\right)$$

$$\begin{cases} 0, & \text{if } (ab > 0 \text{ or } ab = 0) \text{ and } a + b \ge 0 \\ \dfrac{1}{2}, & \text{otherwise} \end{cases} \tag{2.117}$$

$L(a, 0|\rho)$ for $0 \le a \le 1$ and $-1 \le \rho \le 0$.

Values for $a < 0$ can be obtained using $L(a, 0|{-}\rho) = 0.5 - L(-a, 0|\rho)$
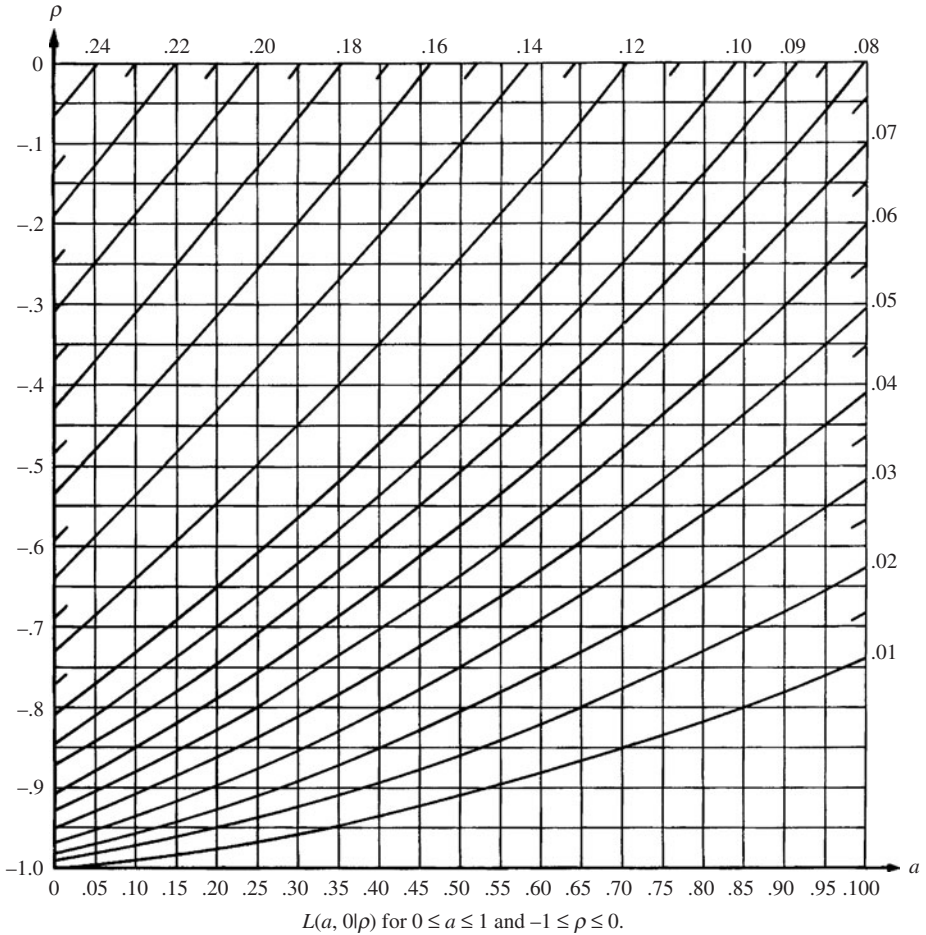
Figure 2.28   Bivariate normal cumulative distribution function. (*After Abramowitz and Stegun, 1972.*)

The relationship between $\Phi(a, b \mid \rho)$ and $L(a, b \mid \rho)$ is

$$\Phi(a, b \mid \rho) = -1 + \Phi(a) + \Phi(b) + L(a, b \mid \rho) \tag{2.118}$$

From the definitions of $\Phi(a, b \mid \rho)$ and $L(a, b \mid \rho)$ and the symmetry of the bivariate normal PDF, the following relations are in order:

$$\Phi(a, \infty \mid \rho) = \Phi(a) \qquad \Phi(\infty, b \mid \rho) = \Phi(b) \tag{2.119a}$$

$$L(a, -\infty \mid \rho) = 1 - \Phi(a) \qquad L(-\infty, b \mid \rho) = 1 - \Phi(b) \tag{2.119b}$$

$L(a, 0|\rho)$ for $0 \leq a \leq 1$ and $0 \leq \rho \leq 1$.

Values for $a < 0$ can be obtained using $L(a, 0|-\rho) = 0.5 - L(-a, 0|\rho)$
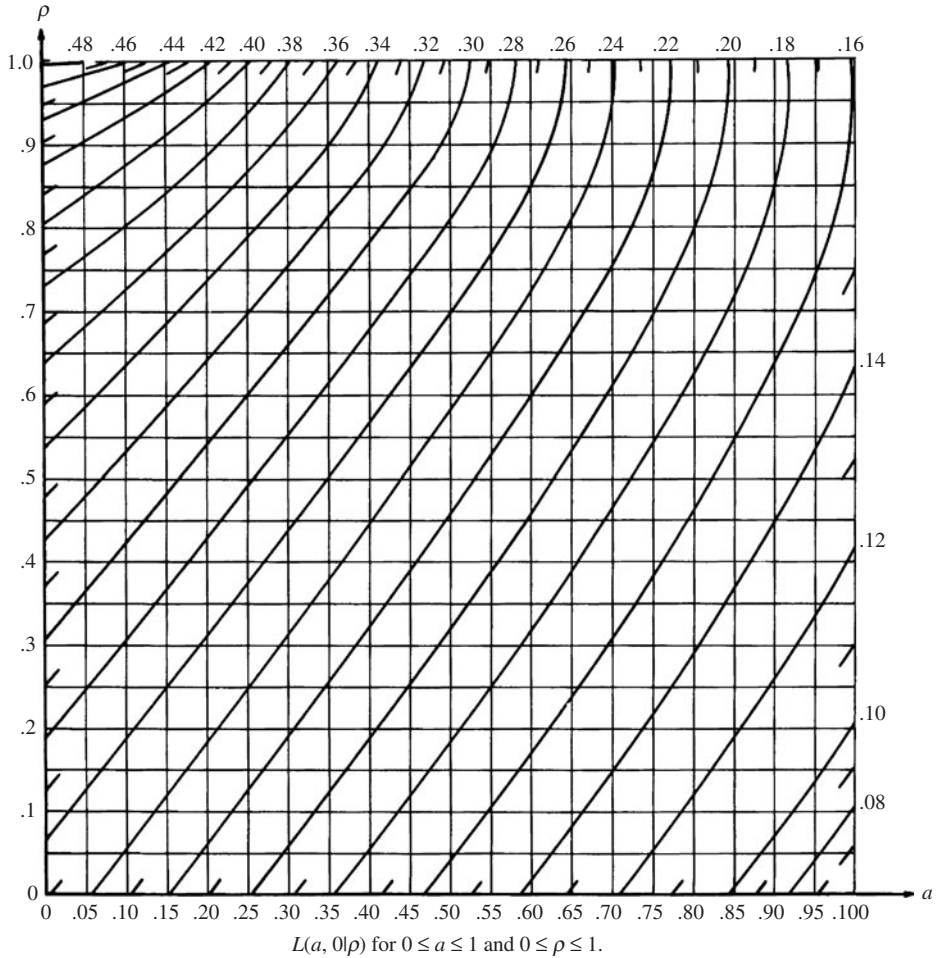
Figure 2.28 (*Continued*)

$$L(a, b \mid \rho) = L(b, a \mid \rho) \tag{2.119c}$$

$$L(-a, b \mid \rho) + L(a, b \mid -\rho) = 1 - \Phi(b) \tag{2.119d}$$

$$L(-h, -k \mid \rho) - L(k, h \mid \rho) = 1 - \Phi(h) - \Phi(k) \tag{2.119e}$$

Example 2.21 Consider two correlated normal random variables X_1 and X_2 with their statistical properties being

$$E(X_1) = 10 \quad \text{Var}(X_1) = 9 \quad E(X_2) = 5 \quad \text{Var}(X_2) = 4 \quad \text{Cov}(X_1, X_2) = 3.6$$

Compute $P(X_1 \leq 13, X_2 \leq 3)$.
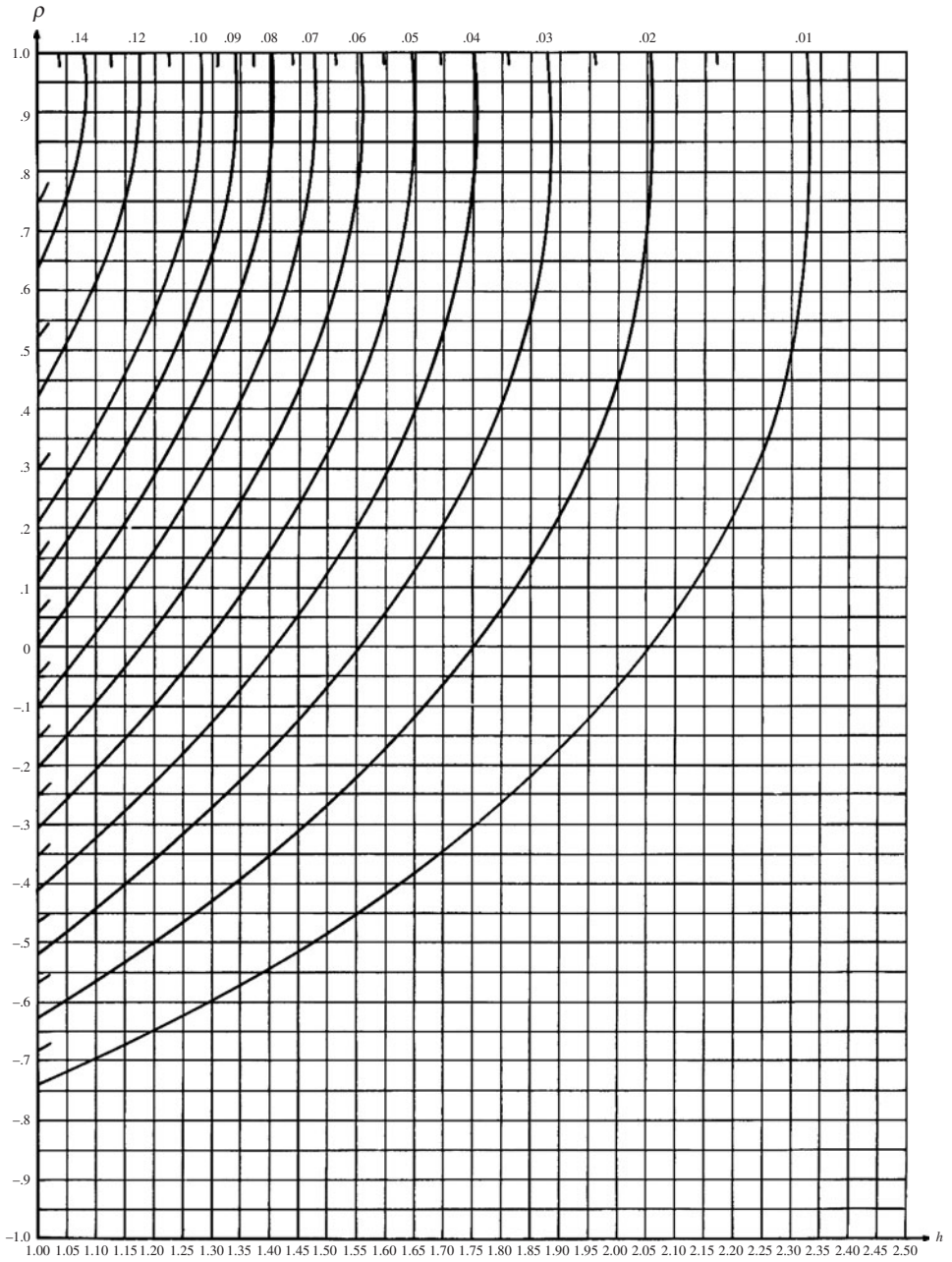
$L(a, 0|\rho)$ for $a \geq 1$ and $-1 \leq \rho \leq 1$.

Values for $a < 0$ can be obtained using $L(a, 0|-\rho) = 0.5 - L(-a, 0|\rho)$

Figure 2.28 (*Continued*)

Solution Based on the given information, the correlation coefficient between X_1 and X_2 is

$$\rho_{1,2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{x_1}\sigma_{x_2}} = \frac{3.6}{\sqrt{9}\sqrt{4}} = 0.6$$

Then

$$P(X_1 \leq 13, X_2 \leq 3) = P\left(Z_1 \leq \frac{13 - 10}{3}, Z_2 \leq \frac{3 - 5}{2} \middle| \rho_{1,2} = 0.6\right)$$

$$= P(Z_1 \leq 1, Z_2 \leq -1 \mid \rho_{1,2} = 0.6)$$

$$= \Phi(a = 1, b = -1 \mid \rho_{1,2} = 0.6)$$

By Eq. (2.118),

$$\Phi(1, -1 | 0.6) = -1 + \Phi(1) + \Phi(-1) + L(1, -1 | 0.6) \qquad \text{(a)}$$

Since $ab = -1 < 0$, according to Eq. (2.117),

$$L(1, -1 \mid 0.6) = L\left(1, 0 \middle| \frac{1.6}{\sqrt{3.2}}\right) + L\left(-1, 0 \middle| \frac{1.6}{\sqrt{3.2}}\right) - \frac{1}{2}$$

$$= L(1, 0 \mid 0.894) + L(-1, 0 \mid 0.894) - 0.5$$

From Fig. 2.28b, $L(1, 0 \mid 0.894) = 0.159$. Since $L(-1, 0 \mid 0.894) = 0.5 - L(1, 0 \mid -0.894)$, and according to Fig. 2.28a, by extrapolation, $L(1, 0 \mid -0.894) = 0.004$,

$$L(-1, 0 \mid 0.894) = 0.5 - 0.004 = 0.496$$

Consequently, $L(1, -1 \mid 0.6) = 0.159 + 0.496 - 0.5 = 0.155$.

According to (a), the bivariate normal probability is $\Phi(1, -1 \mid 0.6) = L(1, -1 \mid 0.6) = 0.155$. Alternatively, the bivariate normal probability can be computed by Eq. (2.121), and the result of the numerical integration for this example is 0.1569.

Multivariate normal probability. Johnson and Kotz (1976) show that if the correlation coefficient ρ_{ij} can be expressed as $\rho_{ij} = \lambda_i \lambda_j$ for all i and j and $|\lambda_j| \leq 1$, then each correlated standard normal random variable Z_k can be represented by

$$Z_k = \lambda_k Z_0' + \sqrt{1 - \lambda_k^2} Z_k' \qquad \text{for } k = 1, 2, \ldots, K$$

where $Z_0', Z_1', Z_2', \ldots, Z_K'$ are *independent* standard normal variables. The inequality $Z_k \leq z_k$ can be expressed as

$$Z_k' \leq \frac{z_k - \lambda_k z_0'}{\sqrt{1 - \lambda_k^2}}$$

Then the multivariate normal probability can be calculated as

$$\Phi(\boldsymbol{z}\,|\,\boldsymbol{R}_x) = \int_{-\infty}^{\infty} \phi(u) \left[\prod_{k=1}^{K} \Phi\left(\frac{z_k - \lambda_k u}{\sqrt{1 - \lambda_k^2}} \right) \right] du \qquad (2.120)$$

As can be seen from Eq. (2.120), the computation of the multivariate normal probability is reduced from multiple integrals to a single integral for which the result can be obtained accurately by various numerical integration techniques. (see Appendix 4A) Under the special case of equicorrelation, that is, $\rho_{ij} = \rho$, for all i and j, the multivariate normal probability can be computed as

$$\Phi(\boldsymbol{z}\,|\,\boldsymbol{R}_x) = \int_{-\infty}^{\infty} \phi(u) \left[\prod_{k=1}^{K} \Phi\left(\frac{z_k - \sqrt{\rho}\,u}{\sqrt{1 - \rho}} \right) \right] du \qquad (2.121)$$

This equation, in particular, can be applied to evaluate the bivariate normal probability.

Under the general unequal correlation case, evaluation of multivariate normal probability by Eq. (2.120) requires solving for K λ's based on $K(K-1)/2$ different values of ρ in the correlation matrix $\boldsymbol{R}_x$. This may not necessarily be a trivial task. Ditlevsen (1984) proposed an accurate algorithm by expanding $\Phi(\boldsymbol{z}\,|\,\boldsymbol{R}_x)$ in a Taylor series about an equal correlation $\rho_{ij} = \rho > 0$, for $i \neq j$. The equicorrelation ρ is determined in such a way that the first-order expansion term $d\Phi(z\,|\,\rho)$ vanishes. The term $d\Phi(z\,|\,\rho)$ can be expressed as

$$d\Phi(z\,|\,\rho) = \frac{1}{1-\rho} \int_{-\infty}^{\infty} \phi(u) \left[\prod_{k=1}^{K} \Phi\left(\frac{z_k - \sqrt{\rho}\,u}{\sqrt{1 - \rho}} \right) \right] \left[\sum_{k=1}^{K-1} \sum_{j=k+1}^{K} a_k(u) a_j(u) \Delta \rho_{kj} \right] du$$

$$(2.122)$$

where

$$a_k(u) = \phi\left(\frac{z_k - \sqrt{\rho}\,u}{\sqrt{1 - \rho}} \right) \Big/ \Phi\left(\frac{z_k - \sqrt{\rho}\,u}{\sqrt{1 - \rho}} \right) \qquad (2.123)$$

and $\Delta \rho_{ij} = \rho_{ij} - \rho$. In computing the value of $d\,\Phi(z\,|\,\rho)$, numerical integration generally is required. However, one should be careful about the possible numerical overflow associated with the computation of $a_k(u)$ as the value of u gets large. It can be shown that by the L'Hospital rule, $\lim_{u \to \infty} a_k(u) = 0$ and $\lim_{u \to -\infty} a_k(u) = -u$.

Determination of the equicorrelation ρ for the expansion point can be made through the iterative procedure, as outlined in Fig. 2.29. A sensible starting value for ρ is the average of ρ_{ij}:

$$\rho = \frac{2}{K(K-1)} \sum_{i<j} \rho_{ij} \qquad (2.124)$$
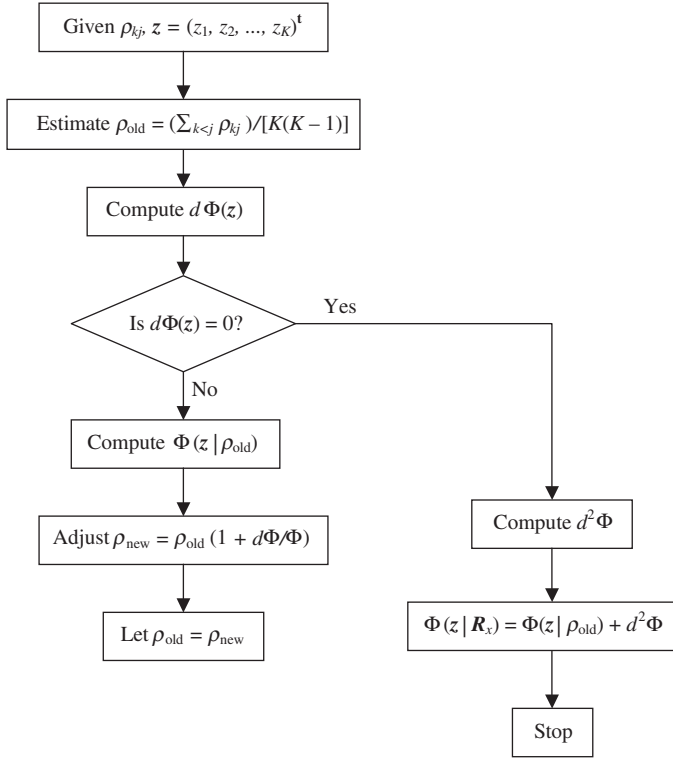
Figure 2.29 Flowchart for determining the equicorrelation in a Taylor series expansion.

Once such ρ is found, Ditlevsen (1984) suggests that the value of $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$ can be estimated accurately by a second-order approximation as

$$\Phi(\boldsymbol{z}|\boldsymbol{R}_x) \approx \Phi(\boldsymbol{z}|\rho) + \frac{1}{2}d^2\Phi(\boldsymbol{z}|\rho) \tag{2.125}$$

in which $\Phi(\boldsymbol{z}|\rho)$ is computed by Eq. (2.121), and the second-order error term is computed as

$$d^2\Phi(\boldsymbol{z}|\rho) = \frac{1}{4(1-\rho)^2} \int_{-\infty}^{\infty} \phi(u) \left[\prod_{k=1}^{K} \Phi\left(\frac{z_k - \sqrt{\rho}u}{\sqrt{1-\rho}}\right)\right]$$

$$\left[\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{r=1}^{K}\sum_{s=1}^{K} a_i(u)a_j(u)a_r(u)a_s(u)(-b_r)^{\delta_{ir}}(-b_s)^{\delta_{js}}\,\Delta\rho_{ij}\,\Delta\rho_{rs}\right] du$$

$$\tag{2.126}$$

where δ_{ir} is the Kronecker's delta, having a value of 1 if $i = r$ and otherwise 0, and

$$b_k(u) = \frac{z_k \sqrt{\rho}u}{a_k(u)\sqrt{1-\rho}} \qquad (2.127)$$

As can be seen, the evaluation of $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$ is reduced from a multiple integral to a single integral, which can be executed efficiently and accurately by many numerical integration algorithms.

For the univariate normal distribution, an asymptotic expansion of $\Phi(z)$ is (Abramowitz and Stegun, 1972) for $z \to \infty$

$$\Phi(-z) \approx \frac{\phi(z)}{z}\left[1 - \frac{1}{z^2} + \frac{1 \times 3}{z^4} - \frac{1 \times 3 \times 5}{z^6} + \cdots + (-1)^n\frac{1 \times 3 \times \cdots \times (2n-1)}{z^{2n}}\right]$$
$$(2.128)$$

This expansion for $\Phi(z)$ is smaller than every summand with an odd number of terms and is larger than every summand with an even number of terms. The truncation error decreases as the number of terms increases. Note that Eq. (2.128) is particularly appropriate for evaluating the normal tail probability. The expansion has been generalized by Ruben (1964) for the multivariate normal distribution as

$$\Phi(-\boldsymbol{z}|\boldsymbol{R}_x) \approx \frac{\exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{t}}\boldsymbol{R}_x^{-1}\boldsymbol{z}\right)}{\sqrt{2\pi}\sqrt{|\boldsymbol{R}_x|}\prod_{k=1}^{K}a_k} \qquad \text{for } |\boldsymbol{z}| \to \infty \qquad (2.129)$$

in which the coefficients a_k are elements in a vector $\boldsymbol{a}$ obtained from

$$\boldsymbol{a} = \boldsymbol{R}_x^{-1}\boldsymbol{z} \qquad (2.130)$$

It should be noted that Eq. (2.130) is valid only when all coefficients a_k are positive. The right-hand-side of Eq. (2.129) provides an upper bound for the multivariate normal probability.

2.7.3 Determination of bounds on multivariate normal probability

Instead of computing the exact value of $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$, several methods have been proposed to determine the bounds on the exact value of $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$. This section describes three such bounds.

Bounds of Rackwitz. The scheme of Rackwitz (1978) is based on the decomposition of a positive correlation coefficient $\rho_{ij} = \lambda_i\lambda_j$, for $i, j = 1, 2, \ldots, K$. The multivariate normal probability $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$ is obtained according to Eq. (2.120). Instead of solving for the exact values for all $K(K-1)/2\lambda$s, Rackwitz selects the smallest three values of $\boldsymbol{z} = (z_1, z_2, \ldots, z_K)^{\mathrm{t}}$ in $\Phi(\boldsymbol{z}|\boldsymbol{R}_x)$ and solves for the corresponding λs that satisfy $\rho_{ij} = \lambda_i\lambda_j$, for $i, j = [1], [2], [3]$ with subscript $[i]$

representing the rank of zs in ascending order, that is, $z_{[1]} \le z_{[2]} \le z_{[3]} \le \cdots \le z_{[K-1]} \le z_{[K]}$. For example, assume that all ρ_{ij}s are positive. Based on the three smallest zs, one can solve for $\lambda_{[i]}$ for $i = 1, 2, 3$ in terms of $\rho_{[i][j]}$ as

$$\lambda_{[1]} = \left(\frac{\rho_{[1][2]} \, \rho_{[1][3]}}{\rho_{[2][3]}} \right)^{1/2} \qquad \lambda_{[2]} = \left(\frac{\rho_{[1][2]} \, \rho_{[2][3]}}{\rho_{[1][3]}} \right)^{1/2} \qquad \lambda_{[3]} = \left(\frac{\rho_{[1][3]} \, \rho_{[2][3]}}{\rho_{[1][2]}} \right)^{1/2}$$

(2.131)

For the remaining λs, their values can be computed as

$$\lambda_{iU} = \max_{j < i} \left| \frac{\rho_{ij}}{\lambda_{jU}} \right| \qquad i = [4], [5], \ldots, [K] \tag{2.132a}$$

$$\lambda_{iL} = \min_{j < i} \left| \frac{\rho_{ij}}{\lambda_{jL}} \right| \qquad i = [4], [5], \ldots, [K] \tag{2.132b}$$

The upper bound and lower bound of $\Phi(\boldsymbol{z} \,|\, \boldsymbol{R}_x)$ can be obtained by Eq. (2.120) along with λs computed by Eqs. (2.132a) and (2.132b), respectively.

Bounds of Ditlevsen. Ditlevsen (1979) proposed an approach for the bounds of the multivariate normal probability as follows:

$$\Phi_U(\boldsymbol{z} \,|\, \boldsymbol{R}_x) = \Phi(z_1) - \sum_{k=2}^{K} \max \left\{ 0, \left[\Phi(-z_k) - \sum_{j=1}^{k-1} \Phi(-z_k, -z_j \,|\, \rho_{kj}) \right] \right\} \tag{2.133a}$$

$$\Phi_L(\boldsymbol{z} \,|\, \boldsymbol{R}_x) = \Phi(z_1) - \sum_{k=2}^{K} \max \left\{ \Phi(-z_k) - \max_{j < k} \left[\Phi(-z_k, -z_j \,|\, \rho_{kj}) \right] \right\} \tag{2.133b}$$

in which $\Phi_U(\boldsymbol{z} \,|\, \boldsymbol{R}_x)$ and $\Phi_L(\boldsymbol{z} \,|\, \boldsymbol{R}_x)$ are the upper and lower bounds of the multivariate normal probability, respectively, and $\Phi(z_k, z_j \,|\, \rho_{kj})$ is the bivariate normal probability. Ditlevsen (1979) further simplified these bounds to involve the evaluation of only the univariate normal probability at the expense of having a more complicated algebraic expression where a narrow bound can be obtained under $|\rho| < 0.6$. For a larger correlation coefficient, Ditlevsen (1982) proposed a procedure using conditioning to obtain a narrow bound. The derivations of various probability bounds for system reliability are presented in Sec. 7.2.5

Example 2.22 Z_1, Z_2, Z_3, Z_4, and Z_5 are correlated standard normal variables with the following correlation matrix:

$$\boldsymbol{R}_x = \begin{bmatrix} 1.00 & 0.80 & 0.64 & 0.51 & 0.41 \\ 0.80 & 1.00 & 0.80 & 0.64 & 0.51 \\ 0.64 & 0.80 & 1.00 & 0.80 & 0.64 \\ 0.51 & 0.64 & 0.80 & 1.00 & 0.80 \\ 0.41 & 0.51 & 0.64 & 0.80 & 1.00 \end{bmatrix}$$

Determine the multivariate probability $P(Z_1 \leq -1, Z_2 \leq -2, Z_3 \leq 0, Z_4 \leq 2, Z_5 \leq 1)$ by Ditlevsen's approach using the Taylor series expansion. Also compute the bounds for the preceding multivariate normal probability using Rackwitz's and Ditlevsen's approaches.

Solution Using Ditlevsen's Taylor series expansion approach, the initial equicorrelation value can be used according to Eq. (2.124) as $\rho = 0.655$. The corresponding multivariate normal probability, based on Eq. (2.121), is $\Phi(z \mid \rho = 0.655) = 0.01707$. From Eq. (2.122), the first-order error, $d\,\Phi(z \mid \rho = 0.655)$, is 0.003958. Results of iterations according to the procedure outlined in Fig. 2.29 are shown below:

i	ρ	$\Phi(\boldsymbol{z} \mid \rho)$	$d\,\Phi(\boldsymbol{z} \mid \rho)$
1	0.6550	0.01707	0.3958×10^{-2}
2	0.8069	0.02100	-0.1660×10^{-3}
3	0.8005	0.02086	-0.3200×10^{-4}
4	0.7993	0.02083	-0.5426×10^{-5}

At $\rho = 0.7993$, the corresponding second-order error term in the Taylor series expansion, according to Eq. (2.126), is

$$d^{\,2}\Phi(\boldsymbol{z} \mid \rho) = 0.01411$$

Based on Eq. (2.125), the multivariate normal probability can be estimated as

$$\Phi(\boldsymbol{z} \mid \boldsymbol{R}_x) = \Phi(\boldsymbol{z} \mid \rho = 0.7993) + 0.5 d^{\,2}\Phi(\boldsymbol{z} \mid \rho = 0.7993)$$

$$= 0.02083 + 0.5(0.01411)$$

$$= 0.02789$$

Using the Rackwitz approach for computing the bounds of the multivariate normal probability, the values of zs are arranged in ascending order as $(z_{[1]}, z_{[2]}, z_{[3]}, z_{[4]}, z_{[5]}) = (-2, -1, 0, 1, 2) = (z_2, z_1, z_3, z_5, z_4)$ with the corresponding correlation matrix as

$$\boldsymbol{R}_{[kj]} = \begin{bmatrix} 1.00 & 0.80 & 0.80 & 0.51 & 0.41 \\ 0.80 & 1.00 & 0.64 & 0.41 & 0.51 \\ 0.80 & 0.64 & 1.00 & 0.64 & 0.80 \\ 0.51 & 0.41 & 0.80 & 1.00 & 0.80 \\ 0.64 & 0.51 & 0.80 & 0.80 & 1.00 \end{bmatrix}$$

The values of λs corresponding to the three smallest zs, that is, -2, -1, and 0, are computed according to Eq. (2.131), and the results are

$$\lambda_{[1]} = 1.00 \qquad \lambda_{[2]} = 0.80 \qquad \lambda_{[3]} = 0.80$$

Using Eq. (2.132a), the values of the remaining λs for computing the upper bound are obtained as

$$\lambda_{[4],U} = \max \left| \frac{\rho_{[41]}}{\overline{\lambda}_{[1]}}, \frac{\rho_{[42]}}{\overline{\lambda}_{[2]}}, \frac{\rho_{[43]}}{\overline{\lambda}_{[3]}} \right| = \max \left| \frac{0.51}{1.00}, \frac{0.41}{0.80}, \frac{0.64}{0.80} \right| = 0.80$$

$$\lambda_{[5],U} = \max \left| \frac{\rho_{[51]}}{\overline{\lambda}_{[1]}}, \frac{\rho_{[52]}}{\overline{\lambda}_{[2]}}, \frac{\rho_{[53]}}{\overline{\lambda}_{[3]}}, \frac{\rho_{[54]}}{\overline{\lambda}_{[4]}} \right| = \max \left| \frac{0.64}{1.00}, \frac{0.51}{0.80}, \frac{0.80}{0.80}, \frac{0.80}{0.80} \right| = 1.00$$

and, by the same token, for the lower bound are

$$\lambda_{[4],L} = 0.51 \qquad \lambda_{[5],L} = 0.6375$$

Applying Eq. (2.120), along with $\lambda_U = (1.0, 0.8, 0.8, 0.8, 1.0)$, one obtains the upper bound for the multivariate normal probability $\Phi_U(\boldsymbol{z}|\boldsymbol{R}_x) = 0.01699$. Similarly, using $\lambda_L = (1.0, 0.8, 0.8, 0.51, 0.6375)$, the lower bound is obtained as $\Phi_L(\boldsymbol{z}|\boldsymbol{R}_x) = 0.01697$.

To use Eqs. (2.133a) and (2.133b) for computing the upper and lower bounds for the multivariate normal probability, the marginal probabilities and each pair of bivariate normal probabilities are computed first, according to Eq. (2.3). The results are

$$\Phi(z_1) = 0.1587 \quad \Phi(z_2) = 0.02275 \quad \Phi(z_3) = 0.5000 \quad \Phi(z_4) = 0.9772 \quad \Phi(z_5) = 0.8413$$

$$\Phi(-z_1, -z_2) = 0.8395 \quad \Phi(-z_1, -z_3) = 0.4816$$

$$\Phi(-z_1, -z_4) = 0.0226 \quad \Phi(-z_1, -z_5) = 0.1523$$

$$\Phi(-z_2, -z_3) = 0.5000 \quad \Phi(-z_2, -z_4) = 0.0228 \quad \Phi(-z_2, -z_5) = 0.1585$$

$$\Phi(-z_3, -z_4) = 0.0227 \quad \Phi(-z_3, -z_5) = 0.1403 \quad \Phi(-z_4, -z_5) = 0.0209$$

The lower and upper bounds of the multivariate probability can be obtained as 0.02070 and 0.02086, respectively.

2.7.4 Multivariate lognormal distributions

Similar to the univariate case, bivariate lognormal random variables have a PDF

$$f_{x_1,x_2}(x_1, x_2) = \frac{1}{2\pi x_1 x_2 \sigma_{\ln x_1} \sigma_{\ln x_2} \sqrt{1 - \rho_{12}'^2}} \exp\left[\frac{-Q'}{2(1 - \rho_{12}'^2)}\right] \tag{2.134}$$

for $x_1, x_2 > 0$, in which

$$Q' = \frac{[\ln(x_1) - \mu_{\ln x_1}]^2}{\sigma_{\ln x_1}^2} + \frac{[\ln(x_2) - \mu_{\ln x_2}]^2}{\sigma_{\ln x_2}^2} - 2\rho_{12}' \frac{[\ln(x_1) - \mu_{\ln x_1}][\ln(x_2)\mu_{\ln x_2}]}{\sigma_{\ln x_1} \sigma_{\ln x_2}}$$

where $\mu_{\ln x}$ and $\sigma_{\ln x}$ are the mean and standard deviation of log-transformed random variables, subscripts 1 and 2 indicate the random variables X_1 and X_2, respectively, and $\rho_{12}' = \text{Corr}(\ln X_1, \ln X_2)$ is the correlation coefficient of the two log-transformed random variables. After log-transformation is made, properties of multivariate lognormal random variables follow exactly as for the multivariate normal case. The relationship between the correlation coefficients in the original and log-transformed spaces can be derived using the moment-generating function (Tung and Yen, 2005, Sec. 4.2) as

$$\text{Corr}(X_1, X_2) = \rho_{12} = \frac{\exp\left(\rho_{12}' \sigma_{\ln x_1} \sigma_{\ln x_2}\right) - 1}{\sqrt{\exp\left(\sigma_{\ln x_1}^2\right) - 1} \sqrt{\exp\left(\sigma_{\ln x_2}^2\right) - 1}} \tag{2.135}$$

Example 2.23   Resolve Example 2.21 by assuming that both X_1 and X_2 are bivariate lognormal random variables.

Solution   Since X_1 and X_2 are lognormal variables,

$$P(X_1 \le 13, X_2 \le 3) = P\left[\ln(X_1) \le \ln(13), \ln(X_2) \le \ln(3)\right]$$

$$= P\left(Z_1 \le \frac{\ln(13) - \mu'_1}{\sigma'_1}, Z_2 \le \frac{\ln(3) - \mu'_2}{\sigma'_2}\bigg|\rho'\right)$$

in which μ'_1, μ'_2, σ'_1, and σ'_2 are the means and standard deviations of $\ln(X_1)$ and $\ln(X_2)$, respectively; ρ' is the correlation coefficient between $\ln(X_1)$ and $\ln(X_2)$. The values of μ_1, μ_2, σ_1, and σ_2 can be computed, according to Eqs. (2.67a) and (2.67b), as

$$\sigma'_1 = \sqrt{\ln(1 + 0.3^2)} = 0.294 \qquad \sigma'_2 = \sqrt{\ln(1 + 0.4^2)} = 0.385$$

$$\mu'_1 = \ln(10) - \frac{1}{2}(0.294)^2 = 2.259 \qquad \mu'_2 = \ln(5) - \frac{1}{2}(0.385)^2 = 1.535$$

Based on Eq. (2.71), the correlation coefficient between $\ln(X_1)$ and $\ln(X_2)$ is

$$\rho' = \frac{\ln[1 + (0.6)(0.3)(0.4)]}{(0.294)(0.385)} = 0.623$$

Then

$$P(X_1 \le 13, X_2 \le 3 \mid \rho = 0.6) = P(Z_1 \le 1.04, Z_2 \le -1.13 \mid \rho' = 0.623)$$

$$= \Phi(a = 1.04, b = -1.13 \mid \rho' = 0.623)$$

From this point forward, the procedure for determining $\Phi(a = 1.04, b = -1.13 \mid \rho' = 0.623)$ is exactly identical to that of Example 2.21. The result from using Eq. (2.121) is 0.1285.

Problems

2.1   Referring to Example 2.4, solve the following problems:
   (a)   Assume that $P(E_1|E_2) = 1.0$ and $P(E_2|E_1) = 0.8$. What is the probability that the flow-carrying capacity of the sewer main is exceeded?
   (b)   If the flow capacity of the downstream sewer main is twice that of its two upstream branches, what is the probability that the flow capacity of the downstream sewer main is exceeded? Assume that if only branch 1 or branch 2 exceeds its corresponding capacity, the probability of flow in the sewer main exceeding its capacity is 0.15.
   (c)   Under the condition of (b), it is observed that surcharge occurred in the downstream sewer main. Determine the probabilities that (i) only branch 1 exceeds its capacity, (ii) only branch 2 is surcharged, and (iii) none of the sewer branches exceed their capacities.

2.2   Referring to Example 2.5, it is observed that surcharge occurred in the downstream sewer main. Determine the probabilities that (a) only branch 1 exceeds its flow-carrying capacity, (b) only branch 2 is surcharged, and (c) none of the sewer branches exceed their capacities.

2.3 A detention basin is designed to accommodate excessive surface runoff temporarily during storm events. The detention basin should not overflow, if possible, to prevent potential pollution of streams or other receiving water bodies.

For simplicity, the amount of daily rainfall is categorized as heavy, moderate, and light (including none). With the present storage capacity, the detention basin is capable of accommodating runoff generated by two consecutive days of heavy rainfall or three consecutive days of moderate rainfall. The daily rainfall amounts around the detention basin site are not entirely independent. In other words, the amount of rainfall on a given day would affect the rainfall amount on the next day.

Let random variable X_t represent the amount of rainfall in any day t. The *transition probability matrix* indicating the conditional probability of the rainfall amount in a given day t, conditioned on the rainfall amount of the previous day, is shown in the following table.

	$X_{t+1} =$		
	H	M	L
H	0.3	0.5	0.2
$X_t = M$	0.3	0.4	0.3
L	0.1	0.3	0.6

(a) For a given day, the amount of rainfall is light. What is the probability that the detention basin will overflow in the next three days? (After Mays and Tung, 1992.)

(b) Compute the probability that the detention basin will overflow in the next three days. Assume that at any given day of the month the probabilities for having the various rainfall amounts are $P(H) = 0.1$, $P(M) = 0.3$, $P(L) = 0.6$.

2.4 Before a section of concrete pipe of a special order can be accepted for installation in a culvert project, the thickness of the pipe needs to be inspected by state highway department personnel for specification compliance using ultrasonic reading. For this project, the required thickness of the concrete pipe wall must be at least 3 in. The inspection is done by arbitrarily selecting a point on the pipe surface and measuring the thickness at that point. The pipe is accepted if the thickness from the ultrasonic reading exceeds 3 in; otherwise, the entire section of the pipe is rejected. Suppose, from past experience, that 90 percent of all pipe sections manufactured by the factory were found to be in compliance with specifications. However, the ultrasonic thickness determination is only 80 percent reliable.

(a) What is the probability that a particular pipe section is well manufactured and will be accepted by the highway department?

(b) What is the probability that a pipe section is poorly constructed but will be accepted on the basis of ultrasonic test?

2.5 A quality-control inspector is testing the sample output from a manufacturing process for concrete pipes for a storm sewer project, wherein 95 percent of the items are satisfactory. Three pipes are chosen randomly for inspection. The successive

quality evaluations may be considered as independent. What is the probability that (a) none of the three pipes inspected are satisfactory and (b) exactly two are satisfactory?

2.6 Derive the PDF for a random variable having a triangular distribution with the lower bound a, mode m, and the upper bound b, as shown in Fig. 2P.1.

2.7 Show that $F_1(x_1) + F_2(x_2) - 1 \leq F_{1,2}(x_1, x_2) \leq \min[F_1(x_1), F_2(x_2)]$

2.8 The Farlie-Gumbel-Morgenstern bivariate uniform distribution has the following joint CDF (Hutchinson and Lai, 1990):

$$F_{x,y}(x, y) = xy[1 + \theta(1 - x)(1 - y)] \qquad \text{for } 0 \leq x, y \leq 1$$

with $-1 \leq \theta \leq 1$. Do the following exercises: (a) derive the joint PDF, (b) obtain the marginal CDF and PDF of X and Y, and (c) derive the conditional PDFs $f_x(x|y)$ and $f_y(y|x)$.

2.9 Refer to Problem 2.8. Compute (a) $P(X \leq 0.5, Y \leq 0.5)$, (b) $P(X \geq 0.5, Y \geq 0.5)$, and (c) $P(X \geq 0.5 \,|\, Y = 0.5)$.

2.10 Apply Eq. (2.22) to show that the first four central moments in terms of moments about the origin are

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_x^2$$

$$\mu_3 = \mu_3' - 3\mu_x\mu_2' + 2\mu_x^3$$

$$\mu_4 = \mu_4' - 4\mu_x\mu_3' + 6\mu_x^2\mu_2' - 3\mu_x^4$$
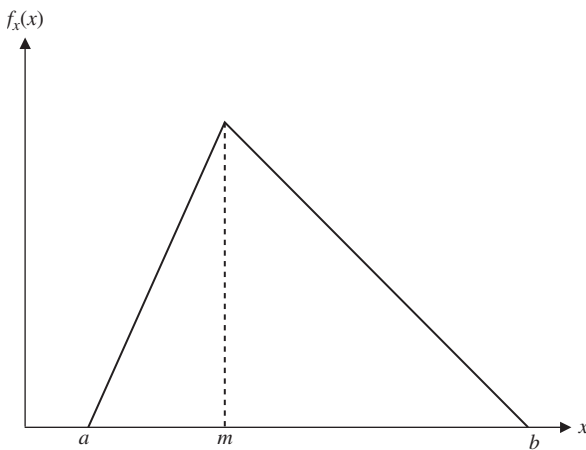
Figure 2P.1 Triangular distribution.

2.11 Apply Eq. (2.23) to show that the first four moments about the origin could be expressed in terms of the first four central moments as

$$\mu_1' = \mu_x$$

$$\mu_2' = \mu_2 + \mu_x^2$$

$$\mu_3' = \mu_3 + 3\mu_x\mu_2 + \mu_x^3$$

$$\mu_4' = \mu_4 + 4\mu_x\mu_3 + 6\mu_x^2\mu_2 + \mu_x^4$$

2.12 Based on definitions of α- and β-moments, i.e., Eqs. (2.26a) and (2.26b), (a) derive the general expressions between the two moments, and (b) write out explicitly their relations for $r = 0, 1, 2$, and 3.

2.13 Refer to Example 2.9. Continue to derive the expressions for the third and fourth L-moments of the exponential distribution.

2.14 A company plans to build a production factory by a river. You are hired by the company as a consultant to analyze the flood risk of the factory site. It is known that the magnitude of an annual flood has a lognormal distribution with a mean of 30,000 ft^3/s and standard deviation 25,000 ft^3/s. It is also known from a field investigation that the stage-discharge relationship for the channel reach is $Q = 1500H^{1.4}$, where Q is flow rate (in ft^3/s) and H is water surface elevation (in feet) above a given datum. The elevation of a tentative location for the factory is 15 ft above the datum (after Mays and Tung, 1992). (a) What is the annual risk that the factory site will be flooded? (b) At this plant site, it is also known that the flood-damage function can be approximated as

$$\text{Damage (in \$1000)} = \begin{cases} 0 & \text{if } H \leq 15 \text{ ft} \\ 40(\ln H + 8)(\ln H - 2.7) & \text{if } H > 15 \text{ ft} \end{cases}$$

What is the annual expected flood damage? (Use the appropriate numerical approximation technique for calculations.)

2.15 Referring to Problem 2.6, assume that Manning's roughness coefficient has a triangular distribution as shown in Fig. 2P.1. (a) Derive the expression for the mean and variance of Manning's roughness. (b) Show that (i) for a symmetric triangular distribution, $\sigma = (b - m)/\sqrt{6}$ and (ii) when the mode is at the lower or upper bound, $\sigma = (b - a)/3\sqrt{2}$.

2.16 Suppose that a random variable X has a uniform distribution (Fig. 2P.2), with a and b being its lower and upper bounds, respectively. Show that (a) $E(X) = \mu_x = (b + a)/2$, (b) $\text{Var}(X) = (b - a)^2/12$, and (c) $\Omega_x = (1 - a/\mu_x)/\sqrt{3}$.

2.17 Referring to the uniform distribution as shown in Fig. 2P.2, (a) derive the expression for the first two probability-weighted moments, and (b) derive the expressions for the L-coefficient of variation.

2.18 Refer to Example 2.8. Based on the conditional PDF obtained in part (c), derive the conditional expectation $E(Y | x)$, and the conditional variance $\text{Var}(Y | x)$.
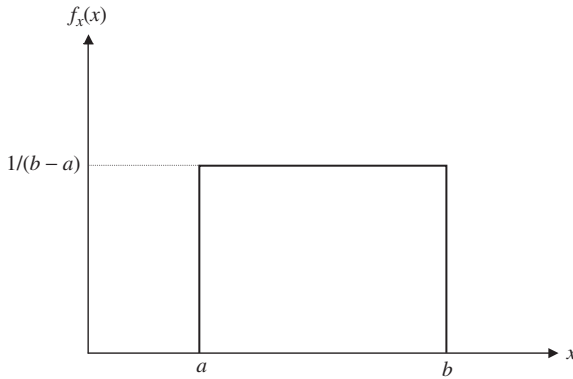
Figure 2P.2 Uniform distribution.

Furthermore, plot the conditional expectation and conditional standard deviation of Y on x with respect to x.

2.19 Consider two random variables X and Y having the joint PDF of the following form:

$$f_{x,y}(x, y) = c \left(5 - \frac{y}{2} + x^2 \right) \qquad \text{for } 0 \leq x, y \leq 2$$

(a) Determine the coefficient c. (b) Derive the joint CDF. (c) Find $f_x(x)$ and $f_y(y)$. (d) Determine the mean and variance of X and Y. (e) Compute the correlation coefficient between X and Y.

2.20 Consider the following hydrologic model in which the runoff Q is related to the rainfall R by

$$Q = a + bR$$

if $a > 0$ and $b > 0$ are model coefficients. Ignoring uncertainties of model coefficients, show that $\text{Corr}(Q, R) = 1.0$.

2.21 Suppose that the rainfall-runoff model in Problem 2.4.11 has a model error, and it can expressed as

$$Q = a + bR + \varepsilon$$

in which ε is the model error term, which has a zero mean and standard deviation of σ_ε. Furthermore, the model error ε is independent of the random rainfall R. Derive the expression for $\text{Corr}(Q, R)$.

2.22 Let $X = X_1 + X_3$ and $Y = X_2 + X_3$. Find $\text{Corr}(X, Y)$, assuming that X_1, X_2, and X_3 are statistically independent.

2.23 Consider two random variables Y_1 and Y_2 that each, individually, is a linear function of two other random variables X_1 and X_2 as follows:

$$Y_1 = a_{11}X_1 + a_{12}X_2 \qquad Y_2 = a_{21}X_1 + a_{22}X_2$$

It is known that the mean and standard deviations of random variable X_k are μ_k and σ_k, respectively, for $k = 1, 2$. (a) Derive the expression for the correlation

coefficient between Y_1 and Y_2 under the condition that X_1 and X_2 are statistically independent. (b) Derive the expression for the correlation coefficient between Y_1 and Y_2 under the condition that X_1 and X_2 are correlated with a correlation coefficient ρ.

2.24 As a generalization to Problem 2.23, consider M random variables $Y_1, Y_2, \ldots, Y_M$ that are linear functions of K other random variables $X_1, X_2, \ldots, X_K$ in a vector form as follows:

$$Y_m = \boldsymbol{a}_m^t \boldsymbol{X} \qquad \text{for } m = 1, 2, \ldots, M$$

in which $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$, a column vector of K random variables Xs and $\boldsymbol{a}_m^t = (a_{m1}, a_{m2}, \ldots, a_{mK})$, a row vector of coefficients for the random variable Y_m. In matrix form, the preceding system of linear equations can be written as $\boldsymbol{Y} = \boldsymbol{A}^t \boldsymbol{X}$. Given that the mean and standard deviations of the random variable X_k are μ_k and σ_k, respectively, for $k = 1, 2, \ldots, K$, (a) derive the expression for the correlation matrix between Ys assuming that the random variable Xs are statistically independent, and (b) derive the expression for the correlation coefficient between Ys under the condition that the random variable Xs are correlated with a correlation matrix $\boldsymbol{R}_x$.

2.25 A coffer dam is to be built for the construction of bridge piers in a river. In an economic analysis of the situation, it is decided to have the dam height designed to withstand floods up to 5000 ft^3/s. From flood frequency analysis it is estimated that the annual maximum flood discharge has a Gumbel distribution with the mean of 2500 ft^3/s and coefficient of variation of 0.25. (a) Determine the risk of flood water overtopping the coffer dam during a 3-year construction period. (b) If the risk is considered too high and is to be reduced by half, what should be the design flood magnitude?

2.26 Recompute the probability of Problem 2.25 by using the Poisson distribution.

2.27 There are five identical pumps at a pumping station. The PDFs of the time to failure of each pump are the same with an exponential distribution as Example 2.6, that is,

$$f(t) = 0.0008 \exp(-0.0008t) \qquad \text{for } t \geq 0$$

The operation of each individual pump is assumed to be independent. The system requires at least two pumps to be in operation so as to deliver the required amount of water. Assuming that all five pumps are functioning, determine the reliability of the pump station being able to deliver the required amount of water over a 200-h period.

2.28 Referring to Example 2.14, determine the probability, by both binomial and Poisson distributions, that there would be more than five overtopping events over a period of 100 years. Compare the results with that using the normal approximation.

2.29 From a long experience of observing precipitation at a gauging station, it is found that the probability of a rainy day is 0.30. What is the probability that the next year would have at least 150 rainy days by looking up the normal probability table?

2.30 The well-known *Thiem equation* can be used to compute the drawdown in a confined and homogeneous aquifer as

$$s_{ik} = \frac{\ln(r_{ok}/r_{ik})}{2\pi T} Q_k = \xi_{ik} Q_k$$

in which s_{ik} is drawdown at the ith observation location resulting from a pumpage of Q_k at the kth production well, r_{ok} is the radius of influence of the kth production well, r_{ik} is the distance between the ith observation point and the kth production well, and T is the transmissivity of the aquifer. The overall effect of the aquifer drawdown at the ith observation point, when more than one production well is in operation, can be obtained, by the principle of linear superposition, as the sum of the responses caused by all production wells in the field, that is,

$$s_i = \sum_{k=1}^{K} s_{ik} = \sum_{k=1}^{K} \xi_{ik} Q_k$$

where K is the total number of production wells in operation. Consider a system consisting of two production wells and one observation well. The locations of the three wells, the pumping rates of the two production wells, and their zones of influence are shown in Fig. 2P.3. It is assumed that the transmissivity of the aquifer has a lognormal distribution with the mean $\mu_T = 4000$ gallons per day per foot (gpd/ft) and standard deviation $\sigma_T = 2000$ gpd/ft (after Mays and Tung, 1992). (a) Prove that the total drawdown in the aquifer field also is lognormally distributed. (b) Compute the exact values of the mean and variance of the total drawdown at the observation point when $Q_1 = 10,000$ gpd and $Q_2 = 15,000$ gpd. (c) Compute the probability that the resulting drawdown at the observation point does not exceed 2 ft. (d) If the maximum allowable probability of the total drawdown exceeding 2 ft is 0.10, find out the maximum allowable total pumpage from the two production wells.

2.31 A frequently used surface pollutant washoff model is based on a first-order decay function (Sartor and Boyd, 1972):

$$M_t = M_0 e^{-cRt}$$

where M_0 is the initial pollutant mass at time $t = 0$, R is runoff intensity (mm/h), c is the washoff coefficient (mm^{-1}), M_t is the mass of the pollutant remaining

Production well 1   Production well 2

Observation well

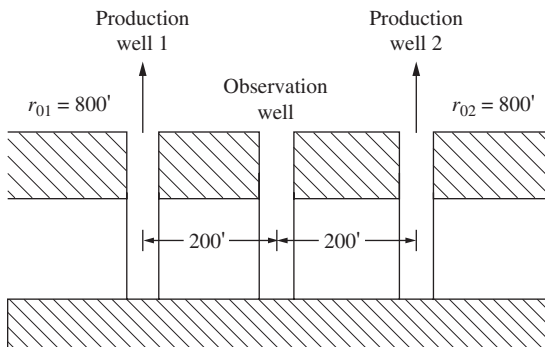$r_{01} = 800'$ $r_{02} = 800'$

200'   200'

Figure 2P.3 Locations of production and observation wells.

on the street surface (kg), and t is time elapsed (in hours) since the beginning of the storm. This model does not consider pollutant buildup and is generally appropriate for the within-storm event analysis. Suppose that $M_0 = 10,000$ kg and $c = 1.84$/cm. The runoff intensity R is a normal random variable with the mean of 10 cm/h and a coefficient of variation of 0.3. Determine the time t such that $P(M_t/M_0 < 0.05) = 0.90$.

2.32 Consider n independent random samples $X_1, X_2, \ldots, X_n$ from an identical distribution with the mean μ_x and variance σ_x^2. Show that the sample mean $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ has the following properties:

$$E(\overline{X}_n) = \mu_x \qquad \text{and} \qquad \text{Var}(\overline{X}_n) = \frac{\sigma_x^2}{n}$$

What would be the sampling distribution of $\overline{X}_n$ if random samples are normally distributed?

2.33 Consider that measured hydrologic quantity Y and its indicator for accuracy S are related to the unknown true quantity X as $Y = SX$. Assume that $X \sim LN(\mu_x, \sigma_x)$, $S \sim LN(\mu_s = 1, \sigma_s)$, and X is independent of S. (a) What is the distribution function for Y? Derive the expressions for the mean and coefficient of variation of Y, that is, μ_y and Ω_y, in terms of those of X and S. (b) Derive the expression for $r_p = y_p/x_p$ with $P(Y \le y_p) = P(X \le x_p) = p$ and plot r_p versus p. (c) Define measurement error as $\varepsilon = Y - X$. Determine the minimum reliability of the measurement so that the corresponding relative absolute error $|\varepsilon/X|$ does not exceed the require precision of 5 percent.

2.34 Consider that measured discharge Q' is subject to measurement error ε and that both are related to the true but unknown discharge Q as (Cong and Xu, 1987)

$$Q' = Q + \varepsilon$$

It is common to assume that (i) $E(\varepsilon \,|\, q) = 0$, (ii) $\text{Var}(\varepsilon \,|\, q) = [\alpha(q)q]^2$, and (iii) random error ε is normally distributed, that is, $\varepsilon \,|\, q \sim N(\mu_{\varepsilon \,|\, q} = 0, \sigma_{\varepsilon \,|\, q})$. (a) Show that $E(Q' \,|\, q) = q$, $E[(Q'/Q) - 1 \,|\, q] = 0$, and $\text{Var}[(Q'/Q - 1)^2 \,|\, q] = \alpha^2(q)$. (b) Under $\alpha(q) = \alpha$, show that $E(Q') = E(Q)$, $\text{Var}(\varepsilon) = \alpha^2 E(Q^2)$, and $\text{Var}(Q') = (1 + \alpha^2)\text{Var}(Q) + \alpha^2 E^2(Q)$. (c) Suppose that it is required that 75 percent of measurements have relative errors in the range of ± 5 percent (precision level). Determine the corresponding value of $\alpha(q)$ assuming that the measurement error is normally distributed.

2.35 Show that the valid range of the correlation coefficient obtained in Example 2.20 is correct also for the general case of exponential random variables with parameters β_1 and β_2 of the form of Eq. (2.79).

2.36 Referring to Example 2.20, derive the range of the correlation coefficient for a bivariate exponential distribution using Farlie's formula (Eq. 2.107).

2.37 The Pareto distribution is used frequently in economic analysis to describe the randomness of benefit, cost, and income. Consider two correlated Pareto random

variables, each of which has the following marginal PDFs:

$$f_k(x_k) = \frac{a\theta_k^a}{x_k^{a+1}} \quad x_k > \theta_k > 0 \quad a > 0, \qquad \text{for } k = 1, 2$$

Derive the joint PDF and joint CDF by Morgenstern's formula. Furthermore, derive the expression for $E(X_1|x_2)$ and the correlation coefficient between X_1 and X_2.

2.38 Repeat Problem 2.37 using Farlie's formula.

2.39 Analyzing the stream flow data from several flood events, it is found that the flood peak discharge Q and the corresponding volume V have the following relationship:

$$\ln(V) = a + b \times \ln(Q) + \varepsilon$$

in which a and b are constants, and ε is the model error term. Suppose that the model error term ε has a normal distribution with mean 0 and standard deviation σ_ε. Then show that the conditional PDF of $V | Q$, $h(v|q)$, is a lognormal distribution. Furthermore, suppose that the peak discharge is a lognormal random variable. Show that the joint PDF of V and Q is bivariate lognormal.

2.40 Analyzing the stream flow data from 105 flood events at different locations in Wyoming, Wahl and Rankl (1993) found that the flood peak discharge Q (in ft^3/s) and the corresponding volume V (in acre-feet, AF) have the following relationship:

$$\ln(V) = \ln(0.0655) + 1.011 \times \ln(Q) + \varepsilon$$

in which ε is the model error term with the assumed $\sigma_\varepsilon = 0.3$. A flood frequency analysis of the North Platte River near Walden, Colorado, indicated that the annual maximum flood has a lognormal distribution with mean $\mu_Q = 1380$ ft^3/s and $\sigma_Q = 440$ ft^3/s. (a) Derive the joint PDF of V and Q for the annual maximum flood. (b) Determine the correlation coefficient between V and Q. (c) Compute $P(Q \geq 2000$ ft^3/s, $V \geq 180$ AF).

2.41 Let $X_2 = a_0 + a_1 Z_1 + a_2 Z_1^2$ and $X_2 = b_0 + b_1 Z_2 + b_2 Z_2^2$ in which Z_1 and Z_2 are bivariate standard normal random variables with a correlation coefficient ρ, that is, $\text{Corr}(Z_1, Z_2) = \rho$. Derive the expression for $\text{Corr}(X_1, X_2)$ in terms of polynomial coefficients and ρ.

2.42 Let X_1 and X_2 be bivariate lognormal random variables. Show that

$$\frac{\exp\left(-\sigma_{\ln x_1}\sigma_{\ln x_2}\right) - 1}{\sqrt{\exp\left(\sigma_{\ln x_1}^2\right) - 1}\sqrt{\exp\left(\sigma_{\ln x_2}^2\right) - 1}} \leq \text{Corr}(X_1, X_2)$$

$$\leq \frac{\exp\left(\sigma_{\ln x_1}\sigma_{\ln x_2}\right) - 1}{\sqrt{\exp\left(\sigma_{\ln x_1}^2\right) - 1}\sqrt{\exp\left(\sigma_{\ln x_2}^2\right) - 1}}$$

What does this inequality indicate?

2.43   Derive Eq. (2.71) from Eq. (2.135):

$$\text{Corr}(\ln X_1, \ln X_2) = \rho'_{12} = \frac{\ln(1 + \rho_{12}\Omega_1\Omega_2)}{\sqrt{\ln\left(1 + \Omega_1^2\right)}\sqrt{\ln\left(1 + \Omega_2^2\right)}}$$

where $\rho_{12} = \text{Corr}(X_1, X_2)$ and $\Omega_k =$ coefficient of variation of X_k, $k = 1, 2$.

2.44   Develop a computer program using Ditlevsen's expansion for estimating the multivariate normal probability.

2.45   Develop computer programs for multivariate normal probability bounds by Rackwitz's procedure and Ditlevsen's procedure, respectively.

References

Abramowitz, M., and Stegun, I.A., eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, 9th ed., Dover Publications, New York.

Ang, A. H. S., and Tang, W. H. (1975). *Probability Concepts in Engineering Planning and Design,* Vol. I: *Basic Principles*, John Wiley and Sons, New York.

Blank, L. (1980). *Statistical Procedures for Engineering, Management, and Science*, McGraw-Hill, New York.

Cong, S. Z., and Y. B. Xu (1987). "The effect of discharge measurement error in flood frequency analysis," *Journal of Hydrology*, 96:237–254.

Consul, P. C. (1989). *Generalized Poisson Distributions: Properties and Application*, Marcel Dekker, New York.

Consul, P. C., and Jain, G. C. (1973). A generalization of Poisson distribution: Properties and application, *Technometrics*, 15:791–799.

DeGroot, M. H. (1975). *Probability and Statistics*, Addison-Wesley, Reading, MA.

Der Kiureghian, A., and Liu, P. L. (1986). Structural reliability under incomplete probability distribution, *Journal of Engineering Mechanics*, ASCE, 112(1):85–104.

Devore, J. L. (1987). *Probability and Statistics for Engineering and Sciences*, 2d ed., Brooks/Cole Publishing, Monterey, CA.

Ditlevsen, O. (1979). Narrow reliability bounds for structural systems, *Journal of Structural Mechanics*, 7(4):435–451.

Ditlevsen, O. (1982). System reliability bounding by conditioning, *Journal of Engineering Mechanics*, ASCE, 108(5):708–718.

Ditlevsen, O. (1984). Taylor expansion of series system reliability, *Journal of Engineering Mechanics*, ASCE, 110(2):293–307.

Donnelly, T. G. (1973). Algorithm 462: Bivariate normal distribution, *Communications*, Association for Computing Machinery, 16:638.

Dowson, D. C., and Wragg, A. (1993). Maximum entropy distribution having prescribed first and second moments, *IEEE Transaction on Information Theory*, 19(9):689–693.

Drane, J. W., Cao, S., Wang, L., and Postelnicu, T. (1993). Limiting forms of probability mass function via recurrence formulas, *The American Statistician*, 47(4):269–274.

Dudewicz, E. (1976). *Introduction to Statistics and Probability*, Holt, Rinehart, and Winston, New York.

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution, *Biometrika*, 47:307–323.

Fisher, R. A., and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceding of the Cambridge Philosphical Society*, 24:180–190.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability-weighted moments: Definitions and relation to parameters of several distribution expressible in inverse form, *Water Resources Research*, 15(6):1049–1054.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire, *Annals of Mathematics*, 44:423–453.

Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

Haan, C. T. (1977). *Statistical Methods in Hydrology*, Iowa State University Press, Ames, IA.

Harr, M. E. (1987). *Reliability-Based Design in Civil Engineering*, McGraw-Hill, New York.

Hosking, J. R. M. (1986). The theory of probability-weighted moments, IBM Research Report, No. 12210, October.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of Royal Statistical Society, Series B*, 52(1):105–124.

Hosking, J. R. M. (1991). Approximations for use in constructing L-moment ratio diagram, IBM Research Report No. 16635, T. J. Watson Research Center, Yorktown Heights, NY.

Hutchinson, T. P., and Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasizing Applications*, Rumsby Scientific Publishing, Adelaide, South Australia.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*, John Wiley and Sons, New York.

Johnson, N. L., and Kotz, S. (1972). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2. John Wiley and Sons, New York.

Johnson, N. L., and Kotz, S. (1976). *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, Inc., New York, NY.

Kite, G. W. (1988). *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, CO.

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

Leemis, L. M. (1986). Relationships among common univariate distributions, *The American Statistician*, 40(2):143–146.

Liu, P. L., and Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances, *Probabilistic Engineering Mechanics*, 1(2):105–112.

Mardia, K. V. (1970a). *Families of Bivariate Distributions*, Griffin, London.

Mardia, K. V. (1970b). A translation family of bivariate distributions and Frechet's bounds, *Sankhya, Series A*, 32:119–121.

Mays, L. W., and Tung, Y. K. (1992). *Hydrosystems Engineering and Management*, McGraw-Hill, New York.

Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen, *Mitteilingsblatt fur Mathematische Statistik*, 8:234–235.

Nataf, A. (1962). Determination des distributions dont les marges sont donnees, *Comptes Rendus de l'Academie des Sciences, Paris*, 225:42–43.

Rackwitz, R. (1978). Close bounds for the reliability of structural sytems, *Berichte zur Zuverlassigkeitstheorie der Bauwerke*, SFB 96, Heft 29/1978, LKI, Technische Universitat Munchen.

Rao, A. R., and Hamed, K. H. (2000). *Flood Frequency Analysis*. CRC Press, Boca Raton, FL.

Royston, P. (1992). Which measures of skewness and kurtosis are best? *Statistics in Medicine*, 11:333–343.

Ruben, H. (1964). An asymptotic expansion for the multivariate normal distribution and Mill's ratio, *Journal of Research National Bureau of Standards*, 68B(1):3–11.

Sartor, J. D., and Boyd, G. B. (1972). Water pollution aspects of street surface contaminants. Report No. EPA-R2-72-081, U.S. Environmental Protection Agency, Washington.

Shen, H. W., and Bryson, M. (1979). Impact of extremal distributions on analysis of maximum loading, in *Reliability in Water Resources Engineering*, ed. by E. A. McBean, K. W. Hipel, and T. E. Unny, Water Resources Publications, Littleton, CO.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgoiu, E. (1993). Frequency analysis of extreme events, in *Handbook of Hydrology*, ed. by D. R. Maidment, McGraw-Hill, New York.

Stuart, A., and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1: *Distribution Theory*, 5th ed., Oxford University Press, New York.

Tietjen, G. (1994). Recursive schemes for calculating cumulative binomial and Poisson probabilities, *The American Statistician*, 48(2):136–137.

Tung, Y. K., and Yen, B. C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, McGraw-Hill, New York.

U.S. Water Resources Council (now called the Interagency Advisory Committee on Water Data) (1982). *Guideline in Determining Flood Flow Frequency*, Bulletin 17B; available from Officie of Water Data Coordination, U.S. Geological Survey, Reston, VA.

Vale, C. D., and Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions, *Psychometrika*, 48(3):65–471.

Vedder, J. D. (1995). An invertible approximation to the normal distribution function, *Computational Statistics and Data Analysis*, 16(2):119–123.

Wahl, K. L., and Rankl, J. G. (1993). A potential problem with mean dimensionless hydrographs at ungaged sites, in *Proceedings*, International Symposium on Engineering Hydrology, San Francisco, CA, July 26–30.

Hydrologic Frequency Analysis

One of the basic questions in many hydrosystems infrastructural designs that an engineer must answer is, "What should be the capacity or size of a system?" The planning goal is not to eliminate all hydro-hazards but to reduce the frequency of their occurrences and thus the resulting damage. If such planning is to be correct, the probabilities of flooding must be evaluated correctly. The problem is made more complex because in many cases the "input" is controlled by nature rather than by humans. For example, variations in the amount, timing, and spatial distribution of precipitation are the underlying reasons for the need for probabilistic approaches for many civil and environmental engineering projects. Our understanding and ability to predict precipitation and its resulting effects such as runoff are far from perfect. How, then, can an engineer approach the problem of design when he or she cannot be certain of the hydrologic load that will be placed on the infrastructure under consideration?

An approach that is used often is a statistical or probabilistic one. Such an approach does not require a complete understanding of the hydrologic phenomenon involved but examines the relationship between magnitude and frequency of occurrence in the hope of finding some statistical regularity between these variables. In effect, the past is extrapolated into the future. This assumes that whatever complex physical interactions control nature, the process does not change with time, and so the historical record can be used as a basis for estimating future events. In other words, the data are assumed to satisfy *statistical stationarity* by which the underlying distributional properties do not change with time, and the historical data series is representative of the storms and watershed conditions to be experienced in the future. An example that violates this statistical stationarity is the progressive urbanization within a watershed that could result in a tendency of increasing peak flow over time.

The hydrologic data most commonly analyzed in this way are rainfall and stream flow records. Frequency analysis was first used for the study of stream flow records by Herschel and Freeman during the period from 1880 to 1890

(Foster, 1935). The first comprehensive study was performed by Fuller (1914). Gumbel (1941, 1942) first applied a particular extreme-value probability distribution to flood flows, whereas Chow (1954) extended the work using this distribution. A significant contribution to the study of rainfall frequencies was made by Yarnell (1936). The study analyzed rainfall durations lasting from 5 minutes to 24 hours and determined their frequency of occurrence at different locations within the continental United States. A similar study was performed by the Miami Conservancy District of Ohio for durations extending from 1 to 6 days (Engineering Staff of Miami Conservancy District, 1937). An extremal probability distribution was applied to rainfall data at Chicago, Illinois, by Chow (1953), and more recent frequency analysis of rainfall data was performed by the U.S. National Weather Service (Hershfield, 1964; U.S. Weather Bureau, 1964; Miller et al., 1973; Frederick et al., 1977, Huff and Angel, 1989, 1992). Low stream flows and droughts also were studied statistically by Gumbel (1954, 1963), who applied an extremal distribution to model the occurrences of drought frequencies. In the United Kingdom, hydrologic frequency analysis usually follows the procedures described in the *Flood Studies Report* of 1975 (National Environment Research Council, 1975). In general, frequency analysis is a useful analytical tool for studying randomly occurring events and need not be limited to hydrologic studies. Frequency analysis also has been applied to water quality studies and to ocean wave studies.

Basic probability concepts and theories useful for frequency analysis are described in Chap. 2. In general, there is no physical rule that requires the use of a particular distribution in the frequency analysis of geophysical data. However, since the maximum or minimum values of geophysical events are usually of interest, extreme-value-related distributions have been found to be most useful.

3.1 Types of Geophysical Data Series

The first step in the frequency-analysis process is to identify the set of data or sample to be studied. The sample is called a *data series* because many events of interest occur in a time sequence, and time is a useful frame of reference. The events are continuous, and thus their complete description as a function of time would constitute an infinite number of data points. To overcome this, it is customary to divide the events into a series of finite time increments and consider the average magnitude or instantaneous values of the largest or smallest within each interval. In frequency analysis, geophysical events that make up the data series generally are assumed to be statistically independent in time. In the United States, the water year concept was developed to facilitate the independence of hydrologic flood series. Throughout the eastern, southern, and Pacific western areas of the United States, the season of lowest stream flow is late summer and fall (August–October) (U.S. Department of Agriculture, 1955). Thus, by establishing the water year as October 1 to September 30, the chance of having related floods in each year is minimized, and the assumption of independence in the flood data is supported. In case time dependence is present

in the data series and should be accounted for, procedures developed in time series analysis (Salas et al., 1980; Salas, 1993) should be applied. This means that the events themselves first must be identified in terms of a beginning and an end and then sampled using some criterion. Usually only one value from each event is included in the data series. There are three basic types of data series extractable from geophysical events:

1. A *complete series*, which includes all the available data on the magnitude of a phenomenon. A complete data series is used most frequently for flow-duration studies to determine the amount of firm power available in a proposed hydropower project or to study the low-flow behavior in water quality management. Such a data series is usually very large, and since in some instances engineers are only interested in the extremes of the distribution (e.g., floods, droughts, wind speeds, and wave heights), other data series often are more practical. For geophysical events, data in a complete series often exhibit significant time dependence, which makes the frequency-analysis procedure described herein inappropriate.

2. An *extreme-value series* is one that contains the largest (or smallest) data value for each of a number of equal time intervals. If, for example, the largest data value in each year of record is used, the extreme-value series is called an *annual maximum series*. If the smallest value is used, the series is called an *annual minimum series*.

3. A *partial-duration series* consists of all data above or below a base value. For example, one might consider only floods in a river with a magnitude greater than 1,000 m^3/s. When the base value is selected so that the number of events included in the data series equals the number of years of record, the resulting series is called an *annual exceedance series*. This series contains the *n* largest or *n* smallest values in *n* years of record.

The selection of geophysical data series is illustrated in Fig. 3.1. Figure 3.1*a* represents the original data; the length of each line indicates the magnitude of the event. Figure 3.1*b* shows an annual maximum series with the largest data value in each year being retained for analysis. Figure 3.1*c* shows the data values that would be included in an annual exceedance series. Since there are 15 years of record, the 15 largest data values are retained. Figure 3.1*d* and *e* illustrate for comparison the rank in descending order of the magnitude of the events in each of the two series. As shown in Fig. 3.1*d* and *e* the annual maximum series and the annual exceedance series form different probability distributions, but when used to estimate extreme floods with return periods of 10 years or more, the differences between the results from the two series are minimal, and the annual maximum series is the one used most commonly. Thus this chapter focuses on the annual maximum series in the following discussion and examples.

Another issue related to the selection of the data series for frequency analysis is the adequacy of the record length. Benson (1952) generated randomly
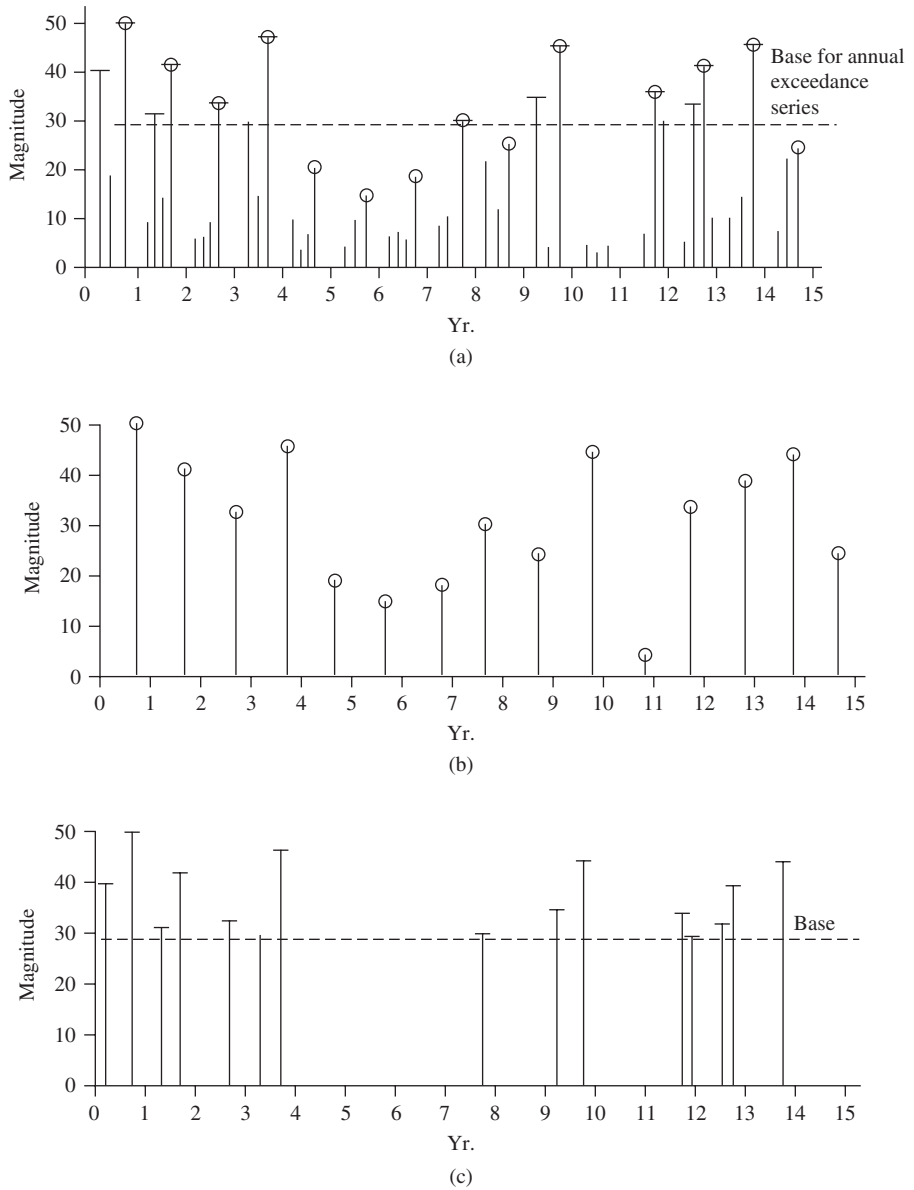
Figure 3.1   Different types of data series: (*a*) original data series; (*b*) annual maximum series; (*c*) annual exceedance series; (*d*) ranked complete series; (*e*) ranked annual maximum and exceedance series (Chow, 1964).

(d)

(e)
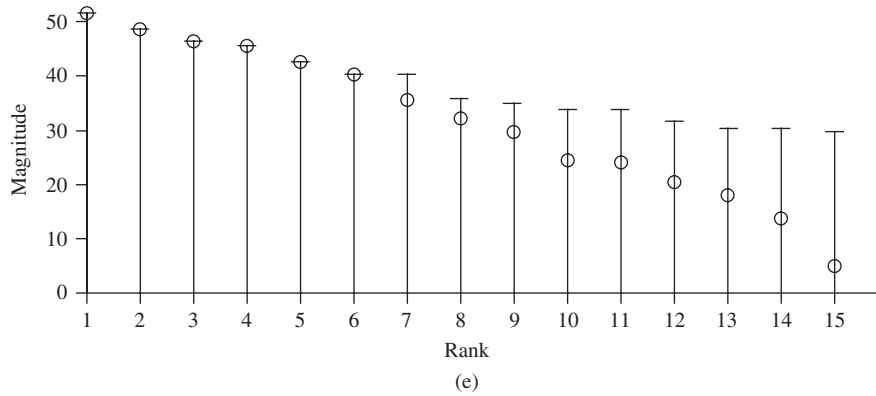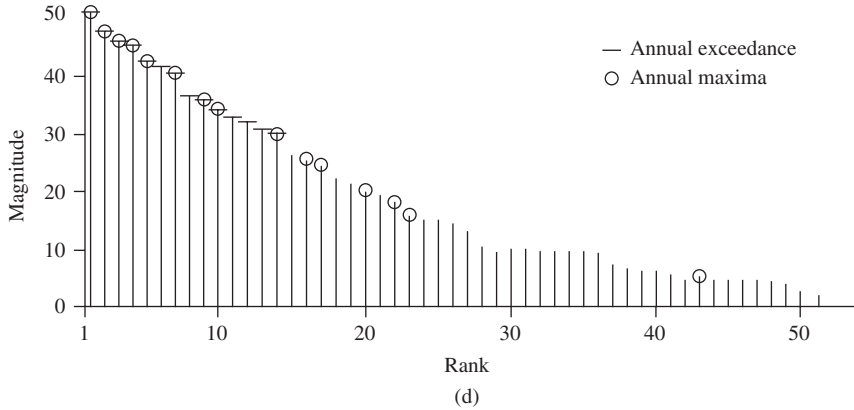
Figure 3.1 (*Continued*)

selected values from known probability distributions and determined the record
length necessary to estimate various probability events with acceptable error
levels of 10 and 25 percent. Benson's results are listed in Table 3.1. Linsley
et al. (1982, p. 358) reported that similar simulation-based studies at Stanford
University found that 80 percent of the estimates of the 100-year flood based
on 20 years of record were too high and that 45 percent of the overestimates

**TABLE 3.1   Number of Years of Record Needed to Obtain Estimates of Specified
Design Probability Events with Acceptable Errors of 10 and 25 Percent**

Design probability	Return period (years)	10% error (years)	25% error (years)
0.1	10	90	18
0.02	50	110	39
0.01	100	115	48

SOURCE: After Benson (1952).

exceeded 30 percent. The U.S. Water Resources Council (1967) recommended that at least 10 years of data should be available before a frequency analysis can be done. However, the results described in this section indicate that if a frequency analysis is done using 10 years of record, a high degree of uncertainty can be expected in the estimate of high-return-period events.

The final issue with respect to the data series used for frequency analysis is related to the problem of data homogeneity. For low-magnitude floods, peak stage is recorded at the gauge, and the discharge is determined from a rating curve established by current meter measurements of flows including similar-magnitude floods. In this case, the standard error of the measurement usually is less than 10 percent of the estimated discharge. For high-magnitude floods, peak stage often is inferred from high-water marks, and the discharge is computed by indirect means. For indirectly determined discharges, the standard error probably is several times larger, on the order of 16 to 30 percent (Potter and Walker, 1981). This is known as the *discontinuous measurement error* (DME) problem. Potter and Walker (1981) demonstrated that, as a result of DME, the probability distribution of measured floods can be greatly distorted with respect to the parent population. This further contributes to the uncertainty in flood frequency analysis.

3.2   Return Period

Hydrosystems engineers have been using the concept of the *return period* (or sometimes called *recurrence interval*) as a substitute for probability because it gives some physical interpretation to the probability. The return period for a given event is defined as the period of time on the *long-term average* at which a given event is equaled or exceeded. Hence, *on average*, an event with a 2-year return period will be equaled or exceeded once in 2 years. The relationship between the probability and return period is given by

$$T = \frac{1}{P(X \geq x_T)} = \frac{1}{1 - P(X < x_T)} \tag{3.1}$$

in which x_T is the value of the variate corresponding to a T-year return period. For example, if the probability that a flood will be equaled or exceeded in a single year is 0.1, that is, $P(X \geq x_T) = 0.1$, the corresponding return period is $1/P(X \geq x_T) = 1/0.1 = 10$ years. Note that $P(X \geq x_T)$ must be the probability that the event is equaled or exceeded in *any one* year and is the same for each year regardless of the magnitudes that occurred in prior years. This is so because the events are independent, and the long-term probabilities are used without regard to the order in which they may occur. A common error or misconception is to assume, for example, that if the 100-year event occurs this year, it will not occur again for the next 100 years. In fact, it could occur again next year and then not be repeated for several hundred years. This misconception resulted in considerable public complaints when the Phoenix area experienced two 50-year and one 100-year floods in a span of 18 months in 1978–1979 and the Milwaukee area experienced 100-year floods in June 1997 and June 1998.

Hence it is more appropriate and less confusing to use the *odds ratio*; e.g., the 100-year event can be described as the value having 1-in-100 chance being exceeded in any one year (Stedinger et al., 1993). In the United States in recent years it has become common practice to refer to the 100-year flood as the 1 percent chance exceedance flood, and similar percent chance exceedance descriptions are used for other flood magnitudes (U.S. Army Corps of Engineers, 1996).

The most common time unit for return period is the year, although semi-annual, monthly, or any other time period may be used. The time unit used to form the time series will be the unit assigned to the return period. Thus an annual series will have a return-period unit of years, and a monthly series will have return-period unit of months. However, one should be careful about compliance with the statistical independence assumption for the data series. Many geophysical data series exhibit serial correlation when the time interval is short, which can be dealt with properly only by time-series analysis procedures (Salas, 1993).

3.3   Probability Estimates for Data Series: Plotting Positions (Rank-order Probability)

As stated previously, the objective of frequency analysis is to fit geophysical data to a probability distribution so that a relationship between the event magnitude and its exceedance probability can be established. The first step in the procedure is to determine the type of data series (i.e., event magnitude) to be used. In order to fit a probability distribution to the data series, estimates of probability (or equivalent return period) must be assigned to each magnitude in the data series.

Consider a data series consisting of the *entire population* of N values for a particular variable. If this series were ranked according to decreasing magnitude, it could be stated that the probability of the largest variate being equaled or exceeded is $1/N$, where N is the total number of variates. Similarly, the exceedance probability of the second largest variate is $2/N$, and so forth. In general,

$$P\left(X \geq x_{(m)}\right) = \frac{1}{T_m} = \frac{m}{N} \tag{3.2}$$

in which m is the rank of the data in descending order, $x_{(m)}$ is the mth largest variate in a data series of size N, and T_m is the return period associated with $x_{(m)}$. In practice, the entire population is not used or available. However, the reasoning leading to Eq. (3.2) is still valid, except that the result is now only an estimate of the exceedance probability based on a sample. Equation (3.2), which shows the ranked-order probability, is called a *plotting position formula* because it provides an estimate of probability so that the data series can be plotted (magnitudes versus probability).

Equation (3.2) is appropriate for data series from the population. Some modifications are made to avoid theoretical inconsistency when it is applied to sample data series. For example, Eq. (3.2) yields an exceedance probability of 1.0 for the smallest variate, implying that all values must be equal or larger. Since only

a sample is used, there is a likelihood that at some future time an event with a lower value could occur. In application, if the lower values in the series are not of great interest, this weakness can be overlooked, and in fact, Eq. (3.2) is used in the analysis of the annual exceedance series. A number of plotting-position formulas have been introduced that can be expressed in a general form as

$$P\left(X \geq x_{(m)}\right) = u_m = \frac{1}{T_m} = \frac{m - a}{n + 1 - b} \tag{3.3}$$

in which $a \geq 0$ and $b \geq 0$ are constants, and n is the number of observations in the sample data series. Table 3.2 lists several plotting-position formulas that have been developed and used in frequency analysis. Perhaps the most popular plotting-position formula is the Weibull formula (with $a = 0$ and $b = 0$):

$$P\left(X \geq x_{(m)}\right) = u_m = \frac{1}{T_m} = \frac{m}{n + 1} \tag{3.4}$$

As shown in Table 3.2, it is noted that although these formulas give different results for the highest values in the series, they yield very similar results for the middle to lowest values, as seen in the last two columns.

Plotting-position formulas in the form of Eq. (3.3) can be categorized into being *probability-unbiased* and *quantile-unbiased*. The probability-unbiased

TABLE 3.2   Plotting-Position Formulas

Name	Formula $P(X \geq x_{(m)})$	$T_m = \dfrac{1}{P(X \geq x_{(m)})}$, for $n = 20$	
		$m = 1$	$m = 10$
California (1923)	$\dfrac{m}{n}$	20.0	2.00
Hazen (1930)	$\dfrac{m - 0.5}{n}$	40.0	2.11
Weibull (1939)	$\dfrac{m}{n + 1}$	41.0	2.10
Leivikov (1955)	$\dfrac{m - 0.3}{n + 0.4}$	29.1	2.10
Blom (1958)	$\dfrac{m - 0.375}{n + 0.25}$	24.5	2.10
Tukey (1962)	$\dfrac{m - 0.333}{n + 0.333}$	30.5	2.10
Gringorten (1963)	$\dfrac{m - 0.44}{n + 0.12}$	35.9	2.10
Cunnane (1978)	$\dfrac{m - 0.4}{n + 0.2}$	33.7	2.10
Hosking et al. (1985)	$\dfrac{m - 0.35}{n}$	30.7	2.07

plotting-position formula is concerned with finding a probability estimate $u_{(m)}$ for the exceedance probability of the mth largest observation such that $E[G(X_{(m)})] = u_{(m)}$, in which $G(X_{(m)}) = P(X \geq X_{(m)})$. In other words, the probability-unbiased plotting position yields the average exceedance probability for the mth largest observation in a sample of size n. If the data are independent random samples regardless of the underlying distribution, the estimator $U_{(m)} = G(X_{(m)})$ will have a beta distribution with the mean $E(U_{(m)}) = m/(n+1)$. Hence the Weibull plotting-position formula is probability-unbiased. On the other hand, Cunnane (1978) proposed quantile-unbiased plotting positions such that average value of the mth largest observation should be equal to $G^{-1}(u_{(m)})$, that is, $E(X_{(m)}) = G^{-1}(u_{(m)})$. The quantile-unbiased plotting-position formula, however, depends on the assumed distribution $G(\cdot)$. For example, referring to Table 3.2, the Blom plotting-position formula gives nearly unbiased quantiles for the normal distribution, and the Gringorton formula gives nearly unbiased quantiles for the Gumbel distribution. Cunnane's formula, however, produces nearly quantile-unbiased plotting positions for a range of distributions.

3.4 Graphic Approach

Once the data series is identified and ranked and the plotting position is calculated, a graph of magnitude x versus probability $[P(X \geq x), P(X < x),$ or $T]$ can be plotted and a distribution fitted graphically. To facilitate this procedure, it is common to use some specially designed probability graph paper rather than linear graph paper. The probability scale in those special papers is chosen such that the resulting probability plot is a straight line. By plotting the data using a particular probability scale and constructing a best-fit straight line through the data, a graphic fit is made to the distribution used in constructing the probability scale. This is a graphic approach to estimate the statistical parameters of the distribution.

Example 3.1 illustrates the graphic approach to the analysis of flood data. The general procedure is as follows:

1. Identify the sample data series to be used. If high-return-period values are of interest, either the annual maximum or exceedance series can be used. If low-return-period values are of interest, use an annual exceedance series.

2. Rank the data series in decreasing order, and compute exceedance probability or return period using the appropriate plotting-position formula.

3. Obtain the probability paper corresponding to the distribution one wishes to fit to the data series.

4. Plot the series, and draw a best-fit straight line through the data. An eyeball fit or a mathematical procedure, such as the least-squares method, can be used. Before doing the fit, make a judgment regarding whether or not to include the unusual observations that do not lie near the line (termed *outliers*).

5. Extend the line to the highest return-period value needed, and read all required return-period values off the line.

Example 3.1 The Boneyard Creek stream gauging station was located near the fire station on the campus of the University of Illinois at Urbana–Champaign. From the USGS Water Supply Papers, the partial duration data of peak discharges above 400 ft^3/s between the water years 1961 and 1975 were obtained and listed below. In addition, for the years when there was no flow in a year exceeding 400 ft^3/s, the peak flow for that year is given in parenthesis (e.g., 1961).

Year	Discharge, ft^3/s	Year	Discharge, ft^3/s
1961	(390)	1969	549, 454
1962	(374)	1970	414, 410
1963	(342)	1971	434, 524
1964	507	1972	505, 415, 406
1965	579, 406, 596	1973	428, 447, 407
1966	416	1974	468, 543, 441
1967	533	1975	591, 497
1968	505		

(a) List the ranked annual maximum series. Also compute and list the corresponding plotting positions (return period) and exceedance probability $P(X \geq x)$.

(b) Plot the annual maximum series on (i) Gumbel paper and (ii) lognormal paper.

(c) Construct a best-fit line through the nonlinear plots, and estimate the flows for return periods of 2, 10, 25, and 50 years.

Solution $n = 15$

(a)

Annual Maximum Discharge (ft^3/s)	Rank (m)	$T_m = \frac{n+1}{m}$ (years)	$P(X \geq x_{(m)})$ $= 1/T_m$	$P(X < x_{(m)})$ $= 1 - 1/T_m$
596	1	16.00	0.0625	0.9375
591	2	8.00	0.1250	0.8750
549	3	5.33	0.1875	0.8125
543	4	4.00	0.2500	0.7500
533	5	3.20	0.3125	0.6875
524	6	2.67	0.3750	0.6250
507	7	2.29	0.4375	0.5625
505	8	2.00	0.5000	0.5000
505	9	1.78	0.5625	0.4375
447	10	1.60	0.6250	0.3750
416	11	1.46	0.6875	0.3125
414	12	1.33	0.7500	0.2500
390	13	1.23	0.8125	0.1875
374	14	1.14	0.8750	0.1250
342	15	1.06	0.9375	0.0625

(b) Plots of the annual maximum flow series on the Gumbel and lognormal probability papers are shown in Fig. 3.2.
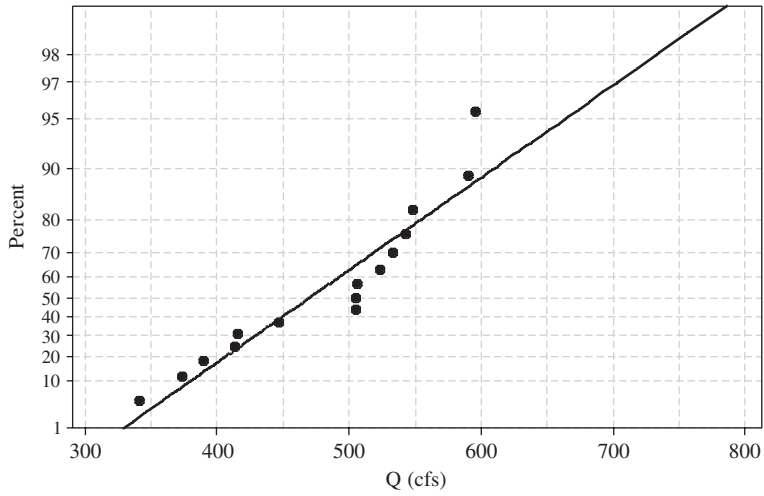
(a)
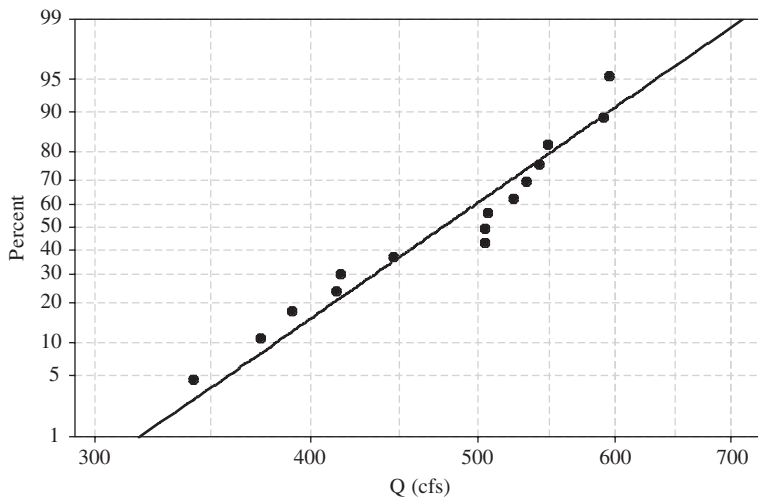
(b)

Figure 3.2 Probability plot for the annual maximum series for 1961–1975 on the Boneyard Creek at Urbana, IL: (*a*) Gumbel probability plot; (*b*) lognormal probability plot.

(c) The following table summarizes the results read from the plots:

	Return Period (years)			
Distribution	2	10	25	50
Gumbel	470	610	680	730
Lognormal	475	590	650	700

3.5   Analytical Approaches

An alternative to the graphic technique is to estimate the statistical parameters of a distribution from the sample data (refer to Sec. 3.6). Then the distribution model can be used to solve for the variate value corresponding to any desired return period or probability as

$$x_T = F_x^{-1}\left(1 - \frac{1}{T}\Big|\theta\right) \tag{3.5}$$

in which $F_x^{-1}(\theta)$ is the inverse cumulative distribution function with the model parameter vector θ. Equation (3.5) can be applied when the inverse distribution function forms are analytically amenable, such as for the Gumbel, generalized extreme value, generalized logistic, and generalized Pareto distributions (see Sec. 2.6.6).

Example 3.2 Consider that the annual maximum floods follow a lognormal distribution with a mean of 490 ft^3/s and a standard deviation of 80 ft^3/s. Determine the flood magnitude with a 1-in-100 chance of being exceeded in any given year.

Solution   From Eqs. (2.67a) and (2.67b), the parameters of a lognormal distribution, for annual maximum flood Q, can be obtained as

$$\sigma_{\ln Q} = \sqrt{\ln\left(\Omega_Q^2 + 1\right)} = \sqrt{\ln\left[\left(\frac{80}{490}\right)^2 + 1\right]} = 0.1622$$

$$\mu_{\ln Q} = \ln(\mu_Q) - \frac{1}{2}\sigma_{\ln Q}^2 = \ln(490) - \frac{1}{2}(0.1622)^2 = 6.1812$$

Since $\ln(Q)$ follows a normal distribution with a mean of $\mu_{\ln Q} = 6.1812$ and a standard deviation of $\sigma_{\ln Q} = 0.1622$ as previously computed, the magnitude of the log-transformed 100-year flood can be calculated by

$$\frac{\ln(q_{100}) - \mu_{\ln Q}}{\sigma_{\ln Q}} = \Phi^{-1}\left(1 - \frac{1}{100}\right) = \Phi^{-1}(0.99) = 2.34$$

Hence $\ln(q_{100}) = \mu_{\ln Q} + 2.34 \times \sigma_{\ln Q} = 6.5607$, and the corresponding 100-year flood magnitude can be calculated as $q_{100} = \exp[\ln(q_{100})] = 706.8$ ft^3/s.

For some distributions, such as Pearson type 3 or log-Pearson type 3, the appropriate probability paper or CDF inverse form is unavailable. In such a case, an analytical approach using the *frequency factor* K_T is applied:

$$x_T = \mu_x + K_T \times \sigma_x \tag{3.6}$$

in which x_T is the variate corresponding to a return period of T, μ_x and σ_x are the mean and standard deviation of the random variable, respectively, and K_T is the frequency factor, which is a function of the return period T or $P(X \geq x_T)$ and higher moments, if required. It is clear that a plot of Eq. (3.6) (x_T versus K_T)

on linear graph paper will yield a straight line with slope of σ_x and intercept μ_x at $K_T = 0$.

In order for Eq. (3.6) to be useful, the functional relationship between K_T and exceedance probability or return period must be determined for the distribution to be used. In fact, the frequency factor $K_T = (x_T - \mu_x)/\sigma_x$ is identical to a standardized variate corresponding to the exceedance probability of $1/T$ for a particular distribution model under consideration. For example, if the normal distribution is considered, then $K_T = z_T = \Phi^{-1}(1 - T^{-1})$. The same applies to the lognormal distribution when the mean and standard deviation of log-transformed random variables are used. Hence the standard normal probability table (Table 2.2) provides values of the frequency factor for sample data from normal and log normal distributions. Once this relation is known, a nonlinear probability or return-period scale can be constructed to replace the linear K_T scale, and thus special graph paper can be constructed for any distribution so that plot of x_T versus P or T will be linear.

Gumbel probability paper has been printed, although it is not readily available from commercial sources. Referring to Eq. (2.85a), the relationship between K_T and T for this distribution can be derived as

$$K_T = -\frac{\sqrt{6}}{\pi}\left[0.5772 + \ln\left(\frac{T}{T-1}\right)\right] \tag{3.7}$$

For Pearson and log-Pearson type 3 distributions, linearization can be accomplished according to Eq. (3.6). However, for this distribution, the frequency factor is a function of both P or T and the skewness coefficient γ_x. This means that a different nonlinear P or T scale is required for each skewness coefficient, and therefore, it is impractical to construct a probability paper for this distribution. However, it should be pointed out that if $\gamma_x = 0$ in log-space, the log-Pearson type 3 reduces to the lognormal distribution, and thus commercial lognormal probability paper can be used. The relationship between frequency factor K_T, T, and γ_x cannot be developed in a closed form, as was done for the Gumbel distribution in Eq. (3.7). However, the relationship can be computed numerically, and the results are given in Table 3.3. For $0.99^{-1} \leq T \leq 100$ and $|\gamma_x| < 2$, the frequency-factor values are well approximated by the Wilson-Hilferty transformation (Stedinger et al., 1993):

$$K_T(\gamma_x) = \frac{2}{\gamma_x}\left\{\left[1 + z_T\left(\frac{\gamma_x}{6}\right) - \left(\frac{\gamma_x}{6}\right)^2\right]^3 - 1\right\} \tag{3.8}$$

in which z_T is the standard normal quantile with exceedance probability of $1/T$.

The procedure for using the frequency-factor method is outlined as follows:

1. Compute the sample mean $\bar{x}$, standard deviation σ_x, and skewness coefficient γ_x (if needed) for the sample.

TABLE 3.3 Frequency Factor (K_T) for Pearson Type 3 Distribution

Skewness coefficient γ_x	Return period in years						
	2	5	10	25	50	100	200
	Exceedence probability						
	0.50	0.20	0.10	0.04	0.02	0.01	0.005
3.0	−0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.9	−0.390	0.440	1.195	2.277	3.134	4.013	4.909
2.8	−0.384	0.460	1.210	2.275	3.114	3.973	4.847
2.7	−0.376	0.479	1.224	2.272	3.093	3.932	4.783
2.6	−0.368	0.499	1.238	2.267	3.071	3.889	4.718
2.5	−0.360	0.518	1.250	2.262	3.048	3.845	4.652
2.4	−0.351	0.537	1.262	2.256	3.023	3.800	4.584
2.3	−0.341	0.555	1.274	2.248	2.997	3.753	4.515
2.2	−0.330	0.574	1.284	2.240	2.970	3.705	4.444
2.1	−0.319	0.592	1.294	2.230	2.942	3.656	4.372
2.0	−0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.9	−0.294	0.627	1.310	2.207	2.881	3.553	4.223
1.8	−0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.7	−0.268	0.660	1.324	2.179	2.815	3.444	4.069
1.6	−0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.5	−0.240	0.690	1.333	2.146	2.743	3.330	3.910
1.4	−0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.3	−0.210	0.719	1.339	2.108	2.666	3.211	3.745
1.2	−0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.1	−0.180	0.745	1.341	2.066	2.585	3.087	3.575
1.0	−0.164	0.758	1.340	2.043	2.542	3.022	3.489
0.9	−0.148	0.769	1.339	2.018	2.498	2.957	3.401
0.8	−0.132	0.780	1.336	1.993	2.453	2.891	3.312
0.7	−0.116	0.790	1.333	1.967	2.407	2.824	3.223
0.6	−0.099	0.800	1.328	1.939	2.359	2.755	3.132
0.5	−0.083	0.808	1.323	1.910	2.311	2.686	3.041
0.4	−0.066	0.816	1.317	1.880	2.261	2.615	2.949
0.3	−0.050	0.824	1.309	1.849	2.211	2.544	2.856
0.2	−0.033	0.830	1.301	1.818	2.159	2.472	2.763
0.1	−0.017	0.836	1.292	1.785	2.107	2.400	2.670
0.0	0	0.842	1.282	1.751	2.054	2.326	2.576
−0.1	0.017	0.846	1.270	1.716	2.000	2.252	2.482
−0.2	0.033	0.850	1.258	1.680	1.945	2.178	2.388
−0.3	0.050	0.853	1.245	1.643	1.890	2.104	2.294
−0.4	0.066	0.855	1.231	1.606	1.834	2.029	2.201
−0.5	0.083	0.856	1.216	1.567	1.777	1.955	2.108
−0.6	0.099	0.857	1.200	1.528	1.720	1.880	2.016
−0.7	0.116	0.857	1.183	1.488	1.663	1.806	1.926
−0.8	0.132	0.856	1.166	1.448	1.606	1.733	1.837
−0.9	0.148	0.854	1.147	1.407	1.549	1.660	1.749
−1.0	0.164	0.852	1.128	1.366	1.492	1.588	1.664
−1.1	0.180	0.848	1.107	1.324	1.435	1.518	1.581
−1.2	0.195	0.844	1.086	1.282	1.379	1.449	1.501
−1.3	0.210	0.838	1.064	1.240	1.324	1.383	1.424
−1.4	0.225	0.832	1.041	1.198	1.270	1.318	1.351
−1.5	0.240	0.825	1.018	1.157	1.217	1.256	1.282
−1.6	0.254	0.817	0.994	1.116	1.166	1.197	1.216

(Continued)

TABLE 3.3   Frequency Factor (K_T) for Pearson Type 3 Distribution (*Continued*)

Skewness coefficient γ_x	Return period in years						
	2	5	10	25	50	100	200
	Exceedence probability						
	0.50	0.20	0.10	0.04	0.02	0.01	0.005
−1.7	0.268	0.808	0.970	1.075	1.116	1.140	1.155
−1.8	0.282	0.799	0.945	1.035	1.069	1.087	1.097
−1.9	0.294	0.788	0.920	0.996	1.023	1.037	1.044
−2.0	0.307	0.777	0.895	0.959	0.980	0.990	0.995
−2.1	0.319	0.765	0.869	0.923	0.939	0.946	0.949
−2.2	0.330	0.752	0.844	0.888	0.900	0.905	0.907
−2.3	0.341	0.739	0.819	0.855	0.864	0.867	0.869
−2.4	0.351	0.725	0.795	0.823	0.830	0.832	0.833
−2.5	0.360	0.711	0.771	0.793	0.798	0.799	0.800
−2.6	0.368	0.696	0.747	0.764	0.768	0.769	0.769
−2.7	0.376	0.681	0.724	0.738	0.740	0.740	0.741
−2.8	0.384	0.666	0.702	0.712	0.714	0.714	0.714
−2.9	0.390	0.651	0.681	0.683	0.689	0.690	0.690
−3.0	0.396	0.636	0.666	0.666	0.666	0.667	0.667

SOURCE: U.S. Water Resources Council (1981).

2. For the desired return period, determine the associated value of K_T for the distribution.

3. Compute the desired quantile value using Eq. (3.6) with $\bar{x}$ replacing μ_x and s_x replacing σ_x, that is,

$$\hat{x}_T = \bar{x} + K_T \times s_x \qquad (3.9)$$

It should be recognized that the basic difference between the graphic and analytical approaches is that each represents a different method of estimating the statistical parameters of the distribution being used. By the analytical approach, a best-fit line is constructed that then sets the statistical parameters. In the mathematical approach, the statistical parameters are first computed from the sample, and effectively, the line thus determined is used. The line determined using the mathematical approach is in general a poorer fit to the observed data than that obtained using the graphic approach, especially if curve-fitting procedures are applied. However, the U.S. Water Resources Council (1967) recommended use of the mathematical approach because

1. Graphic least-squares methods are avoided to reduce the incorporation of the random characteristics of the particular data set (especially in the light of the difficulty in selecting the proper plotting-position formula).

2. The generally larger variance of the mathematical approach is believed to help compensate for the typically small data sets.

Example 3.3 Using the frequency-factor method, estimate the flows with return periods of 2, 10, 25, 50, and 100 years for the Boneyard Creek using the Gumbel and log-Pearson type 3 distributions. Use the historical data in Example 3.1 as a basis for the calculations.

Solution Based on the samples, the method requires determination of the frequency factor K_T in

$$\hat{x}_T = \bar{x} + K_T \times s_x$$

For the Gumbel distribution, values of K_T can be calculated by Eq. (3.7). For the log-Pearson type 3 distribution, Table 3.3 or Eq. (3.8) can be used, which requires computation of the skewness coefficient. The calculations of relevant sample moments are shown in the following table:

Year	q_i(ft^3/s)	q_i^2	q_i^3	$y_i = \ln(q_i)$	y_i^2	y_i^3
		Original scale			Log -Transformed scale	
1961	390	$1.52e+05$	$5.93e+07$	5.97	35.59	212.36
1962	374	$1.40e+05$	$5.23e+07$	5.92	35.10	207.92
1963	342	$1.17e+05$	$4.00e+07$	5.83	34.05	198.65
1964	507	$2.57e+05$	$1.30e+08$	6.23	38.79	241.63
1965	596	$3.55e+05$	$2.12e+08$	6.39	40.84	260.95
1966	416	$1.73e+05$	$7.20e+07$	6.03	36.37	219.33
1967	533	$2.84e+05$	$1.51e+08$	6.28	39.42	247.50
1968	505	$2.55e+05$	$1.29e+08$	6.22	38.75	241.17
1969	549	$3.01e+05$	$1.65e+08$	6.31	39.79	251.01
1970	414	$1.71e+05$	$7.10e+07$	6.03	36.31	218.81
1971	524	$2.75e+05$	$1.44e+08$	6.26	39.21	245.49
1972	505	$2.55e+05$	$1.29e+08$	6.22	38.75	241.17
1973	447	$2.00e+05$	$8.93e+07$	6.10	37.24	227.27
1974	543	$2.95e+05$	$1.60e+08$	6.30	39.65	249.70
1975	591	$3.49e+05$	$2.06e+08$	6.38	40.73	259.92
Sum =	7236	$3.58e+06$	$1.81e+09$	92.48	570.58	3522.88

For the Gumbel distribution,

$$\bar{q} = \frac{7236}{15} = 482.4 \text{ft}^3/\text{s}$$

$$s = \left(\frac{\sum q_i^2 - 15\bar{q}^2}{15 - 1} \right)^{1/2} = (6361.8)^{1/2} = 79.8 \text{ft}^3/\text{s}$$

For the log-Pearson type 3 distribution,

$$\bar{y} = \frac{\sum \ln(q_i)}{n} = \frac{92.48}{15} = 6.165$$

$$s_y = \left(\frac{\sum y_i^2 - 15\bar{y}^2}{15 - 1} \right)^{1/2} = (0.417/14)^{1/2} = 0.173$$

$$g_y = \frac{n}{(n-1)(n-2)} \frac{m_3}{s_y^3} = \frac{15(-0.00336)}{(14)(13)(0.173)^3} = -0.540$$

in which $m_3 = \sum y_i^3 - 3\bar{y} \sum y_i^2 + 2n\bar{y}^3$. The determination of the values of frequency factor corresponding to different return periods is shown in the following table:

Return period (years)	Exceedance probability	Nonexceedance probability	Frequency factor by distribution	
			Gumbel Eq. (3.7)	LP3 Eq. (3.8)
2	0.50	0.50	−0.1643	0.0892
10	0.10	0.90	1.3046	1.2093
25	0.04	0.96	2.0438	1.5526
50	0.02	0.98	2.5923	1.7570
100	0.01	0.99	3.1367	1.9292

Based on the preceding frequency-factor values, the flood magnitude of the various return periods can be determined as

Return period (years)	Frequency curves by distribution (ft^3/s)	
	Gumbel $q_T = 482.4 + 79.8 K_{T,\,\text{EV1}}$	LP3 $q_T = \exp(6.165 + 0.173 K_{T,\,\text{LP3}})$
2	469.3	483.3
10	586.5	586.4
25	645.4	622.2
50	689.2	644.5
100	732.6	663.9

One could compare these results for the Gumbel distribution with those obtained from the graphic approach of Example 3.1.

3.6 Estimation of Distributional Parameters

For a chosen distributional model, its shape and position are completely defined by the associated parameters. By referring to Eq. (3.5), determination of the quantile also requires knowing the values of the parameters $\boldsymbol{\theta}$.

There are several methods for estimating the parameters of a distribution model on the basis of available data. In frequency analysis, the commonly used parameter-estimation procedures are the method of maximum likelihood and the methods of moments (Kite, 1988; Haan, 1977). Other methods, such as method of maximum entropy (Li et al., 1986), have been applied.

3.6.1 Maximum-likelihood (ML) method

This method determines the values of parameters of a distribution model that maximizes the likelihood of the sample data at hand. For a sample of n independent random observations, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^t$, from an identical distribution, that is,

$$x_i \stackrel{\text{iid}}{\sim} f_x(x \mid \boldsymbol{\theta}) \qquad \text{for } i = 1, 2, \ldots, n$$

in which $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^{\mathrm{t}}$, a vector of m distribution model parameters, the likelihood of occurrence of the samples is equal to the joint probability of $\{x_i\}_{i=1,2,\ldots,n}$ calculable by

$$L(\boldsymbol{x} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_x(x_i \,|\, \boldsymbol{\theta}) \tag{3.10}$$

in which $L(\boldsymbol{x}|\boldsymbol{\theta})$ is called the *likelihood function*. The ML method determines the distribution parameters by solving

$$\operatorname*{Max}_{\theta} L(\boldsymbol{x} \,|\, \boldsymbol{\theta}) = \max_{\theta} \ln[L(\boldsymbol{x} \,|\, \boldsymbol{\theta})]$$

or more specifically

$$\operatorname*{Max}_{\theta} \prod_{i=1}^{n} f_x(x_i \,|\, \boldsymbol{\theta}) = \max_{\theta} \sum_{i=1}^{n} \ln[f_x(x_i \,|\, \boldsymbol{\theta})] \tag{3.11}$$

As can be seen, solving for distribution-model parameters by the ML principle is an unconstrained optimization problem. The unknown model parameters can be obtained by solving the following necessary conditions for the maximum:

$$\frac{\partial \left\{ \sum_{i=1}^{n} \ln[f_x(x_i \,|\, \boldsymbol{\theta})] \right\}}{\partial \theta_j} = 0 \qquad \text{for } j = 1, 2, \ldots, m \tag{3.12}$$

In general, depending on the form of the distribution model under consideration, Eq. (3.12) could be a system of nonlinear equations requiring numerical solution. Alternatively, Eq. (3.11) can be solved by a proper direct optimum search scheme, such as the conjugate gradient method or quasi-Newton method (McCormick, 1983 or see Section 8.13.2).

Example 3.4 Referring to Eq. (2.79) for the exponential distribution as

$$f_x(x \,|\, \beta) = \exp(-x/\beta)/\beta \qquad \text{for } x > 0, \beta > 0$$

determine the maximum likelihood estimate for β based on n independent random samples $\{x_i\}_{i=1,2,\ldots,n}$.

Solution The log-likelihood function for the exponential distribution is

$$\ln[L(\boldsymbol{x} \,|\, \beta)] = \sum_{i=1}^{n} \ln\left(\frac{1}{\beta} \mathrm{e}^{-x_i/\beta}\right) = -n \ln(\beta) - \frac{1}{\beta} \sum_{i=1}^{n} x_i$$

The parameter β that maximizes the preceding log-likelihood function, according to the necessary condition, i.e., Eq. (3.12), is

$$\frac{\partial}{\partial \beta} \{\ln[L(\boldsymbol{x} \,|\, \beta)]\} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i = 0$$

Hence the ML estimator of β for the exponential distribution is

$$\hat{\beta}_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

which is the sample mean.

Example 3.5 Consider a set of n independent samples, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^t$, from a normal distribution with the following PDF:

$$f_x(\boldsymbol{x} \mid \alpha, \beta) = \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2} \qquad \text{for} - \infty < x < \infty$$

Determine the ML estimators for the parameters α and β.

Solution The likelihood function for the n independent normal samples is

$$L(\boldsymbol{x} \mid \alpha, \beta) = \left(\frac{1}{\beta\sqrt{2\pi}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \alpha)^2}{2\beta^2}\right]$$

The corresponding log-likelihood function can be written as

$$\ln[L(\boldsymbol{x} \mid \alpha, \beta)] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\beta^2) - \frac{\sum_{i=1}^{n}(x_i - \alpha)^2}{2\beta^2}$$

Taking the partial derivatives of the preceding log-likelihood function with respect to α and β^2 and setting them equal to zero results in

$$\frac{\partial\{\ln[L(\boldsymbol{x} \mid \alpha, \beta)]\}}{\partial\alpha} = \frac{\sum_{i=1}^{n}(x_i - \alpha)}{\beta^2} = 0$$

$$\frac{\partial\{\ln[L(\boldsymbol{x} \mid \alpha, \beta)]\}}{\partial\beta^2} = -\frac{n}{2\beta^2} + \frac{\sum_{i=1}^{n}(x_i - \alpha)^2}{2\beta^4} = 0$$

After some algebraic manipulations, one can easily obtain the ML estimates of normal distribution parameters α and β as

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \hat{\beta}_{\mathrm{ML}}^2 = \frac{\sum_{i=1}^{n}(x_i - \alpha)^2}{n}$$

As can be seen, the ML estimation of the normal parameters for α is the sample mean and for β^2 is a biased variance.

3.6.2 Product-moments-based method

By the moment-based parameter-estimation methods, parameters of a distribution are related to the statistical moments of the random variable. The conventional method of moments uses the product moments of the random variable. Example 3.3 for frequency analysis is typical of this approach. When sample data are available, sample product moments are used to solve for the model parameters. The main concern with the use of product moments is that their reliabilities owing to sampling errors deteriorate rapidly as the order of moment

increases, especially when sample size is small (see Sec. 3.1), which is often the case in many geophysical applications. Hence, in practice only, the first few statistical moments are used. Relationships between product-moments and parameters of distribution models commonly used in frequency analysis are listed in Table 3.4.

3.6.3 L-moments-based method

As described in Sec. 2.4.1, the L-moments are linear combinations of order statistics (Hosking, 1986). In theory, the estimators of L-moments are less sensitive to the presence of outliers in the sample and hence are more robust than the conventional product moments. Furthermore, estimators of L-moments are less biased and approach the asymptotic normal distributions more rapidly and closely. Hosking (1986) shows that parameter estimates from the L-moments are sometimes more accurate in small samples than are the maximum-likelihood estimates.

To calculate sample L-moments, one can refer to the probability-weighted moments as

$$\beta_r = M_{1,r,0} = E\{X[F_x(X)]^r\} \qquad \text{for } r = 0, 1, \ldots \qquad (3.13)$$

which is defined on the basis of nonexceedance probability or CDF. The estimation of β_r then is hinged on how $F_x(X)$ is estimated on the basis of sample data.

Consider n independent samples arranged in ascending order as $X_{(n)} \leq X_{(n-1)} \leq \cdots \leq X_{(2)} \leq X_{(1)}$. The estimator for $F_x(X_{(m)})$ for the mth-order statistic can use an appropriate plotting-position formula as shown in Table 3.2, that is,

$$\hat{F}(X_{(m)}) = 1 - \frac{m - a}{n + 1 - b} \qquad \text{for } m = 1, 2, \ldots, n$$

with $a \geq 0$ and $b \geq 0$. The Weibull plotting-position formula ($a = 0$, $b = 0$) is a probability-unbiased estimator of $F_x(X_{(m)})$. Hosking et al. (1985a, 1985b) show that a smaller mean square error in the quantile estimate can be achieved by using a biased plotting-position formula with $a = 0.35$ and $b = 1$. According to the definition of the β-moment β_r in Eq. (3.13), its sample estimate b_r can be obtained easily as

$$b_r = \frac{1}{n} \sum_{i=1}^{n} x_{(m)}[\hat{F}(x_{(m)})]^r \qquad \text{for } r = 0, 1, \ldots \qquad (3.14)$$

Stedinger et al. (1993) recommend the use of the quantile-unbiased estimator of $F_x(X_{(m)})$ for calculating the L-moment ratios in at-site and regional frequency analyses.

TABLE 3.4 Relations between Moments and Parameters of Selected Distribution Models

Distribution	PDF or CDF	Range	Product moments	L-Moments
Normal	$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$		$\lambda_1 = \mu; \ \lambda_2 = \tilde{\sigma}/\sqrt{\pi};$ $\tau_3 = 0; \tau_4 = 0.1226$
Lognormal	$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_{\ln x}x} \exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\mu_{\ln x}}{\sigma_{\ln x}}\right)^2\right]$	$x > 0$	$\mu_{\ln x} = \ln(\mu_x) - \sigma_{\ln x}^2/2;$ $\sigma_{\ln x}^2 = \ln(\Omega_x^2 + 1);$ $\gamma_x = 3\Omega_x + \Omega_x^3$	Eq. (2.68); Eq. (2.70)
Rayleigh	$f_x(x) = \frac{(x-\xi)}{\alpha^2} \exp\left[-\frac{1}{2}\left(\frac{x-\xi}{\alpha}\right)^2\right]$	$\xi \leq x < \infty$	$\mu = \xi + \sqrt{\pi/2}\,\alpha;$ $\sigma = \sqrt{2 - \frac{\pi}{2}}\,\alpha$	$\lambda_1 = \xi + \alpha\sqrt{\pi/2};$ $\lambda_2 = \frac{1}{2}\alpha\sqrt{\pi}(\sqrt{2}-1)$ $\tau_3 = 0.1140; \tau_4 = 0.1054$
Pearson 3	$f_x(x) = \frac{1}{|\beta|\Gamma(\alpha)}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1} e^{-(x-\xi)/\beta}$	$\alpha > 0$ for $\beta > 0$: $x > \xi$; for $\beta < 0$: $x < \xi$	$\mu = \xi + \alpha\beta; \ \sigma^2 = \alpha\beta^2;$ $\gamma = \text{sign}(\beta)\frac{2}{\sqrt{\alpha}}$	
Exponential	$f_x(x) = e^{-x/\beta}/\beta$	$x > 0$	$\mu = \beta$	$\lambda_1 = \beta; \ \lambda_2 = \beta/2;$ $\tau_3 = 1/3; \tau_4 = 1/6$
Gumbel (EV1 for maxima)	$f_x(x) = \frac{1}{\beta}\exp\left\{-\left(\frac{x-\xi}{\beta}\right) - \exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right\}$	$-\infty < x < \infty$	$\mu = \xi + 0.5772\beta;$ $\sigma^2 = \frac{\pi^2\beta^2}{6}; \ \gamma = 1.1396$	$\lambda_1 = \xi + 0.5772\beta; \ \lambda_2 = \beta\ln(2);$ $\tau_3 = 0.1699; \tau_4 = 0.1504$
Weibull	$f_x(x) = \frac{\alpha}{\beta}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x-\xi}{\beta}\right)^\alpha\right]$	$\alpha, \beta > 0; x > 0$	$\mu = \beta\Gamma\left(1+\frac{1}{\beta}\right);$ $\sigma^2 = \beta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]$	$\lambda_1 = \xi + \beta\,\Gamma\left(1+\frac{1}{\alpha}\right);$ $\lambda_2 = \beta\left(1-2^{-1/\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right)$
Generalized extreme-value (GEV)	$F_x(x) = \exp\left\{-\left[1 - \frac{\alpha(x-\xi)}{\beta}\right]^{1/\alpha}\right\}$	$\alpha > 0: x < \left(\xi + \frac{\beta}{\alpha}\right);$ $\alpha < 0: x > \left(\xi + \frac{\beta}{\alpha}\right)$	$\mu = \xi + \frac{\beta}{\alpha}[1 - \Gamma(1+\alpha)];$ $\sigma^2 = \left(\frac{\beta}{\alpha}\right)^2[\Gamma(1+2\alpha) - \Gamma^2(1+\alpha)]$	$\lambda_1 = \xi + \left(\frac{\beta}{\alpha}\right)[1 - \Gamma(1+\alpha)];$ $\lambda_2 = \frac{\beta}{\alpha}(1 - 2^{-\alpha})\Gamma(1+\alpha);$ $\tau_3 = \frac{2(1-3^{-\alpha})}{(1-2^{-\alpha})} - 3;$ $\tau_4 = \frac{1 - 5(4^{-\alpha}) + 10(3^{-\alpha}) - 6(2^{-\alpha})}{1 - 2^{-\alpha}}$
Generalized Pareto (GPA)	$F_x(x) = 1 - \left[1 - \alpha\frac{(x-\xi)}{\beta}\right]^{1/\alpha}$	$\alpha > 0:$ $\zeta \leq x \leq \left(\xi + \frac{\beta}{\alpha}\right);$ $\alpha < 0: \zeta \leq x < \infty;$	$\mu = \xi + \frac{\beta}{1+\alpha};$ $\sigma^2 = \frac{\beta^2}{(1+\alpha)^2+(1+2\alpha)};$ $\gamma = \frac{2(1-\alpha)(1+2\alpha)^{1/2}}{(1+3\alpha)}$	$\lambda_1 = \xi + \frac{\beta}{1+\alpha};$ $\lambda_2 = \frac{\beta}{(1+\alpha)(2+\alpha)};$ $\tau_3 = \frac{1-\alpha}{3+\alpha};$ $\tau_4 = \frac{(1-\alpha)(2-\alpha)}{(3+\alpha)(4+\alpha)}$

For any distribution, the L-moments can be expressed in terms of the probability-weighted moments as shown in Eq. (2.28). To compute the sample L-moments, the sample probability-weighted moments can be obtained as

$$l_1 = b_0$$
$$l_2 = 2b_1 - b_0$$
$$l_3 = 6b_2 - 6b_1 + b_0 \tag{3.15}$$
$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

where the l_rs are sample estimates of the corresponding L-moments, the λ_rs, respectively. Accordingly, the sample L-moment ratios can be computed as

$$t_2 = \frac{l_2}{l_1} \qquad t_3 = \frac{l_3}{l_2} \qquad t_4 = \frac{l_4}{l_2} \tag{3.16}$$

where t_2, t_3, and t_4 are the sample L-coefficient of variation, L-skewness coefficient, and L-kurtosis, respectively. Relationships between L-moments and parameters of distribution models commonly used in frequency analysis are shown in the last column of Table 3.4.

Example 3.6 Referring to Example 3.3, estimate the parameters of a generalized Pareto (GPA) distribution by the L-moment method.

Solution Since the GPA is a three-parameter distribution, the calculation of the first three sample L-moments is shown in the following table:

Year	q_i (ft³/s)	Ordered $q_{(i)}$ (ft³/s)	Rank (i)	$F(q_{(i)}) =$ $(i - 0.35)/n$	$q_{(i)} \times F(q_{(i)})$	$q_{(i)} \times F(q_{(i)})^2$	$q_{(i)} \times F(q_{(i)})^3$
1961	390	342	1	0.0433	14.82	0.642	0.0278
1962	374	374	2	0.1100	41.14	4.525	0.4978
1963	342	390	3	0.1767	68.90	12.172	2.1504
1964	507	414	4	0.2433	100.74	24.513	5.9649
1965	596	416	5	0.3100	128.96	39.978	12.3931
1966	416	447	6	0.3767	168.37	63.419	23.8880
1967	533	505	7	0.4433	223.88	99.255	44.0030
1968	505	505	8	0.5100	257.55	131.351	66.9888
1969	549	507	9	0.5767	292.37	168.600	97.2260
1970	414	524	10	0.6433	337.11	216.872	139.5210
1971	524	533	11	0.7100	378.43	268.685	190.7666
1972	505	543	12	0.7767	421.73	327.544	254.3922
1973	447	549	13	0.8433	462.99	390.455	329.2836
1974	543	591	14	0.9100	537.81	489.407	445.3605
1975	591	596	15	0.9767	582.09	568.511	555.2459
	Sum =	7236			4016.89	2805.930	2167.710

Note that the plotting-position formula used in the preceding calculation is that proposed by Hosking et al. (1985a) with $a = 0.35$ and $b = 1$.

Based on Eq. (3.14), the sample estimates of β_j, for $j = 0, 1, 2, 3$, are $b_0 = 428.4$, $b_1 = 267.80$, $b_3 = 187.06$, and $b_4 = 144.51$. Hence, by Eq. (3.15), the sample estimates of λ_j, $j = 1, 2, 3, 4$, are $l_1 = 482.40$, $l_2 = 53.19$, $l_3 = -1.99$, and $l_4 = 9.53$, and the corresponding sample L-moment ratios τ_j, for $j = 2, 3, 4$, are $t_2 = 0.110$, $t_3 = -0.037$, and $t_4 = 0.179$.

By referring to Table 3.4, the preceding sample $l_1 = 482.40$, $l_2 = 53.19$, and $t_3 = -0.037$ can be used in the corresponding L-moment and parameter relations, that is,

$$l_1 = 482.40 = \xi + \frac{\beta}{1 + \alpha}$$

$$l_2 = 53.19 = \frac{\beta}{(1 + \alpha)(2 + \alpha)}$$

$$t_3 = -0.037 = \frac{1 - \alpha}{3 + \alpha}$$

Solving backwards starting from t_3, l_2, and then l_1, then, the values of sample parameter estimates of the GPA distribution can be obtained as

$$\hat{\alpha} = 1.154 \qquad \hat{\beta} = 361.36 \qquad \hat{\xi} = 314.64$$

3.7 Selection of Distribution Model

Based on a given sample of finite observations, procedures are needed to help identify the underlying distribution from which the random samples are drawn. Several statistical goodness-of-fit procedures have been developed (D'Agostino and Stephens, 1986). The insensitivity to the tail portion of the distribution of the conventional chi-square test and Kolmogorov-Smirnov test has been well known. Other more powerful goodness-of-fit criteria such as the probability plot correlation coefficient (Filliben, 1975) have been investigated and advocated (Vogel and McMartin, 1991). This and other criteria are described herein.

3.7.1 Probability plot correlation coefficients

The probability plot is a graphic representation of the mth-order statistic of the sample $x_{(m)}$ as a function of a plotting-position $\hat{F}(x_{(m)})$. For each order statistic $X_{(m)}$, a plotting-position formula can be applied to estimate its corresponding nonexceedance probability $\hat{F}(x_{(m)})$, which, in turn, is used to compute the corresponding quantile $Y_m = G^{-1}[\hat{F}(X_{(m)})]$ according to the distribution model $G(\cdot)$ under consideration. Based on a sample with n observations, the *probability plot correlation coefficient* (*PPCC*) then can be defined mathematically as

$$\text{PPCC} = \frac{\sum_{m=1}^{n}(x_{(m)} - \bar{x})(y_m - \bar{y})}{\left[\sum_{m=1}^{n}(x_{(m)} - \bar{x})^2 \sum_{m=1}^{n}(y_m - \bar{y})^2\right]^{0.5}} \tag{3.17}$$

where y_m is the quantile value corresponding to $\hat{F}(x_{(m)})$ from a selected plotting-position formula and an assumed distribution model $G(\cdot)$, that is, $y_m = G^{-1}[\hat{F}(x_{(m)})]$. It is intuitively understandable that if the samples to be tested are actually generated from the hypothesized distribution model $G(\cdot)$, the corresponding plot of $x_{(m)}$ versus y_m would be close to linear. The values of $\hat{F}(x_{(m)})$ for calculating y_m in Eq. (3.17) can be determined by using either a probability- or quantile-unbiased plotting-position formula. The hypothesized distribution model $G(\cdot)$ that yields the highest value of the PPCC should be chosen.

Critical values of the PPCCs associated with different levels of significance for various distributions have been developed. They include normal and lognormal distribution (Fillben, 1975; Looney and Gulledge, 1985; Vogel, 1986), Gumbel distribution (Vogel, 1986), uniform and Weibull distributions (Vogel and Kroll, 1989), generalized extreme-value distribution (Chowdhury et al., 1991), Pearson type 3 distribution (Vogel and McMartin, 1991), and other distributions (D'Agostino and Stephens, 1986). A distribution is accepted as the underlying random mechanism with a specified significance level if the computed PPCC is larger than the critical value for that distribution.

3.7.2   Model reliability indices

Based on the observed $\{x_{(m)}\}$ and the computed $\{y_m\}$, the degree of goodness of fit also can be measured by two reliability indices proposed by Leggett and Williams (1981). They are the *geometric reliability index K_G*,

$$K_G = \frac{1 + \sqrt{\frac{1}{n}\sum_{m=1}^{n}\left[\frac{1-(y_m/x_{(m)})}{1+(y_m/x_{(m)})}\right]^2}}{1 - \sqrt{\frac{1}{n}\sum_{m=1}^{n}\left[\frac{1-(y_m/x_{(m)})}{1+(y_m/x_{(m)})}\right]^2}} \tag{3.18}$$

and the *statistical reliability index K_S*,

$$K_S = \exp\left\{\sqrt{\frac{1}{n}\sum_{m=1}^{n}\left[\log\left(\frac{y_m}{x_{(m)}}\right)\right]^2}\right\} \tag{3.19}$$

When the computed series $\{y_m\}$ perfectly matches with the observed sequence $\{x_{(m)}\}$, the values of K_G and K_S reach their lower bound of 1.0. As the discrepancy between $\{x_{(m)}\}$ and $\{y_m\}$ increases, the values of K_G and K_S increase. Again, for each of K_G and K_S, two different values can be computed, each associated with the use of probability-unbiased and quantile-unbiased plotting-position formulas. The most suitable probability model is the one that is associated with the smallest value of the reliability index.

3.7.3   Moment-ratio diagrams

Relationships between product moments and the parameters of various distributions are shown in Table 3.4, which also can be found elsewhere (Patel et al., 1976; Stedinger et al., 1993). Similarly, the *product-moment ratio diagram*

based on skewness coefficient and kurtosis (Stuart and Ord, 1987, p. 211) can be used to identify the distributions. When sample data are used, sample product moments are used to solve for the model parameters. However, owing to the low reliability of sample skewness coefficient and kurtosis, use of the product-moment ratio diagram for model identification is not reliable. Alternatively, the *L-moment ratio diagram* defined in the (τ_3, τ_4)-space (Fig. 3.3) also can be used for model identification. Namely, one can judge the closeness of the sample L-skewness coefficient and L-kurtosis with respect to the theoretical $\tau_3 - \tau_4$ curve associated with different distribution models. Some types of distance measures can be computed between the sample point of (t_3, t_4) and each theoretical $\tau_3 - \tau_4$ curve. One commonly used distance measure is to compute the shortest distance or distance in L-kurtosis direction fixed at the sample L-skewness coefficient (Pandey et al., 2001). Although it is computationally simple, the

E Exponential	------- Generalized logistic	●—— Lower bound for wakeby
G Gumbel	-·—·- Generalized extreme–value	
L Logistic	------ Generalized pareto	
N Normal	– – – Lognormal	■—— Lower bound for all distributions
U Uniform	—— Gamma	
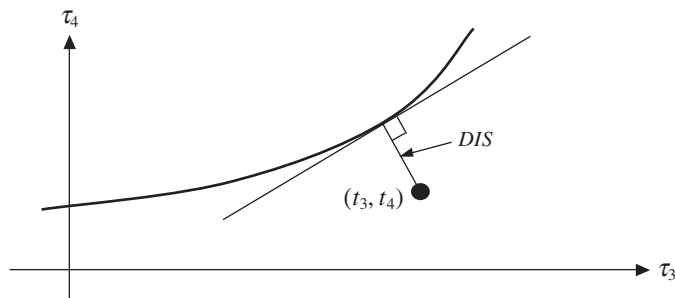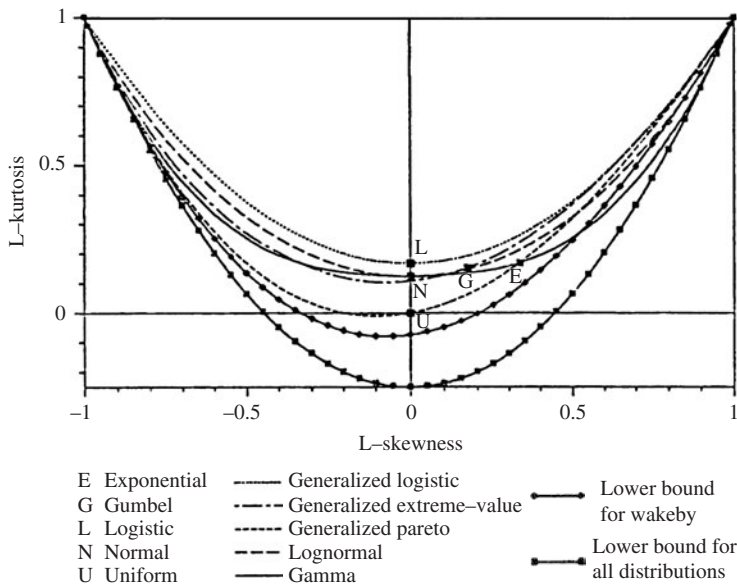
Figure 3.3 L-moment ratio diagram and shortest distance from a sample point.

distance measure could not account for the sampling error in the sample L-skewness coefficient. To consider the effect of sampling errors in both the sample L-skewness coefficient and L-kurtosis, the shortest distance between the sample point (t_3, t_4) and the theoretical $\tau_3 - \tau_4$ curve of each candidate distribution model is computed for the measure of goodness of fit. The computation of the shortest distance requires locating a point on the theoretical $\tau_3 - \tau_4$ curve that minimizes the distance as

$$\text{DIS} = \min_{\tau_3} \sqrt{(t_3 - \tau_3)^2 + [t_4 - \tau_4(\tau_3)]^2} \tag{3.20}$$

Since the theoretical $\tau_3 - \tau_4$ curve for a specified distribution is unique, determination of the shortest distance was accomplished by an appropriate one-dimensional search technique such as the golden-section procedure or others.

Example 3.7 (Goodness of Fit) Referring to the flood data given in Example 3.3, calculate the values of the probability-unbiased PPCCs and the two reliability indices with respect to the generalized Pareto distribution (GPA).

Solution Referring to Table 3.4, the GPA quantile can be obtained easily as

$$x(F) = \xi + \frac{\beta}{\alpha}[1 - (1 - F)^\alpha]$$

According to the model parameter values obtained from Example 3.6, that is, $\hat{\alpha} = 1.154$, $\hat{\beta} = 361.36$, $\hat{\xi} = 314.64$, the GPA quantile can be computed as

$$x(F) = 314.64 + \frac{361.36}{1.154}\left[1 - (1 - F)^{1.154}\right]$$

Using the probability-unbiased plotting position, i.e., the Weibull formula, the corresponding GPA quantiles are calculated and shown in column (4) of the following table. From data in columns (2) and (4), the correlation coefficient can be obtained as 0.9843.

To calculate the two-model reliability indices, the ratios of GPA quantiles y_m to the order flow $q_{(m)}$ are calculated in column (5) and are used in Eqs. (3.18) and (3.19) for K_G and K_S, respectively, as 1.035 and 1.015.

Rank (m) (1)	Ordered $q_{(m)}$ (2)	$F(q_{(m)}) = m/(n+1)$ (3)	y_m (4)	$y_m/q_{(m)}$ (5)
1	342	0.0625	337.1	0.985714
2	374	0.1250	359.4	0.960853
3	390	0.1875	381.4	0.977846
4	414	0.2500	403.1	0.973676
5	416	0.3125	424.6	1.020591
6	447	0.3750	445.7	0.997162
7	505	0.4375	466.6	0.923907
8	505	0.5000	487.1	0.964476
9	507	0.5625	507.2	1.000308
10	524	0.6250	526.8	1.005368
11	533	0.6875	546.0	1.024334
12	543	0.7500	564.5	1.039672
13	549	0.8125	582.4	1.060849
14	591	0.8750	599.4	1.014146
15	596	0.9375	615.0	1.031891

3.7.4 Summary

As the rule for selecting a single distribution model, the PPCC-based criterion would choose the model with highest values, whereas the other two criteria (i.e., reliability index and DIS) would select a distribution model with the smallest value. In practice, it is not uncommon to encounter a case where the values of the adopted goodness-of-fit criterion for different distributions are compatible, and selection of a best distribution may not necessarily be the best course of action, especially in the presence of sampling errors. The selection of acceptable distributions based on the their statistical plausibility through hypothesis testing, at the present stage, can only be done for the PPCCs for which extensive experiments have been done to define critical values under various significance levels (or type I errors) and different distributions.

3.8 Uncertainty Associated with a Frequency Relation

Consider Example 3.2 in which the annual maximum flood peak discharges over a 15-year period on the Boneyard Creek at Urbana, Illinois, were analyzed. Suppose that the annual maximum floods follow the Gumbel distribution. The estimated 25-year flood peak discharge is 656 ft^3/s. It is not difficult to imagine that if one had a second set of 15 years of record, the estimated 25-year flood based on the second 15-year record likely would be different from the first 15-year record. Also, combining with the second 15 years of record, the estimated 25-year flood magnitude based on a total of 30 years of record again would not have the same value as 656 ft^3/s. This indicates that the estimated 25-year flood is subject to uncertainty that is due primarily to the use of limited amount of data in frequency analysis. Furthermore, it is intuitive that the reliability of the estimated 25-year flood, based on a 30-year record, is higher than that based on a 15-year record.

From the preceding discussions one can conclude that using a limited amount of data in frequency analysis, the estimated value of a geophysical quantity of a particular return period x_T and the derived frequency relation are subject to uncertainty. The degree of uncertainty of the estimated x_T depends on the sample size, the extent of data extrapolation (i.e., return period relative to the record length), and the underlying probability distribution from which the data are sampled (i.e., the distribution). Since the estimated design quantity is subject to uncertainty, it is prudent for an engineer to quantify the magnitude of such uncertainty and assess its implications on the engineering design (Tung and Yen, 2005, Sec. 1.5). Further, Benson (1968) noted that the results of the U.S. Water Resources Council study to determine the "best" distribution indicated that confidence limits always should be computed for flood frequency analysis.

In practice, there are two ways to express the degree of uncertainty of a statistical quantity, namely, standard error and confidence interval (confidence limit). Because the estimated geophysical quantities of a particular return period are subject to uncertainty, they can be treated as a random variable associated with a distribution, as shown in Fig. 3.4. Similar to the standard deviation of a
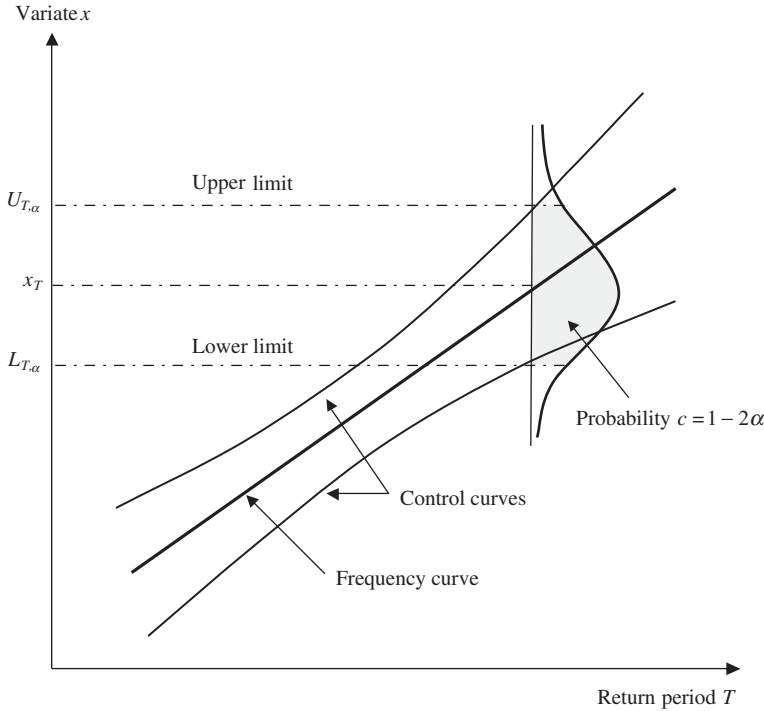
Figure 3.4 Definition of confidence limit.

random variable, the *standard error of estimate* s_e measures the standard deviation of an estimated statistical quantity from a sample, such as $\hat{x}_T$, about the true but unknown event magnitude. On the other hand, the *confidence limit* of an estimated quantity is an interval that has a specified probability (or confidence) to include the true value.

In the context of frequency analysis, the standard error of $\hat{x}_T$ is a function of the distribution of the data series under consideration and the method of determining the distribution parameters. For example, the asymptotic (that is, as $n \rightarrow \infty$) standard error of a T-year event $s_e(\hat{x}_T)$ from a normal distribution can be calculated as (Kite, 1988)

$$s_e(\hat{x}_T) = \left(\frac{2 + z_T^2}{2n} \right)^{1/2} s_x \tag{3.21}$$

in which z_T is the standard normal variate corresponding to the exceedance probability of $1/T$, that is, $\Phi(z_T) = 1 - 1/T$, n is the sample size, and s_x is the sample standard deviation of random variable X. From the Gumbel distribution, the standard error of $\hat{x}_T$ is

$$s_e(\hat{x}_T) = \left[\frac{1}{n} \left(1 + 1.1396\, K_T + 1.1\, K_T^2 \right) \right]^{1/2} s_x \tag{3.22}$$

To construct the confidence interval for $\hat{x}_T$ or for the frequency curve, a *confidence level c* that specifies the desired probability that the specified range will include the unknown true value is predetermined by the engineer. In practice, a confidence level of 95 or 90 percent is used. Corresponding to the confidence level c, the *significance level* α is defined as $\alpha = 1 - c$; for example, if the desired confidence level $c = 90$ percent, the corresponding significance level $\alpha = 10$ percent. In determining the confidence interval, the common practice is to distribute the significance level α equally on both ends of the distribution describing the uncertainty feature of estimated $\hat{x}_T$ (see Fig. 3.4). In doing so, the boundaries of the confidence interval, called *confidence limits*, are defined. Assuming normality for the asymptotic sample distribution for $\hat{x}_T$, the approximated $100(1 - \alpha)$ percent confidence interval for $\hat{x}_T$ is

$$x_{T,\alpha}^L = \hat{x}_T - z_{1-\alpha/2} \times s_e(\hat{x}_T) \qquad x_{T,\alpha}^U = \hat{x}_T + z_{1-\alpha/2} \times s_e(\hat{x}_T) \qquad (3.23)$$

in which $x_{T,\alpha}^L$ and $x_{T,\alpha}^U$ are, respectively, the values defining the lower and upper bounds for the $100(1 - \alpha)$ percent confidence interval, and $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. The confidence interval defined by Eq. (3.23) is only approximate and the approximation accuracy increases with sample size.

Similar to the frequency-factor method, the formulas to compute the upper and lower limits of confidence interval for $\hat{x}_T$ has the same form as Eq. (3.6), except that the frequency-factor term is adjusted as

$$x_{T,\alpha}^L = \bar{x} + K_{T,\alpha}^L \times s_x \qquad x_{T,\alpha}^U = \bar{x} + K_{T,\alpha}^U \times s_x \qquad (3.24)$$

in which $K_{T,\alpha}^L$ and $K_{T,\alpha}^U$ are the *confidence-limit factors* for the lower and upper limits of the $100(1 - \alpha)$ percent confidence interval, respectively. For random samples from a normal distribution, the exact confidence-limit factors can be determined using the noncentral-t variates ζ (Table 3.5). An approximation for $K_{T,\alpha}^L$ with reasonable accuracy for $n \geq 15$ and $\alpha = 1 - c \geq 5$ percent (Chowdhury et al., 1991) is

$$K_{T,\alpha}^L = \zeta_{T,\alpha/2} \approx \frac{z_T + z_{\alpha/2}\sqrt{\frac{1}{n} + \frac{z_T^2}{2(n-1)} - \frac{z_{\alpha/2}^2}{2n(n-1)}}}{1 - \frac{z_{\alpha/2}^2}{2(n-1)}} \qquad (3.25)$$

To compute $K_{T,\alpha}^U$, by symmetry, one only has to change $z_{\alpha/2}$ by $z_{1-\alpha/2}$ in Eq. (3.25). As was the case for Eq. (3.20), the confidence intervals defined by Eqs. (3.24) and (3.25) are most appropriate for samples from populations following a normal distribution, and for nonnormal populations, these confidence limits are only approximate, with the approximation accuracy increasing with sample size.

For Pearson type 3 distributions, the values of confidence-limit factors for different return periods and confidence levels given in Eq. (3.24) can be modified by introducing the scaling factor obtained from a first-order asymptotic

TABLE 3.5 95 Percent Confidence-Limit Factors for Normal Distribution

n	\multicolumn{12}{c}{Return period (years)}											
	\multicolumn{2}{c}{2}		\multicolumn{2}{c}{5}		\multicolumn{2}{c}{10}		\multicolumn{2}{c}{25}		\multicolumn{2}{c}{50}		\multicolumn{2}{c}{100}	
	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$	$K_{T,\alpha}^L$	$K_{T,\alpha}^U$
15	−0.4468	0.4468	0.3992	1.4641	0.7908	2.0464	1.1879	2.6880	1.4373	3.1095	1.6584	3.4919
20	−0.3816	0.3816	0.4544	1.3579	0.8495	1.9101	1.2535	2.5162	1.5085	2.9139	1.7351	3.2743
25	−0.3387	0.3387	0.4925	1.2913	0.8905	1.8257	1.2997	2.4109	1.5586	2.7942	1.7891	3.1415
30	−0.3076	0.3076	0.5209	1.2447	0.9213	1.7672	1.3345	2.3382	1.5965	2.7120	1.8300	3.0504
40	−0.2647	0.2647	0.5613	1.1824	0.9654	1.6898	1.3845	2.2427	1.6510	2.6041	1.8889	2.9310
50	−0.2359	0.2359	0.5892	1.1418	0.9961	1.6398	1.4194	2.1814	1.6892	2.5349	1.9302	2.8546
60	−0.2148	0.2148	0.6100	1.1127	1.0191	1.6042	1.4457	2.1378	1.7179	2.4859	1.9613	2.8006
70	−0.1986	0.1986	0.6263	1.0906	1.0371	1.5772	1.4664	2.1050	1.7406	2.4490	1.9858	2.7599
80	−0.1855	0.1855	0.6396	1.0730	1.0518	1.5559	1.4833	2.0791	1.7591	2.4199	2.0059	2.7279
90	−0.1747	0.1747	0.6506	1.0586	1.0641	1.5385	1.4974	2.0580	1.7746	2.3963	2.0226	2.7019
100	0.1656	0.1656	0.6599	1.0466	1.0746	1.5240	1.5095	2.0404	1.7878	2.3766	2.0370	2.6802

approximation of the Pearson type 3 to normal quantile variance ratio η as (Stedinger et al., 1983)

$$K^L_{T,\alpha} = K_T + \eta(\zeta_{T,\,1-\alpha/2} - z_T) \qquad \text{and} \qquad K^U_{T,\alpha} = K_T + \eta(\zeta_{T,\alpha/2} - z_T) \quad (3.26)$$

where

$$\eta = \sqrt{\frac{1 + \hat{\gamma}_x K_T + \frac{1}{2}(1 + \frac{3}{4}\gamma_x)K_T^2 + n\,\mathrm{var}(\hat{\gamma}_x)(\partial K_T/\partial\gamma_x)^2}{1 + (1/2)z_T^2}} \qquad (3.27)$$

in which $\hat{\gamma}_x$ is the estimated skewness coefficient, and

$$\frac{\partial K_T}{\partial\gamma_x} \approx \frac{1}{6}(z_T^2 - 1)\left[1 - 3\left(\frac{\gamma_x}{6}\right)^2\right] + (z_T^3 - 6z_T)\frac{\gamma_x}{54} + \frac{2}{3}z_T\left(\frac{\gamma_x}{6}\right)^3 \qquad (3.28)$$

A simulation study by Whitley and Hromadka (1997) showed that the approximated formula for the Pearson type 3 distribution is relatively crude and that a better expression could be derived for more accurate confidence-interval determination.

Example 3.8 Referring to Example 3.3, determine the 95 percent confidence interval of the 100-year flood assuming that the sample data are from a lognormal distribution.

Solution In this case, with the 95 percent confidence interval $c = 0.95$, the corresponding significance level $\alpha = 0.05$. Hence $z_{0.025} = \Phi^{-1}(0.025) = -1.960$ and $z_{0.975} = \Phi^{-1}(0.975) = +1.960$. Computation of the 95 percent confidence interval associated with the selected return periods are shown in the table below. Column (4) lists the values of the upper tail of the standard normal quantiles associated with each return period, that is, $K_T = z_T = \Phi^{-1}(1 - 1/T)$. Since random floods are assumed to be lognormally distributed, columns (7) and (8) are factors computed by Eq. (3.25) for defining the lower and upper bounds of the 95 percent confidence interval of different quantiles in log-space, according to Eq. (3.24), as

$$y^L_{T,0.95} = \bar{y} + \zeta_{T,0.025} \times s_y \qquad y^U_{T,0.95} = \bar{y} + \zeta_{T,0.975} \times s_y$$

In the original space, the 95 percent confidence interval can be obtained simply by taking exponentiation as

$$q^L_{T,0.95} = \exp\left(y^L_{T,0.95}\right) \qquad \text{and} \qquad q^U_{T,0.95} = \exp\left(y^U_{T,0.95}\right)$$

as shown in columns (11) and (12), respectively. The curves defining the 95 percent confidence interval, along with the estimated frequency curve, for a lognormal distribution are shown in Fig. 3.5.
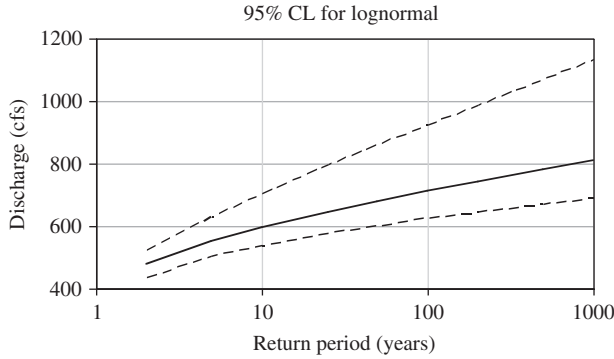
Figure 3.5 95 percent confidence limits for a lognormal distribution applied to the annual maximum discharge for 1961–1975 on the Boneyard Creek at Urbana, IL.

Return period T (years) (1)	Exceedance probability $1 - p = 1/T$ (2)	Nonexceedance probability $p = 1 - 1/T$ (3)	$K_T = z_T$ (4)	y_T (5)	q_T (6)
2	0.5	0.5	0.0000	6.165	475.9
5	0.2	0.8	0.8416	6.311	550.3
10	0.1	0.9	1.2816	6.386	593.7
25	0.04	0.96	1.7505	6.467	643.8
50	0.02	0.98	2.0537	6.520	678.3
100	0.01	0.99	2.3263	6.567	711.0

Return period T (years)	$\zeta_{T,0.025}$ (7)	$\zeta_{T,0.975}$ (8)	$y_{T,0.95}^{L}$ (9)	$y_{T,0.95}^{U}$ (10)	$q_{T,0.95}^{L}$ (11)	$q_{T,0.95}^{U}$ (12)
2	−0.54	0.54	6.071	6.259	433.2	522.8
5	0.32	1.63	6.221	6.446	503.1	630.4
10	0.71	2.26	6.288	6.555	538.1	702.9
25	1.10	2.96	6.355	6.676	575.5	792.8
50	1.34	3.42	6.397	6.755	600.1	858.2
100	1.56	3.83	6.434	6.827	622.8	922.2

In order to define confidence limits properly for the Pearson type 3 distribution, the skewness coefficient must be estimated accurately, thus allowing the frequency factor K_T to be considered a constant and not a statistic. Unfortunately, with the Pearson type 3 distribution, no simple, explicit formula is available for the confidence limits. The Interagency Advisory Committee on Water Data (1982) (hereafter referred to as "the Committee") proposed that the confidence limits for the log-Pearson type 3 distribution could be approximated using a noncentral t-distribution. The committee's procedure is similar to that of Eqs. (3.24) and (3.25), except that $K_{T,\alpha}^{L}$ and $K_{T,\alpha}^{U}$, the *confidence-limit factors* for the lower and upper limits, are computed with the frequency factor K_T replacing Z_T in Eq. (3.25).

Example 3.9 Referring to Example 3.3, determine the 95 percent confidence intervals for the 2-, 10-, 25-, 50-, and 100-year floods assuming that the sample data are from a log-Pearson type 3 distribution.

Solution From Example 3.3, the mean and standard deviation of the logarithms of the peak flows were 6.17 and 0.173, and the number of data n is 15. For the 100-year flood, K_T is 1.8164, and for the 95 percent confidence limits, a is 0.05; thus $Z_{a/2}$ is -1.96. Thus $K_{T,\alpha}^L$ is -0.651, and the lower 95 percent confidence bound is 427.2 ft^3/s. The upper and lower confidence bounds for all the desired flows are listed in the following table:

Return Period T (years)	K_T Eq. (3.8)	$K_{T,0.025}$ Eq. (3.26)	$K_{T,0.975}$ Eq. (3.26)	$q_{T,0.95}^L$ (ft^3/s)	$q_{T,0.95}^U$ (ft^3/s)
2	-0.0907	-0.6513	0.4411	427.2	516.1
10	1.0683	0.5260	1.9503	523.7	670.1
25	1.4248	0.8322	2.4705	552.2	733.2
50	1.6371	1.0082	2.7867	569.3	774.4
100	1.8164	1.1540	3.0565	583.8	811.4

3.9 Limitations of Hydrologic Frequency Analysis

3.9.1 Distribution Selection: Practical Considerations

Many different probability distributions have been proposed for application to hydrologic data. Some of them were proposed because the underlying concept of the distribution matched the goal of hydrologic frequency analysis. For example, the extremal distributions discussed in Sec. 2.6.4 have very favorable properties for hydrologic frequency analysis. Ang and Tang (1984, p. 206) noted that the asymptotic distributions of extremes in several cases tend to converge on certain limiting forms for large sample sizes n, specifically to the double-exponential form or to two single-exponential forms. The extreme value from an initial distribution with an exponentially decaying tail (in the direction of the extreme) will converge asymptotically to the extreme-value type I (Gumbel) distribution form. Distributions with such exponentially decaying tails include the normal distribution and many others listed in Sec. 2.6. This is why Gumbel (1941) first proposed this distribution for floods, and it has gained considerable popularity since then. Also, the properties of the central limit theorem discussed in Sec. 2.6.2 have made the lognormal distribution a popular choice for hydrologic frequency analysis.

In the 1960s, as the number of different approaches to flood frequency analysis were growing, a working group of U.S. government agency hydrologic experts was formed by the U.S. Water Resources Council to evaluate the best/preferred approach to flood frequency analysis. Benson (1968) reviewed the results of this working group and listed the following key results of their study:

1. There is no physical rule that requires the use of any specific distribution in the analysis of hydrologic data.

2. Intuitively, there is no reason to expect that a single distribution will apply to all streams worldwide.

3. No single method of testing the computed results against the original data was acceptable to all those on the working group, and the statistical consultants could not offer a mathematically rigorous method.

Subsequent to this study, the U.S. Water Resources Council (1967) recommended use of the log-Pearson type 3 distribution for all flood frequency analyses in the United States, and this has become the official distribution for all flood frequency studies in the United States. There are no physical arguments for the application of this distribution to hydrologic data. It has added flexibility over two-parameter distributions (e.g., Gumbel, lognormal) because the skewness coefficient is a third independent parameter, and the use of three parameters generally results in a better fit of the data. However, a number of researchers have suggested that the use of data for a single site may be insufficient to estimate the skewness coefficient properly.

Beard (1962) recommended that only average regional skewness coefficients should be applied in flood frequency analysis for a single station unless that record exceeds 100 years. This led the U.S. Water Resources Council (1967) to develop maps of regional skewness coefficient values that are averaged with the at-a-site skewness coefficient as a function of the number of years of record. For details on the procedure, see Interagency Advisory Committee on Water Data (1982). Linsley et al. (1982) noted that although regional skewness coefficients may not make for more reliable analysis, their use does lead to more consistency between values for various streams in the region.

3.9.2 Extrapolation problems

Most often frequency analysis is applied for the purpose of estimating the magnitude of truly rare events, e.g., a 100-year flood, on the basis of short data series. Viessman et al. (1977, pp. 175–176) note that "as a general rule, frequency analysis should be avoided . . . in estimating frequencies of expected hydrologic events greater than twice the record length." This general rule is followed rarely because of the regulatory need to estimate the 100-year flood; e.g., the U.S. Water Resources Council (1967) gave its blessing to frequency analyses using as few as 10 years of peak flow data. In order to estimate the 100-year flood on the basis of a short record, the analyst must rely on extrapolation, wherein a law valid inside a range of p is assumed to be valid outside of p. The dangers of extrapolation can be subtle because the results may look plausible in the light of the analyst's expectations.

The problem with extrapolation in frequency analysis can be referred to as "the tail wagging the dog." In this case, the "tail" is the annual floods of relatively high frequency (1- to 10-year events), and the "dog" is the estimation of extreme floods needed for design (e.g., the floods of 50-, 100-, or even higher-year return periods). When trying to force data to fit a mathematical distribution,

equal weight is given to the low end and high end of the data series when trying to determine high-return-period events. Figure 3.6 shows that small changes in the three smallest annual peaks can lead to significant changes in the 100-year peak owing to "fitting properties" of the assumed flood frequency distribution. The analysis shown in Fig. 3.6 is similar to the one presented by Klemes (1986); in this case, a 26-year flood series for Gilmore Creek at Winona, Minnesota, was analyzed using the log-Pearson type 3 distribution employing the skewness coefficient estimated from the data. The three lowest values in the annual maximum series (22, 53, and 73 ft^3/s) then were changed to values of 100 ft^3/s (as if a crest-stage gauge existed at the site with a minimum flow value of 100 ft^3/s), and the log-Pearson type 3 analysis was repeated. The relatively small absolute change in these three events changed the skewness coefficient from 0.039 to 0.648 and the 100-year flood from 7,030 to 8,530 ft^3/s. As discussed by Klemes (1986), it is illogical that the 1- to 2-year frequency events should have such a strong effect on the rare events.

Under the worst case of hydrologic frequency analysis, the frequent events can be caused by a completely different process than the extreme events. This situation violates the initial premise of hydrologic frequency analysis, i.e., to find some statistical relation between the magnitude of an event and its likelihood of occurrence (probability) without regard for the physical process of flood formation. For example, in arid and semiarid regions of Arizona, frequent events (1- to 5-year events) are caused by convective storms of limited spatial extent, whereas the major floods (>10-year events) are caused by frontal
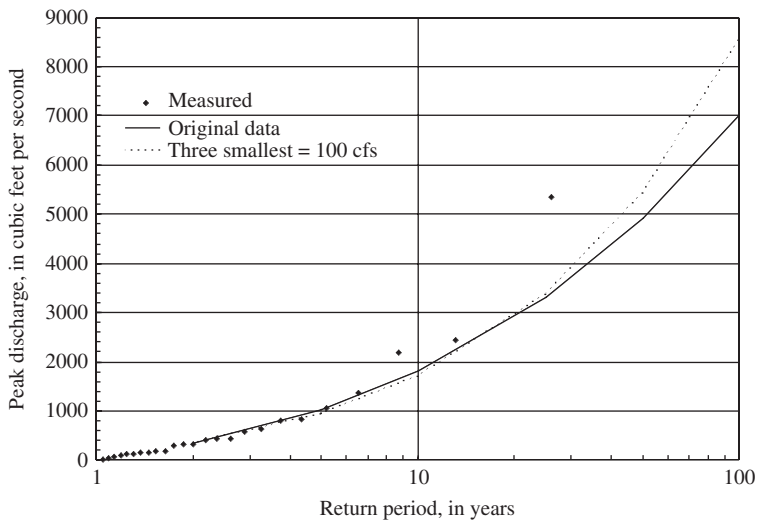
Figure 3.6 Flood frequency analysis for Gilmore Creek at Winona, Minnesota, for 1940–1965 computed with the log-Pearson type 3 distribution fitted to (1) the original annual maximum series and (2) to the original annual maximum series with the three smallest annual peaks set to 100 ft^3/s.

monsoon-type storms that distribute large amounts of rainfall over large areas for several days. Figure 3.7 shows the daily maximum discharge series for the Agua Fria River at Lake Pleasant, Arizona, for 1939–1979 and clearly indicates a difference in magnitude and mechanism between frequent and infrequent floods. In this case estimating the 100-year flood giving equal weight in the statistical calculations to the 100 ft^3/s and the 26,000 ft^3/s flows seems inappropriate, and an analyst should be prepared to use a large safety factor if standard frequency analysis methods were applied.

Another problem with "the tail wagging the dog" results when the watershed experiences substantial changes. For example, in 1954 the Vermilion River, Illinois, Outlet Drainage District initiated a major channelization project involving the Vermilion River, its North Fork, and North Fork tributaries. The project was completed in the summer of 1955 and resulted in changing the natural 35-ft-wide North Fork channel to a trapezoidal channel 100 ft in width and the natural 75-ft-wide Vermilion channel to a trapezoidal channel 166 ft in width. Each channel also was deepened 1 to 6 ft (U.S. Army Corps of Engineers, 1986). Discharges less than about 8,500 ft^3/s at the outlet remain in the modified channel, whereas those greater than 8,500 ft^3/s go overbank. At some higher discharge, the overbank hydraulics dominate the flow, just as they did before the channelization. Thus the more frequent flows are increased by the improved hydraulic efficiency of the channel, whereas the infrequent events are still subject to substantial attenuation by overbank flows. Thus the frequency curve is flattened relative to the pre-channelization condition, where the more frequent events are also subject to overbank attenuation. The pre- and post-channelization flood frequency curves cross in the 25- to 50-year return period
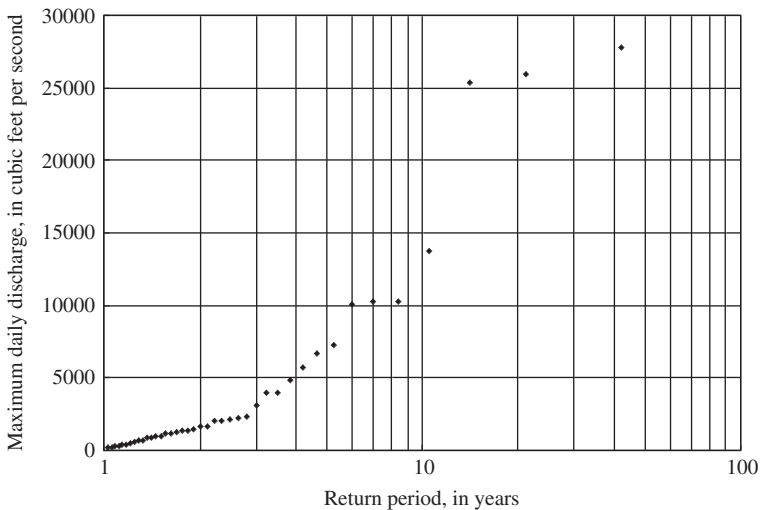
Figure 3.7   Return periods for the annual maximum daily flow of the Agua Fria River at Lake Pleasant, Arizona, for 1939–1979.
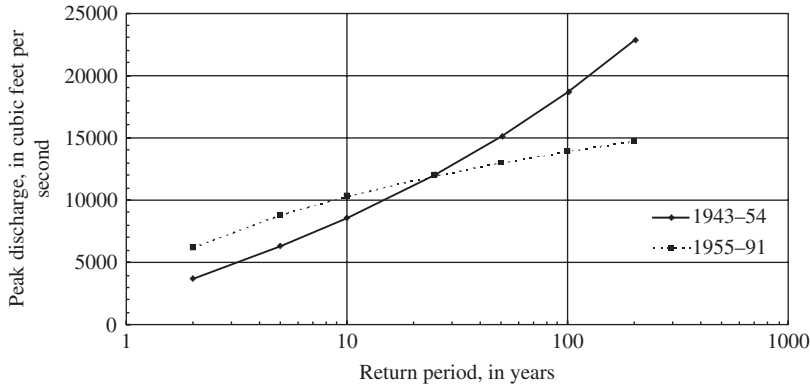
Figure 3.8   Peak discharge frequency for the Vermilion River at Pontiac, Illinois, for pre-channelized (1943–1954) and post-channelized (1955–1991) conditions.

range (Fig. 3.8), resulting in the illogical result that the pre-channelization condition results in a higher 100-year flood than the post-channelization condition. Similar results have been seen for flood flows obtained from continuous simulation applied to urbanizing watersheds (Bradley and Potter, 1991).

3.9.3   The stationarity assumption

Viessman et al. (1977, p. 158) noted that "usually, the length of record as well as the design life for an engineering project are relatively short compared with geologic history and tend to temper, if not justify, the assumption of stationarity." On the other hand, Klemes (1986) noted that there are many known causes for nonstationarity ranging from the dynamics of the earth's motion to human-caused changes in land use. In this context, Klemes (1986) reasons that the notion of a 100-year flood has no meaning in terms of average return period, and thus the 100-year flood is really a reference for design rather than a true reflection of the frequency of an event.

3.9.4   Summary comments

The original premise for the use of hydrologic frequency analysis was to find the optimal project size to provide a certain protection level economically, and the quality of the optimization is a function of the accuracy of the estimated flood level. The preceding discussions in this section have indicated that the accuracy of hydrologic frequency estimates may not be high. For example, Beard (1987) reported that the net result of studies of uncertainties of flood frequency analysis is that standard errors of estimated flood magnitudes are very high—on the order of 10 to 50 percent depending on the stream characteristics and amount of data available.

Even worse, the assumptions of hydrologic frequency analysis, namely, stationarity and homogeneous, representative data, and good statistical

modeling—not extrapolating too far beyond the range of the data—may be violated or stretched in common practice. This can lead to illogical results such as the crossing of pre- and post-change frequency curves illustrated in Fig. 3.8, and the use of such illogical results is based on "a subconscious hope that nature can be cheated and the simple logic of mathematical manipulations can be substituted for the hidden logic of the external world" (Klemes, 1986).

Given the many potential problems with hydrologic frequency analysis, what should be done? Klemes (1986) suggested that if hydrologic frequency theorists were good engineers, they would adopt the simplest procedures and try to standardize them in view of the following facts:

1. The differences in things such as plotting positions, parameter-estimation methods, and even the distribution types, may not matter much in design optimization (Slack et al., 1975). Beard (1987) noted that no matter how reliable flood frequency estimates are, the actual risk cannot be changed. Thus the benefits from protection essentially are a function of investment and are independent of uncertainties in estimating flood frequencies. Moderate changes in protection or zoning do not change net benefits greatly; i.e., the benefit function has a broad, flat peak (Beard, 1987).

2. There are scores of other uncertain factors in the design that must be settled, but in a rather arbitrary manner, so the whole concept of optimization must be taken as merely an expedient design procedure. The material covered in Chaps. 4, 6, 7, and 8 of this book provide methods to consider the other uncertain factors and improve the optimization procedure.

3. Flood frequency analysis is just one convenient way of rationalizing the old engineering concept of a safety factor rather than a statement of hydrologic truth.

Essentially, the U.S. Water Resources Council (1967) was acting in a manner similar to Klemes' approach in that a standardized procedure was developed and later improved (Interagency Advisory Committee on Water Data, 1982). However, rather than selecting and standardizing a simple procedure, the relatively more complex log-Pearson type 3 procedure was selected. Beard (1987) suggested that the U.S. Water Resources Council methods are the best currently available but leave much to be desired.

Problems

Given are the significant independent peak discharges measured on the Saddle River at Lodi, NJ, for two 18-year periods 1948–1965 and 1970–1987. The Saddle River at Lodi has a drainage area of 54.6 mi^2 primarily in Bergen County. The total data record for peak discharge at this gauge is as follows: 1924–1937 annual peak only, 1938–1987 all peaks above a specified base value, 1988–1989 annual peak only (data are missing for 1966, 1968, and 1969, hence the odd data periods).

Water year	Date	Q_p (ft³/s)
1948	11/09/47	830
1949	12/31/48	1030
1950	3/24/50	452
1951	3/31/51	2530
1952	12/21/51	1090
	3/12/52	1100
	4/06/52	1470
	6/02/52	1740
1953	3/14/53	1860
	3/25/53	993
	4/08/53	1090
1954	9/12/54	1270
1955	8/19/55	2200
1956	10/16/55	1530
1957	11/02/56	795
	4/06/57	795
1958	1/22/58	964
	2/28/58	1760
	4/07/58	1100
1959	3/07/59	795
1960	9/13/60	1190
1961	2/26/61	952
1962	3/13/62	1670
1963	3/07/63	824
1964	1/10/64	702
1965	2/08/65	1490
	8/10/65	1020
1970	2/11/70	1770
	4/03/70	2130
1971	8/28/71	3530
	9/12/71	3770
1972	6/19/72	2240
1973	11/09/72	2450
	2/03/73	3210
	6/30/73	1570
1974	12/21/73	2940
1975	5/15/75	2640
	7/14/75	2720
	9/27/75	2350
1976	4/01/76	1590
	7/01/76	2440
1977	2/25/77	3130
	3/23/77	2380
1978	11/09/77	4500
	1/26/78	1980
	3/27/78	1610
1979	1/21/79	2890
	2/26/79	1570
	5/25/79	1760
1980	3/22/80	1840
	4/10/80	2470
	4/29/80	2370
1981	2/20/81	1540
	5/12/81	1900
1982	1/04/82	1980
1983	3/28/83	1800
	4/16/83	2550
1984	10/24/83	1510
	12/13/83	2610
	4/05/84	3350
	5/30/84	2840
	7/07/84	2990
1985	4/26/85	1590
	9/27/85	2120
1986	1/26/86	1850
	8/17/86	1660
1987	12/03/86	2310
	4/04/87	2320

3.1 Determine the annual maximum series.

3.2 Plot the annual maximum series on normal, lognormal, and Gumbel probability papers.

3.3 Calculate the first four product moments and L-moments based on the given peak-flow data in both the original and logarithmic scales.

3.4 Use the frequency-factor approach to the Gumbel, lognormal, and log-Pearson type 3 distributions to determine the 5-, 25-, 50-, and 100-year flood peaks.

3.5   Based on the L-moments obtained in Problem 3.3, determine the 5-, 25-, 50-, and 100-year flood peaks using Gumbel, generalized extreme value (GEV), and lognormal distributions.

3.6   Determine the best-fit distribution for the annual maximum peak discharge series based on the probability-plot correlation coefficient, the two model reliability indices, and L-moment ratio diagram.

3.7   Establish the 95 percent confidence interval for the frequency curve derived based on lognormal and log-Pearson type 3 distribution models.

References

Abramowitz, M., and Stegun, I. A. (1967). Probability functions, in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Washington.

Ang, A. H.-S., and Tang, W. H. (1984). Probability Concepts in Engineering Planning and Design: Decision, Risk and Reliability, Vol. II: Decision, Risk, and Reliability, John Wiley and Sons, New York.

Beard, L. R. (1962). *Statistical Methods in Hydrology*, Civil Works Investigations, U.S. Army Corps of Engineers, Sacramento District.

Benson, M. A. (1968). Uniform flood-frequency estimating methods for federal agencies, *Water Resources Research*, 4(5):891–908.

Benson, M. A. (1968). Uniform flood-frequency analysis of simulated flows, *Water Resources Research*, 4(5):891–908.

Benson, M. A. (1952). Characteristics of frequency curves based on a theoretical 1000-year record, U.S. Geological Survey Water-Supply Paper 1543-A, pp. 51–74.

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variates*. John Wiley and Sons, New York.

Bradley, A. A. and Potter, K. W. (1991). Flood frequency analysis for evaluating watershed conditions with rainfall-runoff models, *Water Resources Bulletin*, 27(1):83–91.

California State Department (1923). Flow in California streams, in *Public Works Bulletin*, 5.

Chow, V. T. (1953). Frequency analysis of hydrology data with special application to rainfall intensities, *University of Illinois Engineering Experiment Station Bulletin*, 414.

Chow, V. T. (1954). The log-probability law and its engineering applications, *Proceedings*, ASCE, 80:1–25.

Chow, V. T. (1964). Statistical and probability analysis of hydrological data. In: Section 8-1, *Handbook of Applied Hydrology*, ed. by V. T. Chow, McGraw-Hill, New York.

Chowdhury, J. U., Stedinger, J. R., and Lu, L.-H., (1991). Goodness-of-fit tests for regional GEV flood distributions, *Water Resources Research*, 27(7):811–831.

Cunnane, C. (1978). Unbiased plotting-positions: A review, *Journal of Hydrology*, 37(3/4):205–222.

D'Agostino, R. B., and Stephens, M. A. (1986). *Goodness-of-Fit Procedures*, Marcel Dekker, New York.

Engineering Staff of Miami Conservancy District (1937). Storm rainfall of eastern U.S., Technical Report, Part V; Miami Conservancy District.

Filliben, J. J. (1975). The probability plot correlation test for normality, *Technometrics*, 17(1):111–117.

Foster, H. A. (1935). Duration curves, *Transactions*, ASCE, 99:1213–1235.

Frederick, R. H., Myers, V. A., and Auciello, E. P. (1977). Five to 60-minute precipitation frequency for the eastern and central United States, *NOAA Technical Memo*, NWS HYDRO-35.

Fuller, W. E. (1914). Flood flows, *Transactions*, ASCE, 77:564–617.

Gringorten, I. I. (1963). A plotting rule for extreme probability paper, *Journal of Geophysical Research*, 68(3):813–814.

Gumbel, E. J. (1941). The return period of flood flows, *American Mathematical Statistics*, XII(2):163–190.

Gumbel, E. J. (1942). Statistical control curves for flood discharges, *Transactions*, AGU, 23:489–500.

Gumbel, E. J. (1954). The statistical theory of droughts, *Proceedings*, ASCE, 80:1–19.

Gumbel, E. J. (1963). Statistical forecast of droughts, *Bulletin*, International Association of the Science of Hydrology, 8th year, 1:5–23.

Haan, C. T. (1977). *Statistical Methods in Hydrology*, Iowa State University Press, Ames, IA.

Hazen, A. (1930). *Flood Flows: A Study of Frequencies and Magnitudes*. John Wiley and Sons, New York.

Hershfield, D. M. (1961). Rainfall frequency atlas of the U.S. for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, Technical Paper No. 40, U.S. Weather Bureau.

Hosking, J. R. M. (1986). The theory of probability-weighted moments, IBM Research Report No. 12210, October.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985a). Estimation of the generalized extreme value distribution by the method of probability-weighted moments, *Technometrics*, 27(3):251–261.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985b). An Appraisal of the regional flood frequency procedure in the U.K. flood studies report, *Hydrological Sciences Journal*, 30(1):85–109.

Huff, F. A., and Angel, J. R. (1989). Frequency distributions and hydroclimatic characteristics of heavy rainstorms in Illinois, *Illinois State Water Survey Bulletin 70*, Illinois State Water Survey, Champaign, IL.

Huff, F. A., and Angel, J. R. (1992). Rainfall frequency atlas of the Midwest, *Illinois State Water Survey Bulletin 71* and *Midwestern Climate Center Research Report 92-03*, Illinois State Water Survey, Champaign, IL.

Interagency Advisory Committee on Water Data (1982). Guidelines for determining flood flow frequency, Bulletin 17-B, Hydrology Subcommittee, Washington.

Kite, G. W. (1988). *Frequency and Risk Analysis in Hydrology*, Water Resources Publications, Littleton, CO.

Klemes, V. (1986). Dilettantism in hydrology: Transition or destiny? *Water Resources Research*, 22(9):177S–188S.

Leggett, R. W., and Williams, L. R. (1981). A reliability index for models, *Ecological Modelling*, 13:303–312.

Leivikov, M. L. (1955). *Meteorology, Hydrology, and Hydrometry*, 2d ed., Sel'khozgiz, Moscow.

Li, Y., Singh, V. P., and Cong, S. (1986). Entropy and probability distributions, in *Hydrologic Frequency Modeling*, ed. by V. P. Singh, D. Reidel Publishing Company, Drodrecht, The Netherlands, pp. 367–382.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H. (1982). *Hydrology for Engineers*, 2nd ed., McGraw-Hill, New York.

Looney, S. W., and Gulledge, T. R., Jr. (1985). Use of the correlation coefficient with normal probability plots, *American Statistician*, 39(1):75–79.

McCormick, G. P. (1983). *Nonlinear Programming: Theory, Algorithm, and Applications*, John Wiley and Sons, New York.

Miller, J. F., Frederick, R. H., and Tracey, R. J. (1973). Precipitation-frequency atlas for the conterminous western United States (by states), *NOAA Atlas 2*, NWS, 11 vols.

National Environment Research Council (1975). *Hydrological Studies*, Vol. 1 of *Flood Studies Report*, National Environment Research Council, London.

Pandey, M. D., van Gelder, P. H. A. J. M., and Vrijling, J. K. (2001). Assessment of an L-kurtosis-based criterion for quantile estimation, *Journal of Hydrologic Engineering*, ASCE, 6(4):284–292.

Patel, J. K., Kapadia, C. H., and Owen, D. B. (1976). *Handbook of Statistical Distributions*, Marcel Dekker, New York, pp. 1–59.

Potter, K. W., and Walker, J. F. (1981) A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements, *Water Resources Research*, 17(5):1505–1509.

Rao, A. R., and Hamed, K. H. (2000). *Flood Frequency Analysis*. CRC Press, Boca Raton, FL.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, CO.

Salas, J. D. (1993). Analysis and modeling of hydrologic time series, in *Handbook of Hydrology*, ed. by D. R. Maidment, McGraw-Hill, New York.

Slack, J. R., Wallis, J. R., and Matalas, N. C. (1975). On the value of information to flood frequency analysis, *Water Resources Research*, 11(5):629–647.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events, *Handbook of Hydrology*, ed. by D. R. Maidment, McGraw-Hill, New York.

Stuart, A., and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1: *Distribution Theory*, Oxford University Press, New York.

Tung, Y. K., and Yen, B. C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, McGraw-Hill, New York.

Tukey, J. W. (1962). The future of data analysis, *Annals of Mathematical Statistics*, 33(1):1–67.

U.S. Army Corps of Engineers (1986). *Feasibility Study for Vermilion River Basin, Illinois*, Rock Island District, Rock Island, IL.

U.S. Army Corps of Engineers (1996). Risk-based analysis for flood damage reduction studies, *Engineering Manual EM 1110-2-1619*, Washington.

U.S. Department of Agriculture (1955). *Yearbook*. U.S. Department of Agriculture, Washington.

U.S. Water Resources Council (1967). A uniform technique for determining flood flow frequencies, *Bulletin No. 15*.

U.S. Weather Bureau (1964). Two-to-ten day precipitation for return periods of 2 to 100 years in the contiguous U.S., Technical Paper No. 49.

Viessman, W., Jr., Knapp, J. W., Lewis, G. K., and Harbaugh, T. E. (1977). *Introduction to Hydrology*, 2d ed., IEP, New York.

Vogel, R. M. (1986). The probability plot correlation coefficient test for normal, lognormal, and Gumbel distributional hypothesis, *Water Resources Research*, 22(4):587–590, corrections, 23(10):2013.

Vogel, R. M. and Kroll, C. N. (1989). Low-flow frequency analysis using probability plot correlation coefficients, *Journal of Water Resources Planning and Managment*, ASCE, 115(3):338–357.

Vogel, R. M., and McMartin, D. E. (1991). Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution, *Water Resources Research*, 27(12):3149–3158.

Weibull, W. (1939). A statistical theory of the strength of materials, Royal Swedish Institute Engineering Research Proceedings, 151:1–45.

Whitley, R., and Hromadka, T. V., II. (1997). Chowdhury and Stedinger's approximate confidence intervals for design floods for a single site, *Stochastic Hydrology and Hydraulics*, 11:51–63.

Yarnell, D. L. (1936). Rainfall intensity-frequency data, U.S. Department of Agriculture, Misc. Publ. 204.

Yen, B. C. (1970). Risks in hydrologic design of engineering project, *Journal of Hydraulic Engineering*, ASCE, 96(HY4):959–966.

Reliability Analysis Considering Load-Resistance Interference

4.1 Basic Concept

The design of a hydrosystem involves analyses of flow processes in hydrology and hydraulics. In a multitude of hydrosystems engineering problems, uncertainties in data and in theory, including design and analysis procedures, warrant a probabilistic treatment of the problems. The risk associated with the potential failure of a hydrosystem is the result of the combined effects of inherent randomness of external loads and various uncertainties involved in the analysis, design, construction, and operational procedures. Hence, to evaluate the probability that a hydrosystem will function as designed requires uncertainty and reliability analyses.

As discussed in Sec. 1.5, failure of an engineering system can be defined as the load L (external forces or demands) on the system exceeding the resistance R (strength, capacity, or supply) of the system. The reliability p_s is defined as the *probability of safe* (or nonfailure) operation, in which the resistance of the structure exceeds or equals to the load, that is,

$$p_s = P(L \leq R) \tag{4.1}$$

in which $P(\cdot)$ denotes the probability. Conversely, *failure probability p_f* can be computed as

$$p_f = P(L > R) = 1 - p_s \tag{4.2}$$

The definitions of reliability and failure probability, Eqs. (4.1) and (4.2), are equally applicable to component reliability, as well as total system

145

reliability. In hydrosystems engineering analyses, the resistance and load frequently are functions of several stochastic basic variables, that is, $L = g(\boldsymbol{X}_L) = g(X_1, X_2, \ldots, X_m)$ and $R = h(\boldsymbol{X}_R) = h(X_{m+1}, X_{m+2}, \ldots, X_K)$, where $X_1, X_2, \ldots, X_K$ are stochastic basic variables defining the load function $g(\boldsymbol{X}_L)$ and the resistance function $h(\boldsymbol{X}_R)$. Accordingly, the failure probability and reliability are functions of stochastic basic variables, that is,

$$p_s = P\left[g(\boldsymbol{X}_L) \leq h(\boldsymbol{X}_R)\right] \tag{4.3}$$

Note that the foregoing presentation of load and resistance in reliability analysis should be interpreted in a very general context. For example, in the design and analysis hydrosystems infrastructures, such as urban drainage systems, the load could be the inflow to the sewer system, whereas the resistance is the sewer conveyance capacity; in water quality assessment, the load may be the concentration or mass of pollutant entering the environmental system, whereas the resistance is the permissible pollutant concentration set by water quality regulations; in the economic analysis of a hydrosystem, the load could be the total cost, whereas the resistance is the total benefit.

Evaluation of reliability or failure probability by Eqs. (4.1) through (4.3) does not consider the time-dependent nature of the load and resistance if statistical properties of the elements in $\boldsymbol{X}_L$ and $\boldsymbol{X}_R$ do not change with time. This procedure generally is applied when the performance of the system subject to a single worst-load event is considered. From the reliability computation viewpoint, this is referred to as *static reliability analysis*.

In general, a hydrosystem infrastructure is expected to serve its designated function over an expected period of time. Engineers frequently are interested in knowing the reliability of the structure over its intended service life. In such circumstances, elements of service period, randomness of load occurrences, and possible change in resistance characteristics over time must be considered. Reliability models incorporating these elements are called *time-dependent reliability models* (Kapur and Lamberson, 1977; Tung and Mays, 1980; Wen, 1987). Computations of the time-dependent reliability of a hydrosystem infrastructure initially require the evaluation of static reliability. Sections 4.3 through 4.6 describe methods for static reliability analysis, and Sec. 4.7 briefly describes some basic methods for dealing with the time-dependent nature of reliability analysis.

As discussed in the preceding chapters, the natural randomness of hydrologic and geophysical variables, such as flood and precipitation, is an important part of the uncertainty in the design of hydrosystems infrastructures. However, other uncertainties also may be significant and should not be ignored. Failure to account for the other uncertainties in the reliability analysis in the past (as discussed in Sec. 1.3) hindered progress in evaluation of failure probability associated with hydrosystems infrastructures. As noted by Cornell (1969) with respect to traditional frequency-based analyses of system safety:

It is important in engineering applications that we avoid the tendency to model only those probabilistic aspects that we think we know how to analyze. It is far better to have an approximated model of the whole problem than an exact model of only a portion of it.

4.2   Performance Functions and Reliability Index

In reliability analysis, Eq. (4.3) alternatively can be written in terms of a performance function $W(\boldsymbol{X}) = W(\boldsymbol{X}_L, \boldsymbol{X}_R)$ as

$$p_s = P\left[W(\boldsymbol{X}_L, \boldsymbol{X}_R) \geq 0\right] = P\left[W(\boldsymbol{X}) \geq 0\right] \tag{4.4}$$

in which $\boldsymbol{X}$ is the vector of basic stochastic variables in the load and resistance functions. In reliability analysis, the system state is divided into the safe (satisfactory) set defined by $W(\boldsymbol{X}) > 0$ and the failure (unsatisfactory) set defined by $W(\boldsymbol{X}) < 0$ (Fig. 4.1). The boundary that separates the safe set and failure set is a surface, called the *failure surface*, defined by the function $W(\boldsymbol{X}) = 0$, called the *limit-state function*. Since the performance function $W(\boldsymbol{X})$ defines the condition of the system, it is sometimes called *system-state function*.
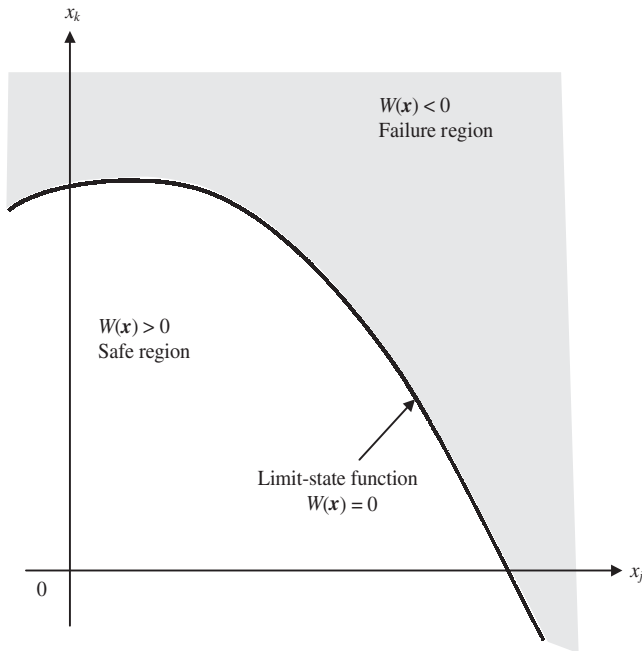
Figure 4.1 System states defined by performance (limit-state) function.

The performance function $W(\boldsymbol{X})$ can be expressed differently as

$$W_1(\boldsymbol{X}) = R - L = h(\boldsymbol{X}_R) - g(\boldsymbol{X}_L) \tag{4.5}$$

$$W_2(\boldsymbol{X}) = (R/L) - 1 = [h(\boldsymbol{X}_R)/g(\boldsymbol{X}_L)] - 1 \tag{4.6}$$

$$W_3(\boldsymbol{X}) = \ln(R/L) = \ln[h(\boldsymbol{X}_R)] - \ln[g(\boldsymbol{X}_L)] \tag{4.7}$$

Referring to Sec. 1.6, Eq. (4.5) is identical to the notion of a safety margin, whereas Eqs. (4.6) and (4.7) are based on safety factor representations.

Example 4.1   Consider the design of a storm sewer system. The sewer flow-carrying capacity Q_C (ft^3/s) is determined by Manning's formula:

$$Q_C = \frac{0.463}{n} \lambda_c \, D^{8/3} \, S^{1/2}$$

where n is Manning's roughness coefficient, λ_c is the model correction factor to account for the model uncertainty, D is the actual pipe diameter (ft), and S is the pipe slope (ft/ft). The inflow Q_L (ft^3/s) to the sewer is the surface runoff whose peak discharge can be estimated by the rational formula

$$Q_L = \lambda_L C i A$$

in which λ_L is the correction factor for model uncertainty, C is the runoff coefficient, i is the rainfall intensity (in/h), and A is the runoff contributing area (acres). In the reliability analysis, the sewer flow-carrying capacity Q_C is the resistance, and the peak discharge of the surface runoff Q_L is the load. The performance functions can be expressed as one of the following three forms:

$$W_1 = Q_C - Q_L = \frac{0.463}{n} \lambda_c D^{8/3} S^{1/2} - \lambda_L C i A$$

$$W_2 = \frac{Q_C}{Q_L} - 1 = \frac{0.463}{n} \lambda_c D^{8/3} S^{1/2} \lambda_L^{-1} C^{-1} i^{-1} A^{-1} - 1$$

$$W_3 = \ln\left(\frac{Q_C}{Q_L}\right) = \ln(0.463) - \ln(n) + \ln(\lambda_c) + \frac{8}{3}\ln(D) + \frac{1}{2}\ln(S) - \ln(\lambda_L)$$
$$- \ln(C) - \ln(i) - \ln(A)$$

Also in the reliability analysis, a frequently used reliability indicator β is called the *reliability index*. The reliability index was first introduced by Cornell (1969) and later formalized by Ang and Cornell (1974). It is defined as the ratio of the mean to the standard deviation of the performance function $W(\boldsymbol{X})$, which is the inverse of the coefficient of variation of the performance function $W(\boldsymbol{X})$,

$$\beta = \frac{\mu_w}{\sigma_w} \tag{4.8}$$

in which μ_w and σ_w are the mean and standard deviation of the performance function, respectively. From Eq. (4.8), assuming an appropriate probability

density function for the random performance function $W(\mathbf{X})$, the reliability then can be computed as

$$p_s = 1 - F_w(0) = 1 - F_{w'}(-\beta) \tag{4.9}$$

in which $F_w(\cdot)$ is the cumulative distribution function of the performance function W, and W' is the standardized performance function defined as $W' = (W - \mu_w)/\sigma_w$. The expressions of reliability p_s for some distributions of $W(\mathbf{X})$ are given in Table 4.1. For distributions not listed, expressions can be found in Sec. 2.6. For practically all probability distributions used in the reliability analysis, the value of the reliability p_s is a strictly increasing function of the reliability index β. In practice, the normal distribution is used commonly for $W(\mathbf{X})$, in which case the reliability can be computed simply as

$$p_s = 1 - \Phi(-\beta) = \Phi(\beta) \tag{4.10}$$

where $\Phi(\cdot)$ is the standard normal CDF the table for which is given in Table 2.2. Without using the normal probability table, the value of $\Phi(\cdot)$ can be computed by various algebraic formulas described in Sec. 2.6.1.

4.3 Direct Integration Method

From Eqs. (4.1) and (4.4) one realizes that the computation of reliability requires knowledge of the probability distributions of the load and resistance or of the performance function W. In terms of the joint PDF of the load and resistance, Eq. (4.1) can be expressed as

$$p_s = \int_{r_1}^{r_2} \left[\int_{\ell_1}^{r} f_{R,L}(r, \ell)\,d\ell \right] dr \tag{4.11a}$$

$$= \int_{\ell_1}^{\ell_2} \left[\int_{\ell}^{r_2} f_{R,L}(r, \ell)\,dr \right] d\ell \tag{4.11b}$$

in which $f_{R,L}(r, \ell)$ is the joint PDF of random load L and resistance R, r and ℓ are dummy arguments for the resistance and load, respectively, and (r_1, r_2) and (ℓ_1, ℓ_2) are the lower and upper bounds for the resistance and load, respectively. The failure probability can be computed as

$$p_f = 1 - p_s = \int_{r_1}^{r_2} \left[\int_{r}^{\ell_2} f_{R,L}(r, \ell)\,d\ell \right] dr \tag{4.12a}$$

$$= \int_{\ell_1}^{\ell_2} \left[\int_{r_1}^{\ell} f_{R,L}(r, \ell)\,dr \right] d\ell \tag{4.12b}$$

This computation of reliability is commonly referred to as *load-resistance interference*.

TABLE 4.1 Reliability Formulas for Selected Distributions

Distribution of W	Probability density function $f_w(w)$	Mean μ_w	Coefficient of variation Ω_w	Reliability $p_s = P(W \geq 0)$
Normal	$\dfrac{1}{\sigma_w\sqrt{2\pi}}\exp\left[-\dfrac{1}{2}\left(\dfrac{w-\mu_w}{\sigma_w}\right)^2\right]$ for $-\infty < w < \infty$	μ_w	σ_w/μ_w	$\Phi\left(\dfrac{\mu_w}{\sigma_w}\right)$
Lognormal	$\dfrac{1}{\sqrt{2\pi}\,w\sigma_{\ln w}}\exp\left[-\dfrac{1}{2}\left(\dfrac{\ln(w)-\mu_{\ln w}}{\sigma_{\ln w}}\right)^2\right]$ for $w > 0$	$\mu_w = \exp\left[\mu_{\ln w}+\dfrac{\sigma_{\ln w}^2}{2}\right]$	$\sqrt{e^{\sigma_{\ln w}^2}-1}$	$\Phi\left(\dfrac{\mu_{\ln w}}{\sigma_{\ln w}}\right)$
Exponential	$\beta e^{-\beta(w-w_o)}$ for $w \geq w_o$	$\dfrac{1}{\beta}+w_o$	$\dfrac{1}{1+\beta\,w_o}$	$e^{-\beta(w-w_o)}$
Gamma	$\dfrac{\beta^\alpha(w-\xi)^{\alpha-1}e^{-\beta(w-\xi)}}{\Gamma(\alpha)}$ for $w \geq \xi$	$\dfrac{\alpha}{\beta}+\xi$	$\dfrac{\sqrt{\alpha}}{\alpha+\beta\xi}$	$1 - IG[\alpha,\beta(w-\xi)]/\Gamma(\alpha)*$
Beta	$\dfrac{1}{B(\alpha,\beta)}\dfrac{(w-a)^{\alpha-1}(b-w)^{\beta-1}}{(b-a)^{\alpha+\beta+1}}$ for $a \leq w \leq b$	$a+\dfrac{\alpha}{\alpha+\beta}(b-a)$	$\sqrt{\dfrac{\alpha\beta}{\alpha+\beta+1}}\dfrac{(b-a)}{(\alpha+\beta)\mu_w}$	$1-\dfrac{B_u\,(\alpha,\beta)^\dagger}{B(\alpha,\beta)}$ $u=\dfrac{w-a}{b-a}$
Triangular	$\dfrac{2}{b-a}\left(\dfrac{w-a}{m-a}\right)$ for $a \leq w \leq m$ $\dfrac{2}{b-a}\left(\dfrac{b-w}{b-m}\right)$ for $m \leq w \leq b$	$\dfrac{a+m+b}{3}$	$\left(\dfrac{1}{2}-\dfrac{ab+am+bm}{6\mu_w^2}\right)^{1/2}$	$=1-\dfrac{(w-a)^2}{(b-a)(m-a)}$ for $a \leq w \leq m$ $=\dfrac{(b-w)^2}{(b-a)(b-m)}$ for $m \leq w \leq b$
Uniform	$\dfrac{1}{b-a}$ for $a \leq w \leq b$	$\dfrac{a+b}{2}$	$\dfrac{1}{\sqrt{3}}\dfrac{b-a}{b+a}$	$\dfrac{b-w}{b-a}$

$*IG(\cdot) = $ incomplete gamma function.
$\dagger B_u(\cdot) = $ incomplete beta function.
SOURCE: After Yen et al. (1986).

Example 4.2 Consider the following joint PDF for the load and resistance:

$$f_{R,L}(r, \ell) = (r + \ell + r\ell)e^{-(r+\ell+r\ell)} \qquad \text{for } r > 0, \ell > 0$$

Compute the reliability p_s.

Solution According to Eq. (4.11), the reliability can be computed as

$$
\begin{aligned}
p_s &= \int_0^\infty \left[\int_0^r (r + \ell + r\ell)e^{-(r+\ell+r\ell)} d\ell \right] dr \\
&= \int_0^\infty \left[-(1 + \ell)e^{-(r+\ell+r\ell)} \right]_0^r dr \\
&= \int_0^\infty \left[e^{-r} - (1+r)e^{-(2r+r^2)} \right] dr = \left[\frac{1}{2}e^{-(2r+r^2)} - e^{-r} \right]_0^\infty = 0.5
\end{aligned}
$$

When the load and resistance are statistically independent, Eq. (4.11) can be reduced to

$$p_s = \int_{r_1}^{r_2} F_L(r) f_R(r)\, dr = E_R[F_L(R)] \tag{4.13a}$$

or
$$p_s = \int_{\ell_1}^{\ell_2} [1 - F_R(\ell)] f_L(\ell)\, d\ell = 1 - E_L[F_R(L)] \tag{4.13b}$$

in which $F_L(\cdot)$ and $F_R(\cdot)$ are the marginal CDFs of random load L and resistance R, respectively, $E_R[F_L(R)]$ is the expected value of the CDF of random load over the possible range of the resistance, and $E_L[F_R(L)]$ is the expected value of the CDF of random resistance over the possible range of the load. Similarly, the failure probability, when the load and resistance are independent, can be expressed as

$$p_f = 1 - p_s = E_R[1 - F_L(R)] = E_L[F_R(L)] \tag{4.14}$$

A schematic diagram illustrating load-resistance interference in the reliability computation, when the load and resistance are independent random variables, is shown in Fig. 4.2.

Example 4.3 Consider that the load and resistance are uncorrelated random variables, each of which has the following PDF:
Load (exponential distribution):

$$f_L(\ell) = 2e^{-2\ell} \qquad \text{for } \ell > 0$$

Resistance (Erlang distribution):

$$f_R(r) = 4r e^{-2r} \qquad \text{for } r > 0$$

Compute the reliability p_s.
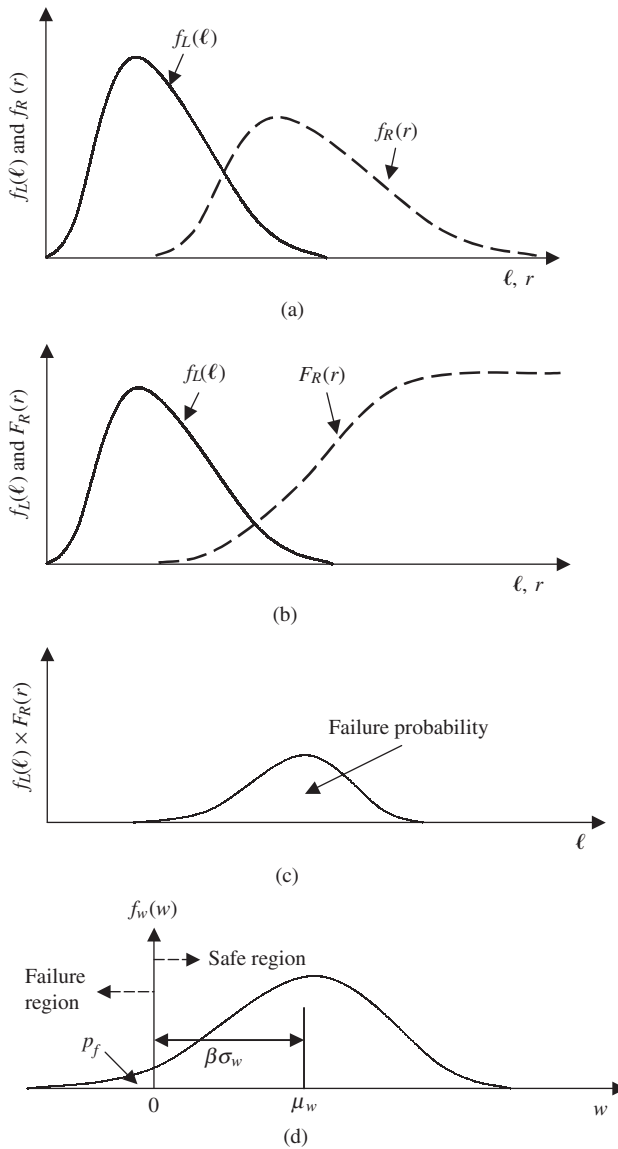
Figure 4.2 Schematic diagram of load-resistance interference for computing failure probability: (*a*) marginal densities of load and resistance; (*b*) PDF of load and CDF of resistance; (*c*) compute $f_L(\ell) \times F_R(r)$ over valid range of load; the area underneath the curve is the failure probability; (*d*) PDF of the performance function; the area left of $w = 0$ is the failure probability.

Solution Since the load and resistance are uncorrelated random variables, the reliability p_s can be computed according to Eq. (4.13a) as

$$p_s = \int_0^\infty (4re^{-2r}) \left[\int_0^r (2e^{-2\ell})\, d\ell \right] dr$$

$$= \int_0^\infty (4re^{-2r})(1 - e^{-2r})\, dr$$

$$= \left[\left(\frac{1}{4} + r \right) e^{-4r} - (1 + 2r)e^{-2r} \right]_0^\infty$$

$$= 0.75$$

In the case that the PDF of the performance function W is known or derived, the reliability can be computed according to Eq. (4.4) as

$$p_s = \int_0^\infty f_w(w)\, dw \tag{4.15}$$

in which $f_w(w)$ is the PDF of the performance function.

Example 4.4 Define the performance function $W = R - L$, in which R and L are independent random variables with their PDFs given in Example 4.2. Determine the reliability p_s using Eq. (4.15).

Solution To use Eq. (4.15) for the reliability computation, it is necessary to first obtain the PDF of the performance function W. Derivation of the PDF of W can be made based on the derived distribution method described in Tung and Yen (2005, Sec. 3.1) as follows: Define $W = R - L$ and $U = L$ from which the original random variables R and L can be expressed in terms of new random variables W and U as $L = U$ and $R = W + U$. By the transformation of variables, the joint PDF of W and U can be expressed as

$$f_{w,u}(w, u) = f_{R,L}(r, \ell)|\boldsymbol{J}|$$

in which the Jacobian matrix $\boldsymbol{J}$ is

$$\boldsymbol{J} = \begin{bmatrix} \dfrac{\partial L}{\partial W} & \dfrac{\partial L}{\partial U} \\[2mm] \dfrac{\partial R}{\partial W} & \dfrac{\partial R}{\partial U} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

The absolute value of the determinant of the Jacobian matrix $|\boldsymbol{J}|$ is equal to one. Hence the joint PDF of W and U is

$$f_{w,u}(w, u) = f_R(r)f_L(\ell)|\boldsymbol{J}| = f_R(w + u)f_L(u)(1) = 8(w + u)e^{-2(w+2u)}$$

for $-\infty < w < \infty$ and $u \geq 0$. Because the marginal PDF associated with the performance function W is needed, it can be obtained from the preceding joint PDF as

$$f_w(w) = \int_0^\infty f_{w,u}(w, u)\, du = \frac{1}{2}\frac{1 + 4w}{e^{2w}} \qquad \text{for } w \geq 0$$

From the derived PDF for W, the reliability can be computed as

$$p_s = \int_0^\infty \frac{1}{2}\left(\frac{1+4w}{e^{2w}}\right)dw = \left[-\left(w+\frac{3}{4}\right)e^{-2w}\right]_0^\infty = 0.75$$

In the conventional reliability analysis of hydraulic engineering design, uncertainty from the hydraulic aspect often is ignored. Treating the resistance or capacity of the hydraulic structure as a constant reduces Eq. (4.11) to

$$p_s = \int_0^{r_o} f_L(\ell)\,d\ell \tag{4.16}$$

in which r_o is the resistance of the hydraulic structure, a deterministic quantity. If the PDF of the hydrologic load is the annual event, such as the annual maximum flood, the resulting annual reliability can be used to calculate the corresponding return period.

To express the reliability in terms of stochastic variables in load and resistance functions, Eq. (4.11) can be written as

$$p_s = \int_{h(x_R)=r_1}^{h(x_R)=r_2}\left[\int_0^{g(x_L)=r} f(\boldsymbol{x}_R, \boldsymbol{x}_L)\,d\boldsymbol{x}_L\right]d\boldsymbol{x}_R \tag{4.17}$$

in which $f(\boldsymbol{x}_L, \boldsymbol{x}_R)$ is the joint PDF of model stochastic basic variables $\boldsymbol{X}$. For independent stochastic basic variables X, Eq. (4.17) can be written as

$$p_s = \int_{h(\boldsymbol{x}_R)=r_1}^{h(\boldsymbol{x}_R)=r_2}\left[\int_0^{g(\boldsymbol{x}_L)=r} \prod_{j=1}^m f_j(x_j)\,d\boldsymbol{x}_L\right]\prod_{k=m+1}^K f_k(x_k)\,d\boldsymbol{x}_R \tag{4.18}$$

in which $f_k(\cdot)$ is the marginal PDF of the stochastic basic variable X_k.

The method of direct integration requires the PDFs of the load and resistance or the performance function to be known or derived. This is seldom the case in practice, especially for the joint PDF, because of the complexity of hydrologic and hydraulic models used in design. Explicit solution of direct integration can be obtained for only a few PDFs, as given in Table 4.1 for the reliability p_s. For most other PDFs, numerical integration may be necessary. Computationally, the direct integration method is analytically tractable for only very few special combinations of probability distributions and functional relationships. For example, the distribution of the safety margin W expressed by Eq. (4.5) has a normal distribution if both load and resistance functions are linear and all stochastic variables are normally distributed. In terms of the safety factor expressed as Eqs. (4.6) and (4.7), the distribution of $W(\boldsymbol{X})$ is lognormal if both load and resistance functions have multiplicative forms involving lognormal stochastic variables. Most of the time, numerical integrations are performed for reliability determination. When using numerical integration (including Monte Carlo simulation described in Chap. 6), difficulty may be encountered when one deals with a multivariate problem. Appendix 4A summarizes a few one-dimensional numerical integration schemes.

Example 4.5   Referring to Example 4.1, the stochastic basic variables n, D, and S in Manning's formula to compute the sewer capacity are independent lognormal random variables with the following statistical properties:

Parameter	Mean	Coefficient of variation
n (ft$^{1/6}$)	0.015	0.05
D (ft)	3.0	0.02
S (ft/ft)	0.005	0.05

Compute the reliability that the sewer can convey the inflow discharge of 35 ft^3/s.

Solution   In this example, the resistance function is $R(n, D, S) = 0.463\, n^{-1} D^{2.67} S^{0.5}$, and the load is $L = 35$ ft^3/s. Since all three stochastic parameters are lognormal random variables, the performance function appropriate for use is

$$
\begin{aligned}
W(n, D, S) &= \ln(R) - \ln(L) \\
&= [\ln(0.463) - \ln(n) + 2.67\ln(D) + 0.5\ln(S)] - \ln(35) \\
&= -\ln(n) + 2.67\ln(D) + 0.5\ln(S) - 4.3319
\end{aligned}
$$

The reliability $p_s = P[W(n, D, S) > 0]$ then can be computed as follows:

Since n, D, and S are independent lognormal random variables, $\ln(n)$, $\ln(D)$, and $\ln(S)$ are independent normal random variables. Note that the performance function $W(n, D, S)$ is a linear function of normal random variables. Then, by the reproductive property of normal random variables as described in Sec. 2.6.1, $W(n, D, S)$ also is a normal random variable with the mean

$$
\mu_w = -\mu_{\ln(n)} + 2.67\mu_{\ln(D)} + 0.5\mu_{\ln(S)} - 4.3319
$$

and the variance

$$
\text{Var}(W) = \text{Var}[\ln(n)] + 2.67^2\text{Var}[\ln(D)] + 0.5^2\text{Var}[\ln(S)]
$$

From Eq. (2.67), the means and variances of log-transformed variables can be obtained as

$$
\text{Var}[\ln(n)] = \ln(1 + 0.05^2) = 0.0025 \qquad \mu_{\ln(n)} = \ln(\mu_n) - 0.5\,\text{Var}[\ln(n)] = -4.201
$$

$$
\text{Var}[\ln(D)] = \ln(1 + 0.02^2) = 0.0004 \qquad \mu_{\ln(D)} = \ln(\mu_D) - 0.5\,\text{Var}[\ln(D)] = 1.0984
$$

$$
\text{Var}[\ln(S)] = \ln(1 + 0.05^2) = 0.0025 \qquad \mu_{\ln(S)} = \ln(\mu_S) - 0.5\,\text{Var}[\ln(S)] = -5.2996
$$

Then the mean and variance of the performance function $W(n, D, S)$ can be computed as

$$
\mu_w = 0.1517 \qquad \text{Var}(W) = 0.005977
$$

The reliability can be obtained as

$$
p_s = P(W > 0) = \Phi\left(\frac{\mu_w}{\sigma_w}\right) = \Phi\left(\frac{0.1517}{\sqrt{0.005977}}\right) = \Phi(1.958) = 0.975
$$

4.4 Mean-Value First-Order Second-Moment (MFOSM) Method

In the first-order methods, the performance function $W(\boldsymbol{X})$, defined on the basis of the loading and resistance functions $g(\boldsymbol{X}_L)$ and $h(\boldsymbol{X}_R)$, are expanded in a Taylor series at a reference point. The second- and higher-order terms in the series expansion are truncated, resulting in an approximation involving only the first two statistical moments of the variables. This simplification greatly enhances the practicality of the first-order methods because, in many problems, it is rather difficult, if not impossible, to find the PDF of the variables, whereas it is relatively simple to estimate the first two statistical moments. The procedure is based on the first-order variance estimation (FOVE) method, which is summarized below. For a detailed description of the method in uncertainty analysis, readers are referred to Tung and Yen (2005, Sec. 5.1).

The *first-order variance estimation* (FOVE) method, also called the *variance propagation method* (Berthouex, 1975), estimates uncertainty features of a model output based on the statistical properties of the model's stochastic basic variables. The basic idea of the method is to approximate a model involving stochastic basic variables by a Taylor series expansion. Consider that a hydraulic or hydrologic performance function $W(\boldsymbol{X})$ is related to K stochastic basic variables because $W(\boldsymbol{X}) = W(X_1, X_2, \ldots, X_K)$, in which $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$, a K-dimensional column vector of variables in which all Xs are subject to uncertainty, the superscript t represents the transpose of a matrix or vector. The Taylor series expansion of the performance function $W(\boldsymbol{X})$ with respect to a selected point of stochastic basic variables $\boldsymbol{X} = \boldsymbol{x}_o$ in the parameter space can be expressed as

$$W(\boldsymbol{X}) = w_o + \sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k} \right]_{\boldsymbol{x}_o} (X_k - x_{ko})$$

$$+ \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \left[\frac{\partial^2 W(\boldsymbol{X})}{\partial X_i \partial X_j} \right]_{\boldsymbol{x}_o} (X_i - x_{io})(X_j - x_{jo}) + \varepsilon \qquad (4.19)$$

in which $w_o = W(\boldsymbol{x}_o)$, and ε represents the higher-order terms. The partial derivative terms are called *sensitivity coefficients*, each representing the rate of change in the performance function value W with respect to the unit change of the corresponding variable at $\boldsymbol{x}_o$.

Dropping the higher-order terms represented by ε, Eq. (4.19) is a second-order approximation of the model $W(\boldsymbol{X})$. Further truncating the second-order terms from it leads to the first-order approximation of W as

$$W(\boldsymbol{X}) \approx w_o + \sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k} \right]_{x_o} (X_k - x_{ko}) \qquad (4.20)$$

or in a matrix form as

$$W(\boldsymbol{X}) = w_o + \boldsymbol{s}_o^t(\boldsymbol{X} - \boldsymbol{x}_o) \tag{4.21}$$

where $\boldsymbol{s}_o = \nabla_x W(\boldsymbol{x}_o)$ is the column vector of sensitivity coefficients with each element representing $\partial W/\partial X_k$ evaluated at $\boldsymbol{X} = \boldsymbol{x}_o$. The mean and variance of W by the first-order approximation can be expressed, respectively, as

$$\mu_w = E[W(\boldsymbol{X})] \approx w_o + \sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k}\right]_{x_o} (\mu_k - x_{ko}) \tag{4.22}$$

and
$$\text{Var}[W(\boldsymbol{X})] \approx \text{Var}\left\{ w_o + \sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k}\right]_{x_o} (X_k - x_{ko}) \right\}$$

$$= \sum_{j=1}^{K} \sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_j}\right]_{x_o} \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k}\right]_{x_o} \text{Cov}(X_j, X_k) \tag{4.23}$$

In matrix forms, Eqs. (4.22) and (4.23) can be expressed as

$$\mu_w \approx w_o + \boldsymbol{s}_o^t(\boldsymbol{\mu}_x - \boldsymbol{x}_o) \tag{4.24}$$
and
$$\sigma_w^2 \approx \boldsymbol{s}_o^t \boldsymbol{C}_x \boldsymbol{s}_o \tag{4.25}$$

in which $\boldsymbol{\mu}_x$ and $\boldsymbol{C}_x$ are the vectors of the means and covariance matrix of the stochastic basic variables $\boldsymbol{X}$, respectively.

Commonly, the first-order variance estimation method consists of taking the expansion point $\boldsymbol{x}_o = \boldsymbol{\mu}_x$ at which the mean and variance of W reduce to

$$\mu_w \approx g(\boldsymbol{\mu}_x) = \overline{w} \tag{4.26}$$
and
$$\sigma_w^2 \approx \boldsymbol{s}^t \boldsymbol{C}_x \boldsymbol{s} \tag{4.27}$$

in which $\boldsymbol{s} = \nabla_x \boldsymbol{W}(\boldsymbol{\mu}_x)$ is a K-dimensional vector of sensitivity coefficients evaluated at $\boldsymbol{x}_o = \boldsymbol{\mu}_x$. When all stochastic basic variables are independent, the variance of model output W could be approximated as

$$\sigma_w^2 \approx \sum_{k=1}^{K} s_k^2 \sigma_k^2 = \boldsymbol{s}^t \boldsymbol{D}_x \boldsymbol{s} \tag{4.28}$$

in which $\boldsymbol{D}_x = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$ is a $K \times K$ diagonal matrix of variances of the involved stochastic basic variables. From Eq. (4.28), the ratio $s_k^2 \sigma_k^2/\text{Var}(W)$ indicates the proportion of the overall uncertainty in the model output contributed by the uncertainty associated with the stochastic basic variable X_k.

The MFOSM method for reliability analysis first applies the FOVE method to estimate the statistical moments of the performance function $W(\boldsymbol{X})$. This is done by applying the expectation and variance operators to the first-order Taylor series approximation of the performance function $W(\boldsymbol{X})$ expanded at the mean values of the stochastic basic variables. Once the mean and standard deviations of $W(\boldsymbol{X})$ are estimated, the reliability is computed according to Eqs. (4.9) or (4.10), with the reliability index β_{MFOSM} computed as

$$\beta_{\text{MFOSM}} = \frac{W(\boldsymbol{\mu}_x)}{\sqrt{\boldsymbol{s}^t \boldsymbol{C}_x \boldsymbol{s}}} \tag{4.29}$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{C}_x$ are the vectors of means and covariance matrix of stochastic basic variables $\boldsymbol{X}$, respectively, and $\boldsymbol{s} = \nabla_x W(\boldsymbol{\mu}_x)$ is the column vector of sensitivity coefficients with each element representing $\partial W/\partial X_k$ evaluated at $\boldsymbol{X} = \boldsymbol{\mu}_x$.

Example 4.6   Manning's formula for determining flow capacity of a storm sewer is

$$Q = 0.463 n^{-1} D^{2.67} S^{0.5}$$

in which Q is flow rate (in ft^3/s), n is the Manning roughness coefficient, D is the sewer diameter (in ft), and S is pipe slope (in ft/ft). Because roughness coefficient n, sewer diameter D, and sewer slope S in Manning's formula are subject to uncertainty owing to manufacturing imprecision and construction error, the sewer flow capacity would be subject to uncertainty. Consider a section of circular sewer pipe with the following features:

Model parameter	Nominal value	Coefficient of variation
Roughness coefficient (n)	0.015	0.05
Pipe diameter (D, ft.)	3.0	0.05
Pipe slope (S, ft/ft)	0.005	0.05

Compute the reliability that the sewer capacity could convey a discharge of 35 ft^3/s. Assume that stochastic model parameters n, D, and S are uncorrelated.

Solution   The performance function for the problem is $W = Q - 35 = 0.463 n^{-1} D^{2.67} S^{0.5} - 35$. The first-order Taylor series expansion of the performance function about $n_o = \mu_n = 0.015$, $D_o = \mu_D = 3.0$, and $S_o = \mu_S = 0.005$, according to Eq. (4.20), is

$$W \approx 0.463(0.015)^{-1}(3)^{2.67}(0.005)^{0.5} + (\partial Q/\partial n)(n - 0.015) + (\partial Q/\partial D)(D - 3.0)$$

$$+ (\partial Q/\partial S)(S - 0.0005) - 35$$

$$= 41.01 - 2733.99(n - 0.015) + 36.50(D - 3.0) + 4100.99(S - 0.005) - 35$$

Based on Eq. (4.26), the approximated mean of the sewer flow capacity is

$$\mu_w \approx 41.01 - 35 = 6.01 \text{ ft}^3/\text{s}$$

Owing to independency of n, D, and S, according to Eq. (4.28), the approximated variance of the performance function W is

$$\sigma_Q^2 \approx (2733.99)^2 \text{Var}(n) + (36.50)^2 \text{Var}(D) + (4100.99)^2 \text{Var}(S)$$

Since

$$\text{Var}(n) = (\Omega_n \mu_n)^2 = (0.05 \times 0.015)^2 = (7.5 \times 10^{-4})^2$$

$$\text{Var}(D) = (\Omega_D \mu_D)^2 = (0.05 \times 3.0)^2 = (1.5 \times 10^{-1})^2$$

$$\text{Var}(S) = (\Omega_S \mu_S)^2 = (0.05 \times 0.005)^2 = 0.00025^2 = 6.25 \times 10^{-8}$$

the variance of the performance function W can be computed as

$$\sigma_Q^2 \approx (2733.99)^2 (7.5 \times 10^{-4})^2 + (36.50)^2 (1.5 \times 10^{-1})^2 + (4100.99)^2 (2.5 \times 10^{-4})^2$$

$$= 2.05^2 + 5.47^2 + 1.03^2 = 35.23 (\text{ft}^3/\text{s})^2$$

Hence the standard deviation of the sewer flow capacity is $\sqrt{35.23} = 5.94$ ft^3/s.

The MFOSM reliability index is $\beta_{\text{MFOSM}} = 6.01/5.94 = 1.01$. Assuming a normal distribution for Q, the reliability that the sewer capacity can accommodate a discharge of 35 ft^3/s is

$$p_s = P[Q > 35] = \Phi(\beta_{\text{MFOSM}}) = \Phi(1.01) = 0.844$$

The corresponding failure probability is $p_f = \Phi(-1.01) = 0.156$.

Yen and Ang (1971), Ang (1973), and Cheng et al. (1986b) indicated that provided that $p_s < 0.99$, reliability is not greatly influenced by the choice of distribution for W, and the assumption of a normal distribution is satisfactory. However, for reliability higher than this value (for example, $p_s = 0.999$), the shape of the tail of a distribution becomes very critical. In such cases, accurate assessment of the distribution of $W(\boldsymbol{X})$ should be used to evaluate the reliability or failure probability. The MFOSM method has been used widely in various hydrosystems infrastructural designs and analyses such as storm sewers (Tang and Yen, 1972; Tang et al., 1975; Yen and Tang, 1976; Yen et al., 1976), culverts (Yen et al., 1980; Tung and Mays, 1980), levees (Tung and Mays, 1981; Lee and Mays, 1986), floodplains (McBean et al., 1984), and open-channel hydraulics (Huang, 1986).

Example 4.7 Referring to Example 4.6, using the same values of the mean and standard deviation for sewer flow capacity, the following table lists the reliabilities and failure probabilities determined by different distributional assumptions for the sewer flow capacity Q to accommodate the inflow discharge of 35 ft^3/s.

Distribution	p_s	p_f
Normal	0.996955	0.003045
Lognormal	0.997704	0.002296
Gumbel	0.999819	0.000191

As can be seen, using different distributional assumptions might result in significant differences in the estimation of failure probability. This results mainly from the fact that the MFOSM method solely uses the first two moments without taking into account the distributional properties of the random variables.

Assuming that stochastic parameters in the sewer capacity formula (that is, n, D, and S) are uncorrelated lognormal random variables, the sewer capacity also is a lognormal random variable. The following table lists the values of the exact reliability index and failure probability and those obtained from the MFOSM by Eq. (4.29) and (4.8). The table indicates that approximation by the MFOSM becomes less and less accurate as the computation approaches the tail portion of the distribution.

Inflow rate (ft^3/s)	MFOSM		Exact	
	β_1	$p_f = \Phi(-\beta_1)$	β_2	$p_f = \Phi(-\beta_2)$
25	5.035	2.384×10^{-7}	6.350	$\gg 0$
30	3.457	2.728×10^{-4}	3.991	3.290×10^{-5}
35	1.880	3.045×10^{-3}	1.996	2.296×10^{-3}
40	0.303	3.810×10^{-1}	0.268	3.943×10^{-1}
45	−1.274	8.988×10^{-1}	−1.256	8.954×10^{-1}

NOTE: $\beta_1 = \mu_w/\sigma_w$, $\beta_2 = \mu_{\ln w}/\sigma_{\ln w}$, and $W = Q -$ inflow.

Application of the MFOSM method is simple and straightforward. However, it possesses certain weaknesses in addition to the difficulties with accurate estimation of extreme failure probabilities mentioned earlier. These weaknesses include

1. *Inappropriateness of the expansion point.* In reliability computation, the concern often is those points in the parameter space that fall on the failure surface or limiting-state surface. In the MFOSM method, the expansion point is located at the mean of the stochastic basic variables that do not necessarily define the critical state of the system. The difference in expansion points and the resulting reliability indices between the MFOSM and its alternative, called the *advanced first-order, second-moment method* (AFOSM), is shown in Fig. 4.3.

2. *Inability to handle distributions with a large skewness coefficient.* Table 4.2 indicates that the discrepancy of the failure probability estimated by the MFOSM method for a lognormally distributed performance function becomes larger as the degree of skewness increases. This mainly due to the fact that the MFOSM method incorporates only the first two moments of the random parameters involved. In other words, the MFOSM method simply ignores any moments higher than the second order. Therefore, for those random variables having asymmetric PDFs, the MFOSM method cannot capture such a feature in the reliability computation.

3. *Generally poor estimations of the mean and variance of nonlinear functions.* This is evident in that the MFOSM method is the first-order representation
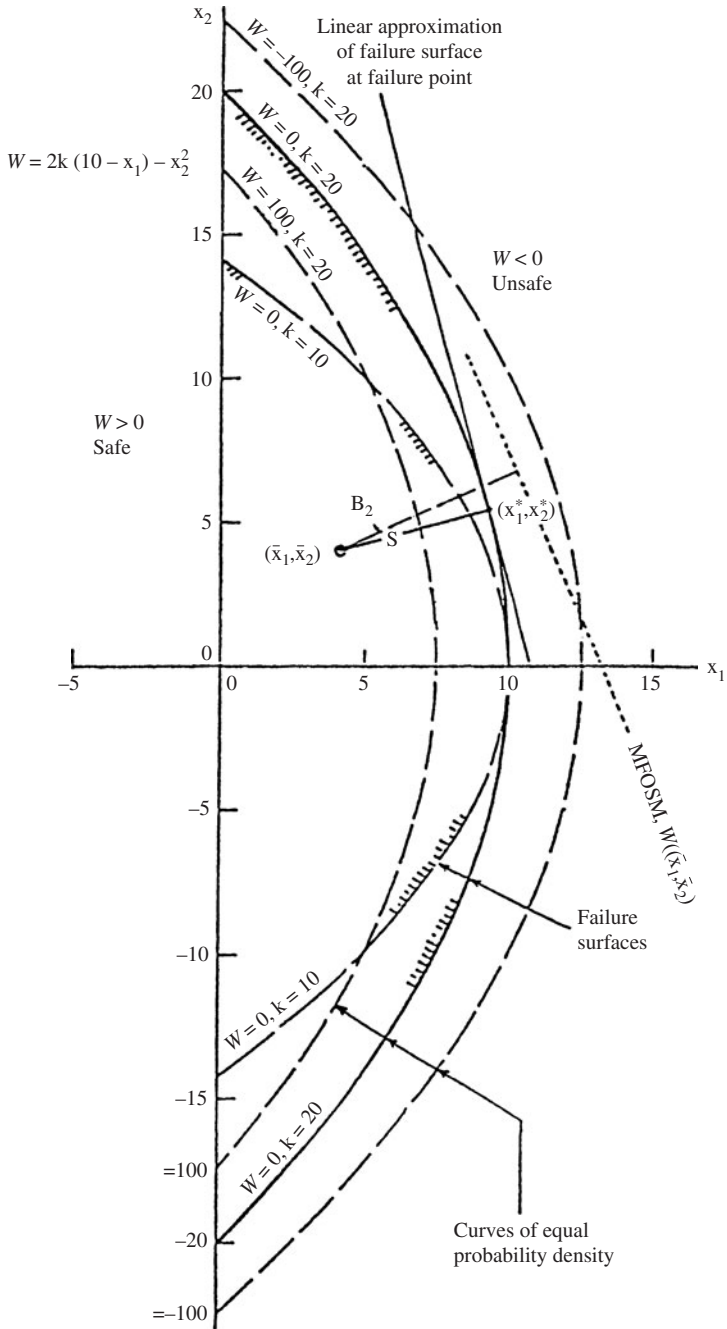
Figure 4.3 Differences in expansion points and reliability indices between the MFOSM and AFOSM methods.

TABLE 4.2   Effect of Skewness on the Accuracy of p_f Estimated by the MFOSM Method

				Exact		MFOSM	
μ_w	σ_w	Ω_w	γ_w	β_{Exact}	p_f	β_{MFOSM}	p_f
1.0	0.3	0.3	0.927	7.70	7.036×10^{-15}	3.00	1.350×10^{-3}
1.0	0.5	0.5	1.625	4.64	1.759×10^{-6}	1.80	3.593×10^{-2}
1.0	1.0	1.0	4.000	2.35	9.402×10^{-3}	0.90	1.841×10^{-1}
1.0	2.0	2.0	27.00	1.18	1.190×10^{-1}	0.45	3.260×10^{-1}

NOTE: $p_f = P(W < 0.1)$, with W being a lognormal random variable.

of the original performance function. In case the performance function is highly nonlinear, linear approximation of such a nonlinear function will not be accurate. Consequently, the estimations of the mean and variance of a nonlinear performance function will be less accurate. The accuracy associated with the estimated mean and variance deteriorates rapidly as the degree of nonlinearity of the performance function increases. For a linear performance function, the FOVE method would produce the exact values for the mean and variance.

4. *Sensitivity of the computed failure probability to the formulation of the performance function W.* Ideally, the computed reliability or failure probability for a system should not depend on the definition of the performance function. However, this is not the case for the MFOSM method. This phenomenon of lack of invariance to the type of performance function is shown in Figs. 4.4 and 4.5. The main reason for this inconsistency is because the MFOSM method would result in different first-order approximations for different forms of the performance function. Consequently, different values of mean and variance will be obtained, resulting in different estimations of reliability and failure probability for the same problem. This behavior of the MFOSM could create an unnecessary puzzle for engineers with regard to which performance function should be used to obtain an accurate estimation of reliability. This is not an easy question to answer, in general, except for a very few simple cases. Another observation that can be made from Figs. 4.4 and 4.5 is that the discrepancies among failure probabilities computed by the MFOSM method using different performance functions become more pronounced as the uncertainties of the stochastic basic variables get larger.

5. *Limited ability to use available probabilistic information.* The reliability index β gives only weak information on the probability of failure, and thus the appropriate system probability distribution must be assumed. Further, the MFOSM method provides no logical way to include available information on basic variable probability distributions.

From these arguments, the general rule of thumb is not to rely on the result of the MFOSM method if any of the following conditions exist: (1) high accuracy
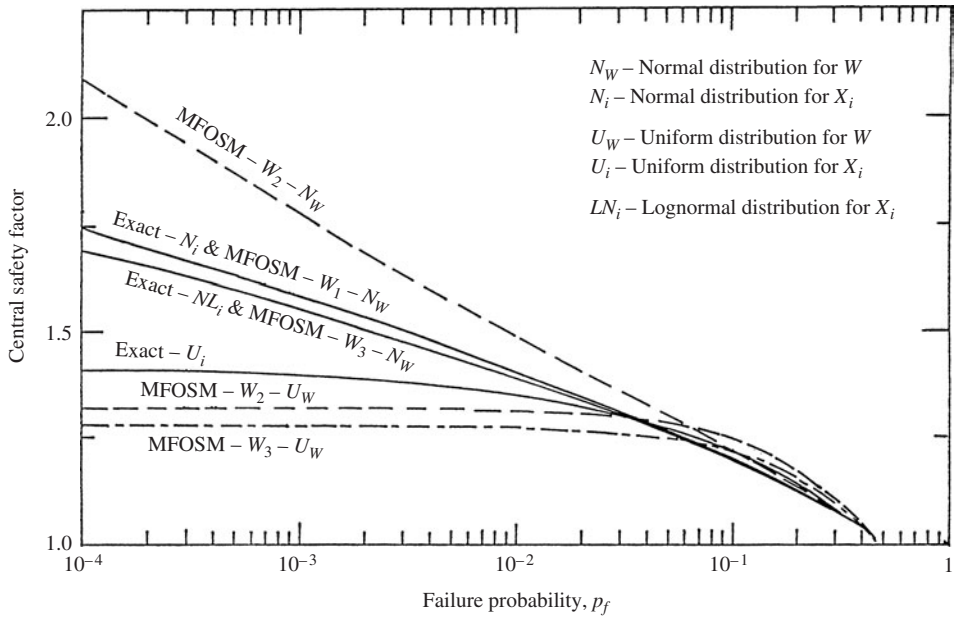
Figure 4.4 Comparison of risk-safety factor curves by different methods using various distributions with $\Omega_L = \Omega_R = 0.1$, where $W_1 = R - L$, $W_2 = (R/L) - 1$, and $W_3 = \ln(R/L)$, and R is the resistance of the system and L is the load placed on the system. (*After Yen et al., 1986.*)
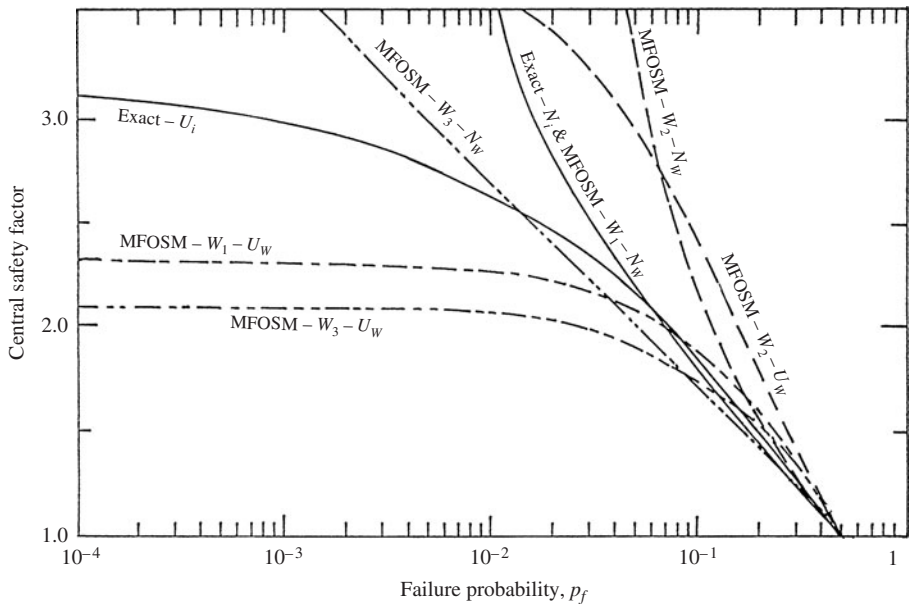
Figure 4.5 Comparison of risk-safety factor curves by different methods using various distributions with $\Omega_L = \Omega_R = 0.3$, where $W_1 = R - L$, $W_2 = (R/L) - 1$, and $W_3 = \ln(R/L)$, and R is the resistance of the system and L is the load placed on the system. (*After Yen et al., 1986.*)

requirement for the estimated reliability or failure probability, (2) high non-linearity of the performance function, and (3) many skewed random variables involved in the performance function. However, Cornell (1969) made a strong defense for the MFOSM method from a practical standpoint as follows:

> An approach based on means and variances may be all that is justified when one appreciates (1) that data and physical arguments are often insufficient to establish the full probability law of a variable; (2) that most engineering analyses include an important component of real, but difficult to measure, professional uncertainty; and (3) that the final output, namely, the decision or design parameters, is often not sensitive to moments higher than the mean and variance.

To reduce the effect of nonlinearity, one way is to include the second-order terms in the Taylor series expansion. This would increase the burden of analysis by having to compute the second-order partial derivatives. Another alternative within the realm of first-order simplicity is given in Sec. 4.5. Section 4.6 briefly describes the basis of the second-order reliability analysis techniques.

4.5 Advanced First-Order Second-Moment (AFOSM) Method

The main thrust of the AFOSM method is to improve the deficiencies associated with the MFOSM method, while keeping the simplicity of the first-order approximation. Referring to Fig. 4.3, the difference in the AFOSM method is that the expansion point $x_* = (x_{L*}, x_{R*})$ for the first-order Taylor series is located on the failure surface defined by the limit-state equation, $W(x) = 0$. In other words, the failure surface is the boundary that separates the system performance from being unsatisfactory (unsafe) or being satisfactory (safe), that is,

$$W(x) \begin{cases} > 0, \text{ system performance is satisfactory (or safe region);} \\ = 0, \text{ limit-state surface (or failure surface);} \\ < 0, \text{ system performance is unsatisfactory (or failure region).} \end{cases}$$

The AFOSM method has been applied to various hydrosystem engineering problems, including storm sewers (Melching and Yen, 1986), dams (Cheng et al., 1982; 1993), sea dikes and barriers (Vrijling, 1987; 1993), freeboard design (Cheng et al., 1986a), bridge scour (Yen and Melching, 1991; Chang, 1994), rainfall-runoff modeling (Melching et al., 1990; Melching, 1992); groundwater pollutant transport (Sitar et al., 1987; Jang et al., 1990), open channel design (Easa, 1992), sediment transport (Bechtler and Maurer, 1992), backwater computations (Cesare, 1991; Singh and Melching, 1993), and water quality modeling (Tung, 1990; Melching and Anmangandla, 1992; Melching and Yoon, 1996; Han et al., 2001; Tolson et al., 2001).

4.5.1 Definitions of stochastic parameter spaces

Before discussing the AFOSM methods, a few notations with regard to the stochastic basic variable space are defined first. In general, the original

stochastic basic variables $\boldsymbol{X}$ could be correlated, non-normal random variables having a vector of mean $\boldsymbol{\mu}_x = (\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_K})^t$ and covariance matrix $\boldsymbol{C}_x$ as shown in Sec. 2.7.2. The original random variables $\boldsymbol{X}$ can be standardized as

$$\boldsymbol{X}' = \boldsymbol{D}_x^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}_x) \tag{4.30}$$

in which $\boldsymbol{X}' = (X_1', X_2', \ldots, X_K')^t$ is a vector of correlated, standardized random variables, and $\boldsymbol{D}_x = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$ is an $K \times K$ diagonal variance matrix. Through the standardization procedure, each standardized variable $\boldsymbol{X}'$ has the mean zero and unit standard deviation. The covariance matrix of $\boldsymbol{X}'$ reduces to the correlation matrix of the original random variables $\boldsymbol{X}$, that is, $\boldsymbol{C}_{x'} = \boldsymbol{R}_x$, as shown in Sec. 2.7.2. Note that if the original random variables $\boldsymbol{X}$ are nonnormal, the standardized ones $\boldsymbol{X}'$ are nonnormal as well. Because it is generally easier to work with the uncorrelated variables in the reliability analysis, the correlated random variables $\boldsymbol{X}$ are often transformed into uncorrelated ones $\boldsymbol{U} = T(\boldsymbol{X})$, with $T(\cdot)$ representing transformation, in general. More specifically, orthogonal transforms often are used to obtain uncorrelated random variables from the correlated ones. Two frequently used orthogonal transforms, namely, *Cholesky decomposition* and *spectral decomposition*, for dealing with correlated random variables are described in Appendix 4B. In probability evaluation, it is generally convenient to deal with normal random variables. For this reason, orthogonal transformation, normal transformation, and standardization procedures are applied to the original random variables $\boldsymbol{X}$ to obtain independent, standardized normal random variables $\boldsymbol{Z}'$. Hence this chapter adopts $\boldsymbol{X}$ for stochastic basic variables in the original scale, $\boldsymbol{X}'$ for the standardized correlated stochastic basic variables, $\boldsymbol{U}$ for the uncorrelated variables, and $\boldsymbol{Z}$ and $\boldsymbol{Z}'$, respectively, for correlated and independent, standardized normal stochastic basic variables.

4.5.2 Determination of design point (Most probable failure point)

In cases for which several stochastic basic variables are involved in a performance function, the number of possible combinations of such variables satisfying $W(\boldsymbol{x}) = 0$ is infinite. From the design viewpoint, one is more concerned with the combination of stochastic basic variables that would yield the lowest reliability or highest failure probability. The point on the failure surface associated with the lowest reliability is the one having the shortest distance to the point where the means of the stochastic basic variables are located. This point is called the *design point* (Hasofer and Lind, 1974) or the *most probable failure point* (Shinozuka, 1983).

Consider that $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$ are K uncorrelated stochastic basic variables having a vector mean $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{D}_x$. The original stochastic basic variables $\boldsymbol{X}$ can be standardized into $\boldsymbol{X}'$ according to Eq. (4.30). The standardization procedure maps the failure surface in the original $\boldsymbol{x}$-space onto the corresponding failure surface in $\boldsymbol{x}'$-space, as shown in Fig. 4.6. Hence the design point in $\boldsymbol{x}'$-space is the one that has the shortest distance from the
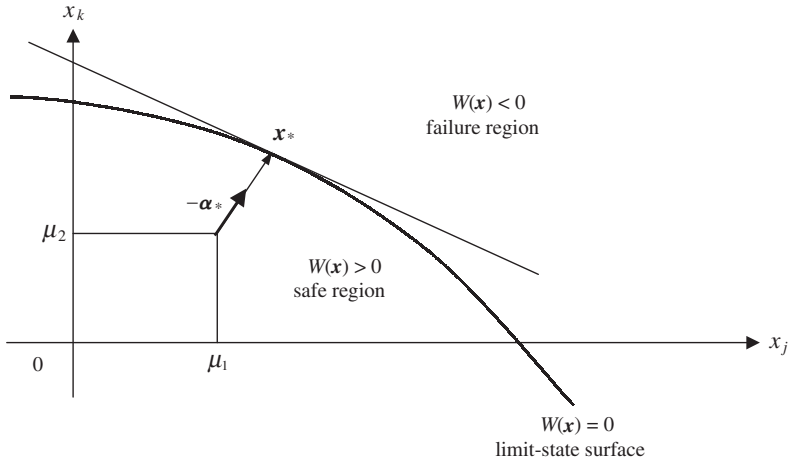
(a)
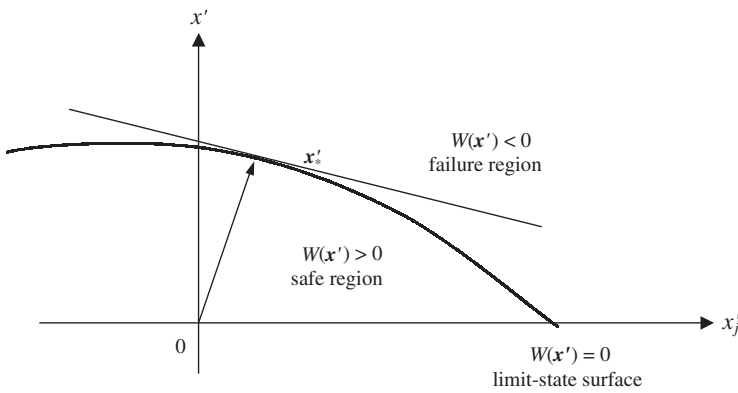
(b)

Figure 4.6 Performance function in the original and standardized spaces:
(*a*) original space; (*b*) standardized space.

failure surface $W'(\boldsymbol{x}') = 0$ to the origin $\boldsymbol{x}' = \boldsymbol{0}$. Such a point can be found by solving

$$\text{Minimize} \qquad |\boldsymbol{x}'| = (\boldsymbol{x}'^{t}\boldsymbol{x}')^{1/2} = \sqrt{\sum_{k=1}^{K} x_k'^2} \qquad (4.31a)$$

$$\text{subject to} \qquad W'(\boldsymbol{x}') = 0 \qquad (4.31b)$$

This constrained nonlinear minimization problem can be converted into an unconstrained minimization problem using the *Lagrangian function*:

$$\text{Minimize } L(\boldsymbol{x}', \xi) = (\boldsymbol{x}'^{t}\boldsymbol{x}')^{1/2} + \xi W'(\boldsymbol{x}') \qquad (4.32)$$

in which ξ is the *Lagrangian multiplier*, which is unrestricted in sign. The solution to Eq. (4.32) can be obtained by solving the following two equations simultaneously, that is,

$$\nabla_{x'} L(\boldsymbol{x}'_*, \xi_*) = \left[\frac{\partial L(\boldsymbol{x}', \xi)}{\partial \boldsymbol{x}'}\right]_{(\boldsymbol{x}'_*, \xi_*)} = \frac{\boldsymbol{x}'_*}{|\boldsymbol{x}'_*|} + \xi_* \nabla_{x'} W'(\boldsymbol{x}'_*) = \boldsymbol{0} \qquad (4.33a)$$

$$\nabla_{\xi} L(\boldsymbol{x}'_*, \xi_*) = \left[\frac{\partial L(\boldsymbol{x}', \xi)}{\partial \xi}\right]_{(\boldsymbol{x}'_*, \xi_*)} = W'(\boldsymbol{x}'_*) = 0 \qquad (4.33b)$$

in which $\nabla_x = (\partial/\partial x_1, \partial/\partial x_2, \dots, \partial/\partial x_K)^t$ is a gradient operator. From Eq. (4.33a), the design point $\boldsymbol{x}'_*$ can be expressed as

$$\boldsymbol{x}'_* = -\xi_* |\boldsymbol{x}'_*| \nabla_{x'} W'(\boldsymbol{x}'_*) \qquad (4.34)$$

Furthermore, from Eq. (4.34), the distance between the origin $\boldsymbol{x}' = \boldsymbol{0}$ and the design point $\boldsymbol{x}'_*$ can be obtained as

$$|\boldsymbol{x}'_*| = |\xi_*| |\boldsymbol{x}'_*| [\nabla^t_{x'} W'(\boldsymbol{x}'_*) \nabla_{x'} W'(\boldsymbol{x}'_*)]^{1/2} = |\xi_*| |\boldsymbol{x}'_*| |\nabla_{x'} W'(\boldsymbol{x}'_*)| \qquad (4.35)$$

from which the value of the optimal Lagrangian multiplier $\boldsymbol{x}_*$ can be determined as

$$\xi_* = \text{sign}[W'(\boldsymbol{0})] |\nabla_{x'} W'(\boldsymbol{x}'_*)|^{-1} \qquad (4.36)$$

Substituting Eq. (4.36) into Eq. (4.34) determines the location of the design point as

$$\boldsymbol{x}'_* = -\text{sign}[W'(\boldsymbol{0})] |\boldsymbol{x}'_*| \frac{\nabla_{x'} W'(\boldsymbol{x}'_*)}{|\nabla_{x'} W'(\boldsymbol{x}'_*)|} = -\text{sign}[W'(\boldsymbol{0})] |\boldsymbol{x}'_*| \boldsymbol{\alpha}_* \qquad (4.37)$$

in which $\boldsymbol{\alpha}_* = \nabla_{x'} W'(\boldsymbol{x}'_*)/|\nabla_{x'} W'(\boldsymbol{x}'_*)|$ is a unit vector emanating from the design point $\boldsymbol{x}'_*$ and pointing toward the safe region. Referring to Fig. 4.6, where the mean point $\boldsymbol{\mu}_x$ is located in the safe region, hence $W'(\boldsymbol{0}) > 0$ [or $W(\boldsymbol{\mu}_x) > 0$], and the corresponding $-\text{sign}[W'(\boldsymbol{0})]\boldsymbol{\alpha}_*$ is a unit vector emanating from the origin $\boldsymbol{x}' = \boldsymbol{0}$ and pointing to the design point $\boldsymbol{x}_*$. The elements of $\boldsymbol{\alpha}_*$ are called the *directional derivatives* representing the value of the cosine angle between the gradient vector $\nabla_{x'} W'(\boldsymbol{x}'_*)$ and axes of the standardized variables. Geometrically, Eq. (4.37) shows that the vector, $\boldsymbol{x}'_*$ is perpendicular to the tangent hyperplane passing through the design point. The shortest distance can be expressed as

$$|\boldsymbol{x}'_*| = -\text{sign}[W'(\boldsymbol{0})] \boldsymbol{\alpha}^t_* \boldsymbol{x}'_* = -\text{sign}[W'(\boldsymbol{0})] \frac{\sum_{k=1}^{K} \left[\frac{\partial W'(\boldsymbol{x}')}{\partial x'_k}\right]_{\boldsymbol{x}'_*} x'_{k*}}{\sqrt{\sum_{j=1}^{K} \left[\frac{\partial W'(\boldsymbol{x}')}{\partial x'_j}\right]^2_{\boldsymbol{x}'_*}}} \qquad (4.38)$$

Recall that $X_k = \mu_k + \sigma_k X'_k$, for $k = 1, 2, \ldots, K$. By the chain rule in calculus,

$$\frac{\partial W'(\boldsymbol{X}')}{\partial X'_k} = \frac{\partial W(\boldsymbol{X})}{\partial X_k} \frac{\partial X_k}{\partial X'_k} = \frac{\partial W(\boldsymbol{X})}{\partial X_k} \sigma_k \tag{4.39a}$$

or in matrix form as

$$\nabla_{x'} W'(\boldsymbol{X}') = \boldsymbol{D}_x^{1/2} \nabla_x W(\boldsymbol{X}) \tag{4.39b}$$

Then Eq. (4.38) can be written, in terms of the original stochastic basic variables $\boldsymbol{X}$, as

$$|\boldsymbol{x}'_*| = \text{sign}[W'(\boldsymbol{0})] \frac{\sum_{k=1}^{K} \left[\frac{\partial W(\boldsymbol{x})}{\partial x_k} \right]_{\boldsymbol{x}_*} (\mu_k - x_{k*})}{\sqrt{\sum_{j=1}^{K} \left[\frac{\partial W(\boldsymbol{x})}{\partial x_j} \right]_{\boldsymbol{x}_*}^2 \sigma_j^2}} \tag{4.40}$$

in which $\boldsymbol{x}_* = (x_{1*}, x_{2*}, \ldots, x_{K*})^t$ is the point in the original variable $\boldsymbol{x}$-space that can be easily determined from the design point $\boldsymbol{x}'_*$ in x'-space as $\boldsymbol{x}_* = \boldsymbol{\mu}_x + \boldsymbol{D}_x^{1/2} \boldsymbol{x}'_*$. It will be shown in the next subsection that the shortest distance from the origin to the design point $|\boldsymbol{x}'_*|$, in fact, is the absolute value of the reliability index based on the first-order Taylor series expansion of the performance function $W(\boldsymbol{X})$ with the expansion point at $\boldsymbol{x}_*$.

Example 4.8 (Linear performance function) Consider that the failure surface is a hyperplane given by

$$W(\boldsymbol{X}) = a_0 + \sum_{k=1}^{K} a_k X_k$$

or in vector form as $W(\boldsymbol{X}) = a_0 + \boldsymbol{a}^t \boldsymbol{X} = 0$, with a's being the coefficients and $\boldsymbol{X}$ being the random variables. Assume that $\boldsymbol{X}$ are uncorrelated random variables with the mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{D}_x$. It can be shown that the MFOSM reliability index computed by Eq. (4.29) with $\mu_w = a_0 + \boldsymbol{a}^t \boldsymbol{\mu}_x$ and $\sigma_w^2 = \boldsymbol{a}^t \boldsymbol{D}_x \boldsymbol{a}$ is the AFOSM reliability index.

To show that the original random variables $\boldsymbol{X}$ are first standardized by Eq. (4.30), therefore, in terms of the standardized random variables $\boldsymbol{X}'$, the preceding linear failure surface can be expressed as

$$W'(\boldsymbol{X}') = b_0 + \boldsymbol{b}^t \boldsymbol{X}' = 0$$

in which $b_0 = a_0 + \boldsymbol{a}^t \boldsymbol{\mu}_x$ and $\boldsymbol{b}^t = \boldsymbol{a}^t \boldsymbol{D}_x^{1/2}$. In Fig. 4.7, let the lower half space containing the origin of $\boldsymbol{x}'$-space be designated as the safe region. This would require $b_0 = a_0 + \boldsymbol{a}^t \boldsymbol{\mu}_x > 0$.

Referring to Fig. 4.7, the gradient of $W'(\boldsymbol{X}')$ is $\boldsymbol{b}$, which is a vector perpendicular to the failure hyperplane defined by $W'(\boldsymbol{X}') = 0$ pointing in the direction of the safe set. Therefore, the vector $-\boldsymbol{a} = -\boldsymbol{b}/\sqrt{\boldsymbol{b}^t \boldsymbol{b}}$ is a unit vector emanating from $\boldsymbol{x}' = \boldsymbol{0}$ toward the failure region, as shown in Fig. 4.7. For any vector $\boldsymbol{x}'$ landing on the
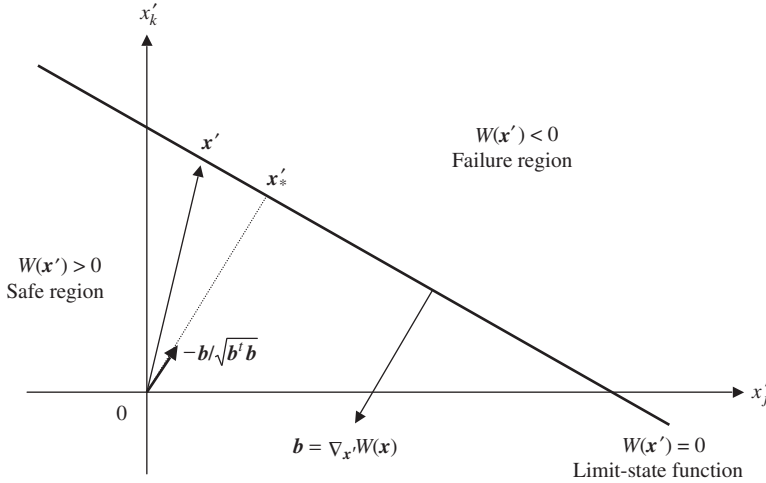
Figure 4.7 A linear performance function in the standardized space.

failure hyperplane defined by $W'(\boldsymbol{x}') = 0$, the following relationship holds:

$$\frac{-2\boldsymbol{b}^t \boldsymbol{x}'}{\sqrt{\boldsymbol{b}^t \boldsymbol{b}}} = \frac{b_0}{\sqrt{\boldsymbol{b}^t \boldsymbol{b}}}$$

Note that the left-hand side is the length of the vector $\boldsymbol{x}'$ projected on the unit vector $-\boldsymbol{b}/\sqrt{\boldsymbol{b}^t \boldsymbol{b}}$, which is the shortest distance from $\boldsymbol{x}' = \boldsymbol{0}$ to the failure hyperplane. Therefore, $b_0/\sqrt{\boldsymbol{b}^t \boldsymbol{b}}$ is the reliability index, that is,

$$\beta = \frac{b_0}{\sqrt{\boldsymbol{b}^t \boldsymbol{b}}} = \frac{a_0 + \boldsymbol{a}^t \mu_x}{\sqrt{\boldsymbol{a}^t \boldsymbol{D}_x \boldsymbol{a}}} = \frac{\mu_w}{\sigma_w}$$

As shown, when the performance function is linear involving uncorrelated stochastic basic variables, the reliability index is the ratio of the expected value of the performance function to its standard deviation. Furthermore, the MFOSM method would yield the same results as the AFOSM method.

4.5.3 First-order approximation of performance function at the design point

Referring to Eqs. (4.20) and (4.21), the first-order approximation of the performance function $W(\boldsymbol{X})$, taking the expansion point $\boldsymbol{x}_o = \boldsymbol{x}_*$, is

$$W(\boldsymbol{X}) \approx \sum_{k=1}^{K} s_{k*}(X_k - x_{k*}) = \boldsymbol{s}_*^t(\boldsymbol{X} - \boldsymbol{x}_*) \qquad (4.41)$$

in which $\boldsymbol{s}_* = (s_{1*}, s_{2*}, \ldots, s_{K*})^t$, a vector of sensitivity coefficients of the performance function $W(\boldsymbol{X})$ evaluated at the expansion point $\boldsymbol{x}_*$ that lies on the

limit-state surface, that is,

$$s_{k^*} = \left[\frac{\partial W(\boldsymbol{X})}{\partial X_k} \right]_{X=x_*} \qquad \text{for } k = 1, 2, \ldots, K$$

note that $W(\boldsymbol{x}_*)$ is not on the right-hand-side of Eq. (4.41) because $W(\boldsymbol{x}_*) = 0$. Hence, at the expansion point $\boldsymbol{x}_*$, the expected value and the variance of the performance function $W(\boldsymbol{X})$ can be approximated according to Eqs. (4.24) and (4.25) as

$$\mu_w \approx \boldsymbol{s}_*^t(\boldsymbol{\mu}_x - \boldsymbol{x}_*) \tag{4.42}$$

$$\sigma_w^2 \approx \boldsymbol{s}_*^t \boldsymbol{C}_x \boldsymbol{s}_* \tag{4.43}$$

in which $\boldsymbol{\mu}_x$ and $\boldsymbol{C}_x$ are the mean vector and covariance matrix of the stochastic basic variables, respectively. If the stochastic basic variables are uncorrelated, Eq. (4.43) reduces to

$$\sigma_w^2 = \sum_{k=1}^{K} s_{k*}^2 \sigma_k^2 \tag{4.44}$$

in which σ_k is the standard deviation of the kth stochastic basic variable X_k.

Since $\boldsymbol{\alpha}_* = \boldsymbol{s}_*/|\boldsymbol{s}_*|$, when stochastic basic variables are uncorrelated, the standard deviation of the performance function $W(\boldsymbol{X})$ alternatively can be expressed in terms of the directional derivatives as

$$\sigma_w = \sum_{k=1}^{K} \alpha_{k*} s_{k*} \sigma_k \tag{4.45}$$

where α_{k*} is the directional derivative for the kth stochastic basic variable at the expansion point $\boldsymbol{x}_*$

$$\alpha_{k*} = \frac{s_{k*}\sigma_k}{\sqrt{\sum_{j=1}^{K} s_{j*}^2 \sigma_j^2}} \qquad \text{for } k = 1, 2, \ldots, K \tag{4.46a}$$

or, in matrix form,

$$\boldsymbol{\alpha}_* = \frac{\boldsymbol{D}_x^{1/2} \nabla_x \boldsymbol{W}(\boldsymbol{x}_*)}{|\boldsymbol{D}_x^{1/2} \nabla_x \boldsymbol{W}(\boldsymbol{x}_*)|} \tag{4.46b}$$

which is identical to the one defined in Eq. (4.37) according to Eq. (4.39). With the mean and standard deviation of the performance function $W(\boldsymbol{X})$ computed

at $\boldsymbol{x}_*$, the AFOSM reliability index β_{AFOSM} given in Eq. (4.34) can be determined as

$$\beta_{\text{AFOSM}} = \frac{\mu_w}{\sigma_w} = \frac{\sum_{k=1}^{K} s_{k*}(\mu_k - x_{k*})}{\sum_{k=1}^{K} \alpha_{k*} s_{k*} \sigma_k} \tag{4.47}$$

The reliability index β_{AFOSM} also is called the *Hasofer-Lind reliability index*.

Once the value of β_{AFOSM} is computed, the reliability can be estimated by Eq. (4.10) as $p_s = \Phi(\beta_{\text{AFOSM}})$. Since $\beta_{\text{AFOSM}} = \text{sign}[W'(\mathbf{0})]|\boldsymbol{x}'_*|$, the sensitivity of β_{AFOSM} with respect to the uncorrelated, standardized stochastic basic variables is

$$\nabla_{x'} \beta_{\text{AFOSM}} = \text{sign}[W'(\mathbf{0})]\nabla_{x'}|\boldsymbol{x}'_*| = \text{sign}[W'(\mathbf{0})]\frac{\boldsymbol{x}'_*}{|\boldsymbol{x}'_*|} = -\boldsymbol{\alpha}_* \tag{4.48}$$

Note that $\nabla_{x'}\beta$ is a vector showing the direction along which the rate change in the value of the reliability index β increases most rapidly. This direction is indicated by $-\boldsymbol{\alpha}_*$ regardless whether the position of the mean of the stochastic basic variables $\boldsymbol{\mu}_x$ is in the safe region $W'(\mathbf{0}) > 0$ or failure zone $W'(\mathbf{0}) < 0$. As shown in Fig. 4.8, the vector $-\boldsymbol{\alpha}_*$ points to the failure region, and moving along $-\boldsymbol{\alpha}_*$ would result in a more negative-valued $W'(\boldsymbol{x}')$. This is, geometrically, equivalent to pushing the limit-state surface $W'(\boldsymbol{x}') = 0$ further away from $\boldsymbol{x}' = \mathbf{0}$ in Fig. 4.8a and closer to $\boldsymbol{x}' = \mathbf{0}$ in Fig. 4.8b. Hence, moving along the direction of $-\boldsymbol{\alpha}_*$ at the design point $\boldsymbol{x}_*$ would make the value of the reliability index β more positive under $W'(\mathbf{0}) > 0$, whereas the value of β would be less negative under $W'(\mathbf{0}) < 0$.

In both cases, the value of the reliability index increases along $-\boldsymbol{\alpha}_*$. Algebraically, as one moves along $-\boldsymbol{\alpha}_*$, the current value of the limit-state surface $W'(\boldsymbol{x}')$ changes from 0 to a negative value, that is, $W'(\boldsymbol{x}') = -c$, for $c > 0$. This implies a new limit state for the system defined by $W'(\boldsymbol{x}') = R(\boldsymbol{x}') - L(\boldsymbol{x}') + c = 0$. The introduction of a positive-valued c in the performance function could mean an increase in resistance, that is, $W'(\boldsymbol{x}') = [R(\boldsymbol{x}') + c] - L(\boldsymbol{x}') = 0$, or a decrease in load, that is, $W'(\boldsymbol{x}') = R(\boldsymbol{x}') - [L(\boldsymbol{x}') - c] = 0$. In either case, the reliability index and the corresponding reliability for the system would increase along the direction of $-\boldsymbol{\alpha}_*$.

Equation (4.48) indicates that moving along the direction of $\boldsymbol{\alpha}_*$ at the design point $\boldsymbol{x}_*$, the values of the reliability index would decrease and that $-\alpha_{k*}$ is the rate of change in β_{AFOSM} owing to a *one standard deviation* change in stochastic basic variable X_k at $\boldsymbol{X} = \boldsymbol{x}_*$. Therefore, the relationship between $\nabla_{x'}\beta$ and $\nabla_x \beta$ can be expressed as

$$-\alpha_{k*} = \left(\frac{\partial \beta_{\text{AFOSM}}}{\partial X'_k}\right)_{\boldsymbol{x}'_*} = \left(\frac{\partial \beta_{\text{AFOSM}}}{\partial X_k}\right)_{\boldsymbol{x}_*} \sigma_k \qquad \text{for } k = 1, 2, \ldots, K \tag{4.49a}$$

or, in matrix form, as

$$\nabla_x \beta_{\text{AFOSM}} = \boldsymbol{D}_x^{-1/2} \nabla_{x'} \beta_{\text{AFOSM}} = -\boldsymbol{D}_x^{-1/2} \boldsymbol{\alpha}_* \tag{4.49b}$$
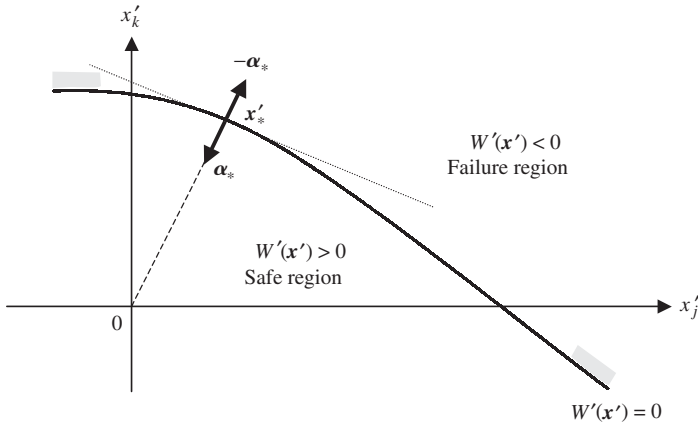
(a)
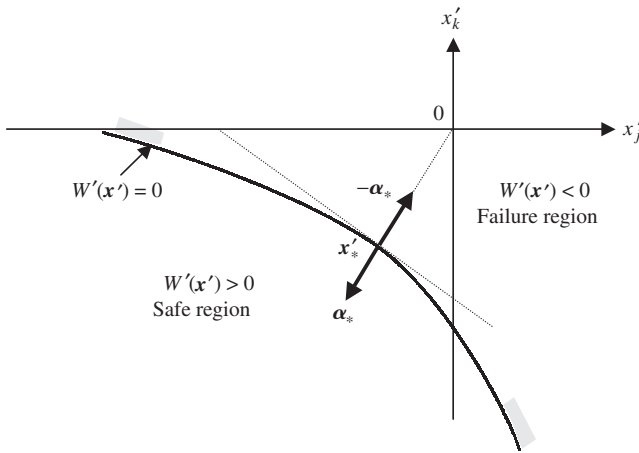
(b)

Figure 4.8 Sensitivity of reliability index: (*a*) under $W'(\mathbf{0}) > 0$ (that is, $W(\boldsymbol{\mu}_x) > 0$); (*b*) under $W'(\mathbf{0}) < 0$ (that is, $W(\boldsymbol{\mu}_x) < 0$).

It also can be shown easily that the sensitivity of reliability or failure probability with respect to each stochastic basic variable along the direction of $\boldsymbol{\alpha}_*$ can be computed as

$$
\begin{aligned}
\left(\frac{\partial p_s}{\partial X'_k}\right)_{\boldsymbol{x}'_*} &= -\alpha_{k*}\phi(\beta_{\text{AFOSM}}) \\
\left(\frac{\partial p_s}{\partial X_k}\right)_{\boldsymbol{x}_*} &= -\frac{\alpha_{k*}\phi(\beta_{\text{AFOSM}})}{\sigma_k}
\end{aligned}
\tag{4.50a}
$$

or in matrix form as

$$\nabla_{x'_*} p_s = -\phi(\beta_{\text{AFOSM}})\boldsymbol{\alpha}_*$$

$$\nabla_{x_*} p_s = \phi(\beta_{\text{AFOSM}})\nabla_{x_*}\beta_{\text{AFOSM}} = -\phi(\beta_{\text{AFOSM}})\boldsymbol{D}_x^{-1/2}\boldsymbol{\alpha}_*$$

(4.50b)

These sensitivity coefficients would reveal the relative importance of each stochastic basic variable for their effects on reliability or failure probability.

4.5.4 Algorithms of AFOSM for independent normal parameters

Hasofer-Lind algorithm.   In the case that $\boldsymbol{X}$ are independent normal stochastic basic variables, standardization of $\boldsymbol{X}$ according to Eq. (4.30) reduces them to independent standard normal random variables $\boldsymbol{Z}'$ with mean 0 and covariance matrix $\boldsymbol{I}$, with $\boldsymbol{I}$ being a $K \times K$ identity matrix. Referring to Fig. 4.8, based on the geometric characteristics at the design point on the failure surface, Hasofer and Lind (1974) proposed the following recursive equation for determining the design point $\boldsymbol{z}'_*$.

$$\boldsymbol{z}'_{(r+1)} = -\left(-\boldsymbol{\alpha}_{(r)}^t\,\boldsymbol{z}'_{(r)}\right)\boldsymbol{\alpha}_{(r)} - \frac{W'(\boldsymbol{z}'_{(r)})}{|\nabla_{z'}W'(\boldsymbol{z}'_{(r)})|}\boldsymbol{\alpha}_{(r)} \qquad \text{for } r = 1, 2, \ldots \quad (4.51)$$

in which subscripts (r) and $(r + 1)$ represent the iteration numbers, and $-\boldsymbol{\alpha}$ denotes the unit gradient vector on the failure surface pointing to the failure region. Referring to Fig. 4.9, the first terms of Eq. (4.51), $-(-\boldsymbol{\alpha}_{(r)}^t\boldsymbol{z}'_{(r)})\boldsymbol{\alpha}_{(r)}$, is a projection vector of the old solution vector $\boldsymbol{z}'_{(r)}$ onto the vector $-\boldsymbol{\alpha}_{(r)}$ emanating from the origin. The quantity $W'(\boldsymbol{z}'_{(r)})/|\nabla W'(\boldsymbol{z}'_{(r)})|$ is the step size to move from $W'(\boldsymbol{z}'_{(r)})$ to $W'(\boldsymbol{z}') = 0$ along the direction defined by the vector $-\boldsymbol{\alpha}_{(r)}$. The second term is a correction that further adjusts the revised solution closer to the limit-state surface. It would be more convenient to rewrite the preceding recursive equation in the original $\boldsymbol{x}$-space as

$$\boldsymbol{x}_{(r+1)} = \boldsymbol{\mu}_x + \boldsymbol{D}_x\boldsymbol{s}_{(r)}\frac{(\boldsymbol{x}_{(r)} - \boldsymbol{\mu}_x)^t\boldsymbol{s}_{(r)} - W(\boldsymbol{x}_{(r)})}{\boldsymbol{s}_{(r)}^t\boldsymbol{D}_x\boldsymbol{s}_{(r)}} \qquad \text{for } r = 1, 2, 3, \ldots \quad (4.52)$$

Based on Eq. (4.52), the Hasofer-Lind AFOSM reliability analysis algorithm for problems involving uncorrelated, normal stochastic variables the can be outlined as follows:

Step 1: Select an initial trial solution $\boldsymbol{x}_{(r)}$.

Step 2: Compute $W(\boldsymbol{x}_{(r)})$ and the corresponding sensitivity coefficient vector $\boldsymbol{s}_{(r)}$.

Step 3: Revise solution point $\boldsymbol{x}_{(r+1)}$, according to Eq. (4.52).

Step 4: Check if $\boldsymbol{x}_{(r)}$ and $\boldsymbol{x}_{(r+1)}$ are sufficiently close. If yes, compute the reliability index β_{AFOSM} according to Eq. (4.47) and the corresponding reliability
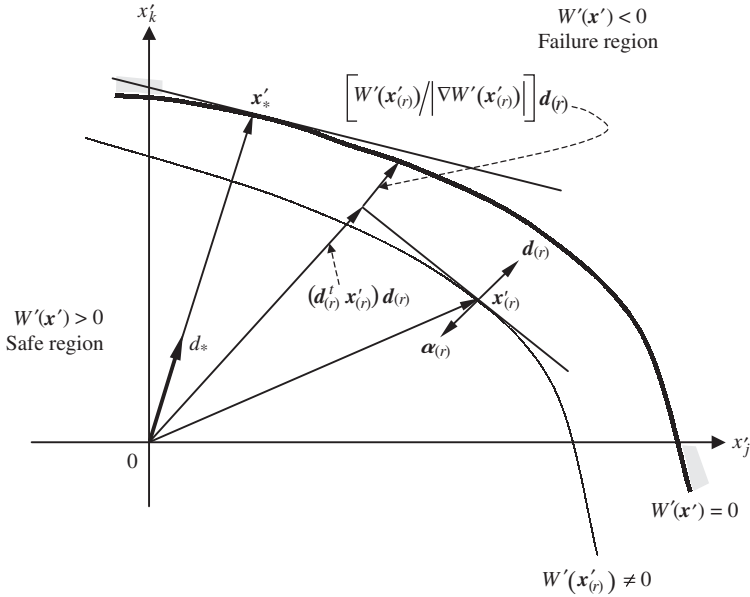
Figure 4.9 Geometric interpretation of Hasofer-Lind algorithm in standardized space.

$p_s = \Phi(\beta_{\mathrm{AFOSM}})$; then, go to step 5. Otherwise, update the solution point by letting $\boldsymbol{x}_{(r)} = \boldsymbol{x}_{(r+1)}$ and return to step 2.

Step 5: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables according to Eqs. (4.48), (4.49), and (4.50).

It is possible that a given performance function might have several design points. In the case that there are J such design points, the reliability can be calculated as

$$p_s = [\Phi(\beta_{\mathrm{AFOSM}})]^{J} \tag{4.53}$$

Ang-Tang algorithm.   The core of the updating procedure of Ang and Tang (1984) relies on the fact that according to Eq. (4.47), the following relationship should be satisfied:

$$\sum_{k=1}^{K} s_{k*} \left(\mu_k - x_{k*} - \alpha_{k*} \beta_* \sigma_k \right) = 0 \tag{4.54}$$

Since the variables $\boldsymbol{X}$ are random and uncorrelated, Eq. (4.35) defines the failure point within the first-order context. Hence Eq. (4.47) can be decomposed into

$$x_{k*} = \mu_k - \alpha_{k*} \beta_* \sigma_k \qquad \text{for } k = 1, 2, \ldots, K \tag{4.55}$$

Ang and Tang (1984) present the following iterative procedure to locate the design point x_* and the corresponding reliability index β_{AFOSM} under the condition that stochastic basic variables are independent normal random variables. The Ang-Tang AFOSM reliability algorithm for problems involving uncorrelated normal stochastic variables has the following steps (Fig. 4.10):

Step 1: Select an initial point $x_{(r)}$ in the parameter space. For practicality, the point μ_x where the means of stochastic basic variables are located is a viable starting point.
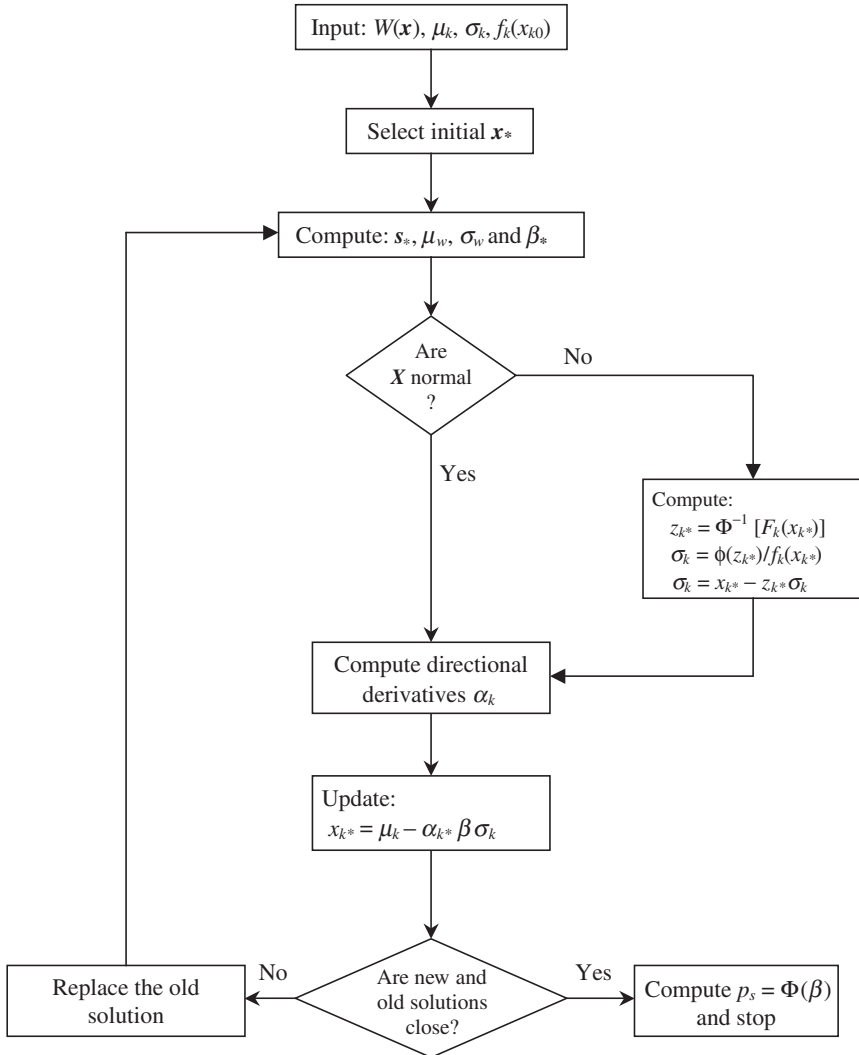
Figure 4.10 Flowchart of the Ang-Tang AFOSM reliability analysis involving uncorrelated variables.

Step 2: At the selected point $\boldsymbol{x}_{(r)}$, compute the mean of the performance function $W(\boldsymbol{X})$ by

$$\mu_w = W(\boldsymbol{x}_{(r)}) + \boldsymbol{s}_{(r)}^t (\boldsymbol{\mu}_x - \boldsymbol{x}_{(r)}) \qquad (4.56)$$

and the variance according to Eq. (4.44).

Step 3: Compute the corresponding reliability index $\beta_{(r)}$ according to Eq. (4.34).

Step 4: Compute the values of directional derivative α_k for all $k = 1, 2, \ldots, K$ according to Eq. (4.46).

Step 5: Revise the location of expansion point $\boldsymbol{x}_{(r+1)}$ according to Eq. (4.56) using α_k and $\beta_{(r)}$ obtained from steps 3 and 4.

Step 6: Check if the revised expansion point $\boldsymbol{x}_{(r+1)}$ differs significantly from the previous trial expansion point $\boldsymbol{x}_{(r)}$. If yes, use the revised expansion point as the new trial point by letting $\boldsymbol{x}_{(r)} = \boldsymbol{x}_{(r+1)}$, and go to step 2 for an additional iteration. Otherwise, the iteration procedure is considered complete, and the latest reliability index is β_{AFOSM} and is to be used in Eq. (4.10) to compute the reliability p_s.

Step 7: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables according to Eqs. (4.48), (4.49), and (4.50).

Referring to Eq. (4.8), the reliability is a monotonically increasing function of the reliability index β, which, in turn, is a function of the unknown failure point. The task to determine the critical failure point $\boldsymbol{x}_*$ that minimizes the reliability is equivalent to minimizing the value of the reliability index β. Low and Tang (1997), based on Eqs. (4.31a) and (4.31b) developed an optimization procedure in Excel by solving

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\mathrm{Min}} & \beta = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu}_x)^t \boldsymbol{C}_x^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_x)} \\ \mathrm{subject\ to} & W(\boldsymbol{x}) = 0 \end{array} \qquad (4.57)$$

Owing to the nature of nonlinear optimization, both AFOSM-HL and AFOSM-AT algorithms do not necessarily converge to the true design point associated with the minimum reliability index. Madsen et al. (1986) suggested that different initial trial points be used and that the smallest reliability index be chosen to compute the reliability. To improve the convergence of the Hasofer-Lind algorithm, Liu and Der Kiureghian (1991) proposed a modified objective function for Eq. (4.31a) using a nonnegative merit function.

Example 4.9 (Uncorrelated, normal) Refer to Example 4.5 for a storm sewer reliability analysis problem with the following data:

Variable	Mean	Coefficient of variation
n (ft$^{1/6}$)	0.015	0.05
D (ft)	3.0	0.02
S (ft/ft)	0.005	0.05

Assume that all three stochastic basic variables are independent normal random variables. Compute the reliability that the sewer can convey an inflow discharge of 35 ft^3/s using the AFOSM-HL algorithm.

Solution   The initial solution is taken to be the means of the three stochastic basic variables, namely, $\boldsymbol{x}_{(1)} = \boldsymbol{\mu}_x = (\mu_n, \mu_D, \mu_S)^t = (0.015, 3.0, 0.005)^t$. The covariance matrix for the three stochastic basic variables is

$$\boldsymbol{D}_x = \begin{bmatrix} \sigma_n^2 & 0 & 0 \\ 0 & \sigma_D^2 & 0 \\ 0 & 0 & \sigma_S^2 \end{bmatrix} = \begin{bmatrix} 0.00075^2 & 0 & 0 \\ 0 & 0.06^2 & 0 \\ 0 & 0 & 0.00025^2 \end{bmatrix}$$

For this example, the performance function $Q_C - Q_L$ is

$$W(n, D, S) = Q_C - Q_L = 0.463 n^{-1} D^{8/3} S^{1/2} - 35$$

Note that because the $W(\mu_n, \mu_D, \mu_S) = 6.010 > 0$, the mean point $\boldsymbol{\mu}_x$ is located in the safe region. At $\boldsymbol{x}_{(1)} = \boldsymbol{\mu}_x$, the value of the performance function $W(n, D, S) = 6.010$, which is not equal to zero. This implies that the solution point $\boldsymbol{x}_{(1)}$, does not lie on the limit-state surface. By Eq. (4.52), the new solution $\boldsymbol{x}_{(2)}$ can be obtained as $\boldsymbol{x}_{(2)} = (0.01592, 2.921, 0.004847)$. Then one checks the difference between the two consecutive solution points as

$$\delta = |\boldsymbol{x}_{(1)} - \boldsymbol{x}_{(2)}| = [(0.01592 - 0.015)^2 + (2.921 - 3.0)^2 + (0.004847 - 0.005)^2]^{0.5}$$
$$= 0.07857$$

which is considered large, and therefore, the iteration continues. The following table lists the solution point $\boldsymbol{x}_{(r)}$, its corresponding sensitivity vector $\boldsymbol{s}_{(r)}$, and the vector of directional derivatives $\boldsymbol{\alpha}_{(r)}$ in each iteration. The iteration stops when the difference between the two consecutive solutions is less than 0.001 and the value of the performance function is less than 0.001.

Iteration	Var.	$\boldsymbol{x}_{(r)}$	$\boldsymbol{s}_{(r)}$	$\boldsymbol{\alpha}_{(r)}$	$\boldsymbol{x}_{(r+1)}$
$r = 1$	n	0.1500×10^{-01}	$-0.2734 \times 10^{+04}$	$-0.6468 \times 10^{+00}$	0.1592×10^{-01}
	D	0.3000×10^{-01}	$0.3650 \times 10^{+02}$	$0.6907 \times 10^{+00}$	$0.2921 \times 10^{+01}$
	S	0.5000×10^{-02}	$0.4101 \times 10^{+04}$	$0.3234 \times 10^{+00}$	0.4847×10^{-02}
		$\delta = 0.7857 \times 10^{-01}$	$W = 0.6010 \times 10^{+01}$	$\beta = 0.0000 \times 10^{+00}$	
$r = 2$	n	0.1592×10^{-01}	$-0.2226 \times 10^{+04}$	$-0.6138 \times 10^{+00}$	0.1595×10^{-01}
	D	$0.2921 \times 10^{+01}$	$0.3239 \times 10^{+02}$	$0.7144 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4847×10^{-02}	$0.3656 \times 10^{+04}$	$0.3360 \times 10^{+00}$	0.4827×10^{-02}
		$\delta = 0.9584 \times 10^{-02}$	$W = 0.4421 \times 10^{+00}$	$\beta = 0.1896 \times 10^{+01}$	

(*Continued*)

Iteration	Var.	$\boldsymbol{x}_{(r)}$	$\boldsymbol{s}_{(r)}$	$\boldsymbol{\alpha}_{(r)}$	$\boldsymbol{x}_{(r+1)}$
$r = 3$	n	0.1595×10^{-01}	$-0.2195 \times 10^{+04}$	$-0.6118 \times 10^{+00}$	0.1594×10^{-01}
	D	0.2912×10^{-01}	$0.3209 \times 10^{+02}$	$0.7157 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4827×10^{-02}	$0.3625 \times 10^{+04}$	$0.3369 \times 10^{+00}$	0.4827×10^{-02}
		$\delta = 0.1919 \times 10^{-03}$	$W = 0.2151 \times 10^{-02}$	$\beta = 0.2056 \times 10^{+01}$	
$r = 4$	n	0.1594×10^{-01}	$-0.2195 \times 10^{+04}$	$-0.6119 \times 10^{+00}$	0.1594×10^{-01}
	D	$0.2912 \times 10^{+01}$	$0.3210 \times 10^{+02}$	$0.7157 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4827×10^{-02}	$0.3626 \times 10^{+04}$	$0.3369 \times 10^{+00}$	0.4827×10^{-02}
		$\delta = 0.3721 \times 10^{-05}$	$W = 0.2544 \times 10^{-06}$	$\beta = 0.2057 \times 10^{+01}$	

After four iterations, the solution converges to the design point $\boldsymbol{x}_* = (n_*, D_*, S_*)^t = (0.01594, 2.912, 0.004827)^t$. At the design point $\boldsymbol{x}_*$, the mean and standard deviation of the performance function W can be estimated, by Eqs. (4.42) and (4.43), respectively, as

$$\mu_{w^*} = 5.536 \qquad \text{and} \qquad \sigma_{w^*} = 2.691$$

The reliability index then can be computed as $\beta_* = \mu_{w^*}/\sigma_{w^*} = 2.057$, and the corresponding reliability and failure probability can be computed, respectively, as

$$p_s = \Phi(\beta_*) = 0.9802 \qquad p_f = 1 - p_s = 0.01983$$

Finally, at the design point $\boldsymbol{x}_*$, the sensitivity of the reliability index and reliability with respect to each of the three stochastic basic variables can be computed by Eqs. (4.49) and (4.50). The results are shown in columns (4) to (7) of the following table:

Variable (1)	$\boldsymbol{x}_*$ (2)	$\boldsymbol{\alpha}_*$ (3)	$\partial\beta/\partial x'$ (4)	$\partial p_s/\partial x'$ (5)	$\partial\beta/\partial x$ (6)	$\partial p_s/\partial x$ (7)	$x\partial\beta/\beta\partial x$ (8)	$x\partial p_s/p_s\partial x$ (9)
n	0.01594	−0.6119	0.6119	0.02942	815.8	39.22	6.323	0.638
D	2.912	0.7157	−0.7157	−0.03441	−11.9	−0.57	−16.890	−1.703
S	0.00483	0.3369	−0.3369	−0.01619	−1347.0	−64.78	−3.161	−0.319

From the preceding table, the quantities $\partial\beta/\partial x'_k$ and $\partial p_s/\partial x'_k$ show the sensitivity of the reliability index and reliability for one standard deviation change in the k-th stochastic basic variable, whereas $\partial\beta/\partial x_k$ and $\partial p_s/\partial x_k$ correspond to one unit change of the k-th stochastic basic variables in the original space. As can be seen, the sensitivity of β and p_s associated with Manning's roughness coefficient is positive, whereas those for pipe size and slope are negative. This indicates that an increase in Manning's roughness coefficient would result in an increase in β and p_s, whereas an increase in slope and/or pipe size would decrease β and p_s. The indication is confusing from a physical viewpoint because an increase in Manning's roughness coefficient would decrease the flow-carrying capacity of the sewer, whereas, on the other hand, an increase in pipe diameter and/or pipe slope would increase the sewer's conveyance capacity. The problem is that the sensitivity coefficients for β and p_s are taken relative to the design point on the failure surface; i.e., a larger Manning's would be farther from the system's mean condition, thus resulting in a larger value of β. However, larger values of pipe diameter or slope would be closer to the system's mean condition, thus resulting in a smaller value of β. Thus the sign of the sensitivity coefficients is deceiving, but their magnitude is useful, as described in the following paragraphs.

Furthermore, one can judge the relative importance of each stochastic basic variable based on the absolute values of the sensitivity coefficients. It is generally difficult to draw a meaningful conclusion based on the relative magnitude of $\partial\beta/\partial x$ and $\partial p_s/\partial x$ because units of different stochastic basic variables are not the same. Therefore, sensitivity measures not affected by the dimension of the stochastic basic variables, such as $\partial\beta/\partial x'$ and $\partial p_s/\partial x'$, generally are more useful. With regard to a one standard deviation change, for example, pipe diameter is significantly more important than pipe slope.

An alternative sensitivity measure, called the *relative sensitivity* or the *partial elasticity* (Breitung, 1993), is defined as

$$s_{k\%} = \frac{\partial y/y}{\partial x_k/x_k} = \left(\frac{\partial y}{\partial x_k}\right)\left(\frac{x_k}{y}\right) \qquad \text{for } k = 1, 2, \ldots, K \qquad (4.58)$$

in which $s_{k\%}$ is a dimensionless quantity measuring the percentage change in the dependent variable y due to 1 percent change in the variable x_k. The last two columns of the preceding table show the percentage change in β and p_s owing to 1 percent change in Manning's roughness, pipe diameter, and pipe slope. As can be observed, the most important stochastic basic variable in Manning's formula affecting the sewer's conveyance reliability is pipe diameter.

Figure 4.11 indicates that application of the AFOSM method removes the undesirable noninvariant behavior of the MFOSM method. The AFOSM method described in this section is suitable for the case that all stochastic basic variables in the load and resistance functions are independent normal random variables. In reality, stochastic basic variables in a performance function may be nonnormal and correlated. In the following two subsections, procedures to treat stochastic basic variables that are nonnormal and correlated are discussed.
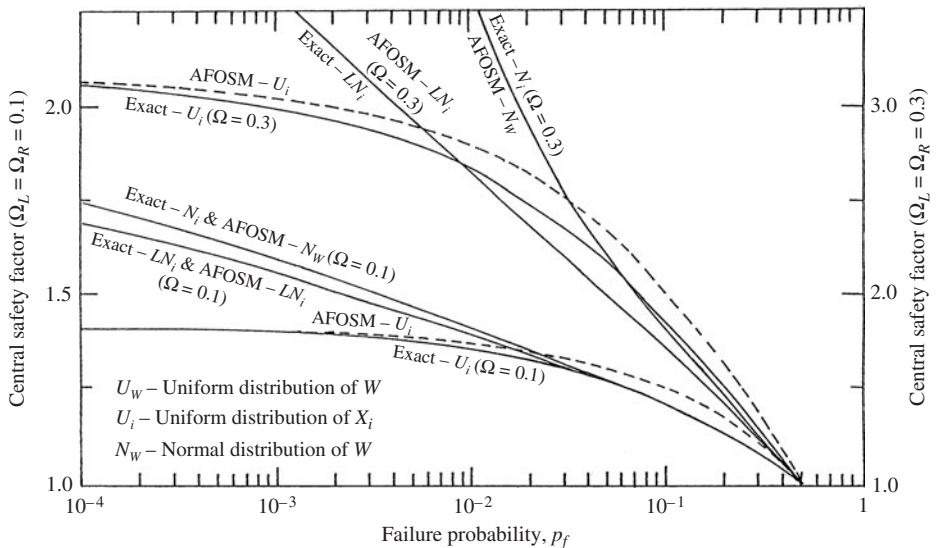
Figure 4.11 Comparison of the AFOSM reliability method with the exact solution.

4.5.5 Treatment of nonnormal stochastic variables

When nonnormal random variables are involved, it is advisable to transform them into equivalent normal variables. Rackwitz (1976) and Rackwitz and Fiessler (1978) proposed an approach that transforms a nonnormal distribution into an equivalent normal distribution so that the probability content is preserved. That is, the value of the CDF of the transformed equivalent normal distribution is the same as that of the original nonnormal distribution at the design point $\boldsymbol{x}_*$. Later, Ditlvesen (1981) provided the theoretical proof of the convergence property of the normal transformation in the reliability algorithms searching for the design point. Table 4.3 presents the normal equivalent for some nonnormal distributions commonly used in reliability analysis.

By the Rackwitz (1976) approach, the normal transform at the design point $\boldsymbol{x}_*$ satisfies the following condition:

$$F_k(x_{k^*}) = \Phi\left(\frac{x_{k^*} - \mu_{k^*N}}{\sigma_{k^*N}}\right) = \Phi(z_k^*) \qquad \text{for } k = 1, 2, \ldots, K \qquad (4.59)$$

in which $F_k(x_{k^*})$ is the marginal CDF of the stochastic basic variable X_k having values at x_{k^*}, μ_{k^*N} and σ_{k^*N} are the mean and standard deviationas of the normal equivalent for the kth stochastic basic variable at $X_k = x_{k^*}$, and $z_{k^*} = \Phi^{-1}[F_k(x_{k^*})]$ is the standard normal quantile. Equation (4.59) indicates that the marginal probability content in both the original and normal transformed spaces must be preserved. From Eq. (4.59), the following equation is obtained:

$$\mu_{k^*N} = x_{k^*} - z_{k^*}\sigma_{k^*N} \qquad (4.60)$$

Note that μ_{k^*N} and σ_{k^*N} are functions of the expansion point $\boldsymbol{x}_*$. To obtain the standard deviation in the equivalent normal space, one can take the derivative on both sides of Eq. (4.59) with respect to x_k, resulting in

$$f_k(x_{k^*}) = \frac{1}{\sigma_{k^*N}}\phi\left(\frac{x_k^* - \mu_{k^*N}}{\sigma_{k^*N}}\right) = \frac{\phi(z_k^*)}{\sigma_{k^*N}}$$

in which $f_k(\cdot)$ and $\phi(\cdot)$ are the marginal PDFs of the stochastic basic variable X_k and the standard normal variable Z_k, respectively. From this equation, the normal equivalent standard deviation σ_{k^*N} can be computed as

$$\sigma_{k^*N} = \frac{\phi(z_{k^*})}{f_k(x_{k^*})} \qquad (4.61)$$

Therefore, according to Eqs. (4.60) and (4.61), the mean and standard deviation of the normal equivalent of the stochastic basic variable X_k can be calculated.

It should be noted that the normal transformation uses only the marginal distributions of the stochastic basic variables without regarding their correlations. Therefore, it is, in theory, suitable for problems involving independent

TABLE 4.3 Normal Transformation of Selected Distributions

Distribution of X	PDF, $f_x(x_*)$	Equivalent standard normal variable $z_N = \Phi^{-1}[F_x(x_*)]$	σ_N
Lognormal	$\dfrac{1}{\sqrt{2\pi}\,x_*\sigma_{\ln x}}\exp\left\{-\dfrac{1}{2}\left[\dfrac{\ln(x_*)-\mu_{\ln x}}{\sigma_{\ln x}}\right]^2\right\}$ $x > 0$	$\dfrac{\ln(x_*)-\mu_{\ln x}}{\sigma_{\ln x}}$	$x_*\sigma_{\ln x}$
Exponential	$\beta e^{-\beta(x_*-x_o)}\quad x>x_o$	$\Phi^{-1}\left(1-e^{-\beta(x_*-x_o)}\right)$	$\dfrac{1}{\beta\sqrt{2\pi}}\exp\left[-\dfrac{z_*^2}{2}+\beta(x_*-x_o)\right]$
Gamma	$\dfrac{\beta^\alpha(x_*-\xi)^{\alpha-1}e^{-\beta(x_*-\xi)}}{\Gamma(\alpha)}\qquad x_*>\xi$	$\Phi^{-1}\left(1-e^{-\beta(x_*-\xi)}\sum_{j=0}^{\alpha-1}\dfrac{[\beta(x_*-\xi)]^j}{j!}\right)$	$\dfrac{\phi(z_*)}{f_x(x_*)}$
Type 1 extremal (max)	$\dfrac{1}{\beta}\exp\left[-\left(\dfrac{x-\xi}{\beta}\right)-\exp\left(\dfrac{x-\xi}{\beta}\right)\right]$ $-\infty < x < \infty$	$\Phi^{-1}\left\{\exp\left[-\exp\left(-\dfrac{x-\xi}{\beta}\right)\right]\right\}$	$\dfrac{\phi(z_*)}{f_x(x_*)}$
Triangular	$\begin{cases}\dfrac{2}{b-a}\left(\dfrac{x_*-a}{m-a}\right) & a\le x\le m \\[2ex] \dfrac{2}{b-a}\left(\dfrac{b-x_*}{b-m}\right) & m\le x\le b\end{cases}$	$\begin{cases}\Phi^{-1}\left[\dfrac{(x_*-a)^2}{(b-a)(m-a)}\right] & a\le x\le m \\[2ex] \Phi^{-1}\left[1-\dfrac{(b-x_*)^2}{(b-a)(b-m)}\right] & m\le x\le b\end{cases}$	$\dfrac{\phi(z_*)}{f_x(x_*)}$
Uniform	$\dfrac{1}{b-a}\qquad a\le x\le b$	$\Phi^{-1}\left(\dfrac{x_*-a}{b-a}\right)$	$(b-a)\phi(z_*)$

NOTE: In all cases, $\mu_N = x_* - z_*\sigma_N$.
SOURCE: After Yen et al. (1986).

nonnormal random variables. When stochastic basic variables are nonnormal but correlated, additional considerations must be given in the normal transformation (see Sec. 4.5.7).

To incorporate the normal transformation for nonnormal uncorrelated stochastic basic variables, the Hasofer-Lind AFOSM algorithm for problems having uncorrelated nonnormal stochastic variables involves the following steps:

Step 1: Select an initial trial solution $\boldsymbol{x}_{(r)}$.

Step 2: Compute the mean and standard deviation of the normal equivalent using Eqs. (4.60) and (4.61) for those nonnormal stochastic basic variables. For normal stochastic basic variables, $\mu_{kN,(r)} = \mu_k$ and $\sigma_{kN,(r)} = \sigma_k$.

Step 3: Compute $W(\boldsymbol{x}_{(r)})$ and the corresponding sensitivity coefficient vector $\boldsymbol{s}_{x,(r)}$.

Step 4: Revise solution point $\boldsymbol{x}_{(r+1)}$ according to Eq. (4.52) with the mean and standard deviations of nonnormal stochastic basic variables replaced by their normal equivalents, that is,

$$\boldsymbol{x}_{(r+1)} = \boldsymbol{\mu}_{N,(r)} + \boldsymbol{D}_{N,(r)}\boldsymbol{s}_{x,(r)} \frac{(\boldsymbol{x}_{(r)} - \boldsymbol{\mu}_{N,(r)})^t \boldsymbol{s}_{x,(r)} - W(\boldsymbol{x}_{(r)})}{\boldsymbol{s}^t_{x,(r)}\boldsymbol{D}_{N,(r)}\boldsymbol{s}_{x,(r)}} \tag{4.62}$$

Step 5: Check if $\boldsymbol{x}_{(r)}$ and $\boldsymbol{x}_{(r+1)}$ are sufficiently close. If yes, compute the reliability index β_{AFOSM} according to Eq. (4.47) and the corresponding reliability $p_s = \Phi(\beta_{\mathrm{AFOSM}})$; then, go to step 5. Otherwise, update the solution point by letting $\boldsymbol{x}_{(r)} = \boldsymbol{x}_{(r+1)}$ and return to step 2.

Step 6: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables according to Eqs. (4.48), (4.49), and (4.50) with $\boldsymbol{D}_x$ replaced by $\boldsymbol{D}_{xN}$ at the design point $\boldsymbol{x}_*$.

As for the Ang-Tang AFOSM algorithm, the iterative algorithms described previously can be modified as follows (also see Fig. 4.10):

Step 1: Select an initial point $\boldsymbol{x}_{(r)}$ in the parameter space.

Step 2: Compute the mean and standard deviation of the normal equivalent using Eqs. (4.60) and (4.61) for those nonnormal stochastic basic variables. For normal stochastic basic variables, $\mu_{kN,(r)} = \mu_k$ and $\sigma_{kN,(r)} = \sigma_k$.

Step 3: At the selected point $\boldsymbol{x}_{(r)}$, compute the mean and variance of the performance function $W(\boldsymbol{x}_{(r)})$ according to Eqs. (4.56) and (4.44), respectively.

Step 4: Compute the corresponding reliability index $\beta_{(r)}$ according to Eq. (4.8).

Step 5: Compute the values of the normal equivalent directional derivative $\alpha_{kN,(r)}$, for all $k = 1, 2, \ldots, K$, according to Eq. (4.46), in that the standard

deviations of nonnormal stochastic basic variables σ_k's are replaced by the corresponding $\sigma_{kN,(r)}$'s.

Step 6: Using $\beta_{(r)}$ and $\alpha_{kN,(r)}$ obtained from steps 3 and 5, revise the location of expansion point $x_{(r+1)}$ according to

$$x_{k,(r+1)} = \mu_{kN,(r)} - \alpha_{kN,(r)}\beta_{(r)}\sigma_{kN,(r)} \qquad k = 1, 2, \ldots, K \qquad (4.63)$$

Step 7: Check if the revised expansion point $x_{(r+1)}$ differs significantly from the previous trial expansion point $x_{(r)}$. If yes, use the revised expansion point as the new trial point by letting $x_{(r)} = x_{(r+1)}$, and go to step 2 for another iteration. Otherwise, the iteration is considered complete, and the latest reliability index $\beta_{(r)}$ is used to compute the reliability $p_s = \Phi(\beta_{(r)})$.

Step 8: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables according to Eqs. (4.47), (4.48), and (4.49) with D_x replaced by D_{xN} at the design point x_*.

Example 4.10 (Independent, nonnormal) Refer to the data in Example 4.9 for the storm sewer reliability analysis problem. Assume that all three stochastic basic variables are independent random variables having different distributions. Manning's roughness n has a normal distribution; pipe diameter D, lognormal; and pipe slope S, Gumbel distribution. Compute the reliability that the sewer can convey an inflow discharge of 35 ft^3/s by the Hasofer-Lind algorithm.

Solution The initial solution is taken to be the means of the three stochastic basic variables, namely, $x_{(1)} = \mu_x = (\mu_n, \mu_D, \mu_S)^t = (0.015, 3.0, 0.005)^t$. Since the stochastic basic variables are not all normally distributed, the Rackwitz normal transformation is applied. For Manning's roughness, no transformation is required because it is a normal stochastic basic variable. Therefore, $\mu_{n,N,(1)} = \mu_n = 0.015$ and $\sigma_{n,N,(1)} = \sigma_n = 0.00075$.

For pipe diameter, which is a lognormal random variable, the variance and the mean of log-transformed pipe diameter can be computed, according to Eqs. (2.67a) and (2.67b), as

$$\sigma_{\ln D}^2 = \ln(1 + 0.02^2) = 0.0003999$$

$$\mu_{\ln D} = \ln(3.0)\frac{0.0003999}{2} = 1.09841$$

The standard normal variate z_D corresponding to $D = 3.0$ ft. is

$$z_D = [\ln(3) - \mu_{\ln D}]/\sigma_{\ln D} = 0.009999$$

Then, according to Eqs. (4.60) and (4.61), the standard deviation and the mean of normal equivalent at $D = 3.0$ ft. are, respectively,

$$\mu_{D,N,(1)} = 2.999 \qquad \sigma_{D,N,(1)} = 0.05999$$

For pipe slope, the two parameters in the Gumbel distribution, according to Eqs. (2.86a) and (2.86b), can be computed as

$$\beta = \frac{\sigma_S}{\sqrt{1.645}} = 0.0001949$$

$$\xi = \mu_S - 0.577\beta = 0.004888$$

The value of reduced variate $Y = (S - \xi)/\beta$ at $S = 0.005$ is $Y = 0.577$, and the corresponding value of CDF by Eq. (2.85a) is $F_{EV1}(Y = 0.577) = 0.5703$. According to Eq. (4.59), the standard normal quantile corresponding to the CDF of 0.5703 is $Z = 0.1772$. Based on the available information, the values of PDFs for the standard normal and Gumbel variables, at $S = 0.005$, can be computed as $\phi(Z = 0.1722) = 0.3927$ and $f_{EV1}(Y = 0.577) = 1643$. Then, by Eqs. (4.61) and (4.60), the normal equivalent standard deviation and the mean for the pipe slope, at $S = 0.005$, are

$$\mu_{S,N,(1)} = 0.004958 \qquad \sigma_{S,N,(1)} = 0.000239$$

At $\boldsymbol{x}_{(1)} = (0.015, 3.0, 0.005)^t$, the normal equivalent mean vector for the three stochastic basic variables is

$$\boldsymbol{\mu}_{N,(1)} = (\mu_{n,N,(1)}, \mu_{D,N,(1)}, \mu_{S,N,(1)})^t = (0.015, 2.999, 0.004958)^t$$

and the covariance matrix is

$$\boldsymbol{D}_{N,(1)} = \begin{bmatrix} \sigma_{n,N}^2 & 0 & 0 \\ 0 & \sigma_{D,N}^2 & 0 \\ 0 & 0 & \sigma_{S,N}^2 \end{bmatrix} = \begin{bmatrix} 0.00075^2 & 0 & 0 \\ 0 & 0.0599^2 & 0 \\ 0 & 0 & 0.000239^2 \end{bmatrix}$$

At $\boldsymbol{x}_{(1)}$, the sensitivity vector $\boldsymbol{s}_{x,(1)}$ is

$$\boldsymbol{s}_{x,(1)} = (\partial W/\partial n, \partial W/\partial D, \partial W/\partial S)^t = (-2734, 36.50, 4101)^t$$

and the value of the performance function $W(n, D, S) = 6.010$, is not equal to zero. This implies that the solution point $\boldsymbol{x}_{(1)}$ does not lie on the limit-state surface. Applying Eq. (4.62) using normal equivalent means $\boldsymbol{\mu}_N$ and variances $\boldsymbol{D}_{xN}$ and the new solution $\boldsymbol{x}_{(2)}$ can be obtained as $\boldsymbol{x}_{(2)} = (0.01590, 2.923, 0.004821)^t$. Then one checks the difference between the two consecutive solutions as

$$\delta = |\boldsymbol{x}_{(1)} - \boldsymbol{x}_{(2)}| = [(0.0159 - 0.015)^2 + (2.923 - 3.0)^2 + (0.004821 - 0.005)^2]^{0.5}$$

$$= 0.07729$$

which is considered large, and therefore, the iteration continues. The following table lists the solution point $\boldsymbol{x}_{(r)}$, its corresponding sensitivity vector $\boldsymbol{s}_{x,(r)}$, and the vector of directional derivatives $\boldsymbol{\alpha}_{N,(r)}$ in each iteration. The iteration stops when the difference between the two consecutive solutions is less than 0.001 and the value of the performance function is less than 0.001.

Iteration	Var.	$x_{(r)}$	$\mu_{N,(r)}$	$\sigma_{N,(r)}$	$s_{x,(r)}$	$\alpha_{N,(r)}$	$x_{(r+1)}$
$r = 1$	n	0.1500×10^{-01}	0.1500×10^{-01}	0.7500×10^{-03}	$-0.2734 \times 10^{+04}$	$-0.6497 \times 10^{+00}$	0.1590×10^{-01}
	D	$0.3000 \times 10^{+01}$	$0.2999 \times 10^{+01}$	0.5999×10^{-01}	$0.3650 \times 10^{+02}$	$0.6938 \times 10^{+00}$	$0.2923 \times 10^{+01}$
	S	0.5000×10^{-02}	0.4958×10^{-02}	0.2390×10^{-03}	$0.4101 \times 10^{+04}$	$0.3106 \times 10^{+00}$	0.4821×10^{-02}
			$\delta = 0.7857 \times 10^{-01}$	$W = 0.6010 \times 10^{+01}$	$\beta = 0.0000 \times 10^{+00}$		
$r = 2$	n	0.1590×10^{-01}	0.1500×10^{-01}	0.7500×10^{-03}	$-0.2229e^{+04}$	$-0.6410e^{+00}$	0.1598×10^{-01}
	D	$0.2923 \times 10^{+01}$	$0.2998 \times 10^{+01}$	0.5845×10^{-01}	$0.3237 \times 10^{+02}$	$0.7255 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4821×10^{-02}	0.4944×10^{-02}	0.1778×10^{-03}	$0.3675 \times 10^{+04}$	$0.2505 \times 10^{+00}$	0.4853×10^{-02}
			$\delta = 0.1113 \times 10^{-01}$	$W = 0.4371 \times 10^{+00}$	$\beta = 0.1894 \times 10^{+01}$		
$r = 3$	n	0.1598×10^{-01}	0.1500×10^{-01}	0.7500×10^{-03}	$-0.2190 \times 10^{+04}$	$-0.6369 \times 10^{+00}$	0.1598×10^{-01}
	D	$0.2912e^{+01}$	$0.2998e^{+01}$	0.5823×10^{-01}	$0.3210 \times 10^{+02}$	$0.7247 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4853×10^{-02}	0.4950×10^{-02}	0.1880×10^{-03}	$0.3607 \times 10^{+04}$	$0.2630e^{+00}$	0.4849×10^{-02}
			$\delta = 0.1942 \times 10^{-04}$	$W = 0.2147 \times 10^{-02}$	$\beta = 0.2049 \times 10^{+01}$		
$r = 4$	n	0.1598×10^{-01}	0.1500×10^{-01}	0.7500×10^{-03}	$-0.2190 \times 10^{+04}$	$-0.6373 \times 10^{+01}$	0.1598×10^{-01}
	D	$0.2912 \times 10^{+01}$	$0.2998 \times 10^{+01}$	0.5823×10^{-01}	$0.3210 \times 10^{+02}$	$0.7249 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4849×10^{-02}	0.4949×10^{-02}	0.1867×10^{-03}	$0.3609 \times 10^{+04}$	$0.2614 \times 10^{+00}$	0.4849×10^{-02}
			$\delta = 0.2553 \times 10^{-04}$	$W = 0.3894 \times 10^{-05}$	$\beta = 0.2050 \times 10^{+01}$		

After four iterations, the solution converges to the design point $\boldsymbol{x}_* = (n_*, D_*, S_*)^t = (0.01598, 2.912, 0.004849)^t$. At the design point $\boldsymbol{x}_*$, the mean and standard deviation of the performance function W can be estimated by Eqs. (4.42) and (4.43), respectively, as

$$\mu_{w^*} = 5.285 \qquad \text{and} \qquad \sigma_{w^*} = 2.578$$

The reliability index then can be computed as $\beta_* = \mu_{w^*}/\sigma_{w^*} = 2.050$, and the corresponding reliability and failure probability can be computed, respectively, as

$$p_s = \Phi(\beta_*) = 0.9798 \qquad p_f = 1 - p_s = 0.02019$$

Finally, at the design point $\boldsymbol{x}_*$, the sensitive of the reliability index and reliability with respect to each of the three stochastic basic variables can be computed by Eqs. (4.49) and (4.50). The results are shown in columns (4) to (7) in the following table:

Variable (1)	$\boldsymbol{x}$ (2)	$\alpha_{N,*}$ (3)	$\partial\beta/\partial z$ (4)	$\partial p_s/\partial z$ (5)	$\partial\beta/\partial x$ (6)	$\partial p_s/\partial x$ (7)	$x\partial\beta/\beta\partial x$ (8)	$x\partial p_s/p_s\partial x$ (9)
n	0.01594	−0.6372	0.6372	0.03110	849.60	41.46	6.623	0.6762
D	2.912	0.7249	−0.7249	−0.03538	−12.45	−0.61	−17.680	−1.8060
S	0.00483	0.2617	−0.2617	−0.01277	−1400.00	−68.32	−3.312	−0.3381

The sensitivity analysis yields a similar indication about the relative importance of the stochastic basic variables, as in Example 4.9.

4.5.6 Treatment of correlated normal stochastic variables

When some of the stochastic basic variables involved in the performance function are correlated, transformation of correlated variables to uncorrelated ones is made. Consider that the stochastic basic variables in the performance function are multivariate normal random variables with the mean matrix $\boldsymbol{\mu}_x$ and the covariance matrix $\boldsymbol{C}_x$. Without losing generality, the original stochastic basic variables are standardized according to Eq. (4.30) as $\boldsymbol{X}' = \boldsymbol{D}_x^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}_x)$.

Therefore, the standardized stochastic basic variables $\boldsymbol{X}'$ have the mean 0 and covariance matrix equal to the correlation matrix $\boldsymbol{R}_x$. That is, $\boldsymbol{C}_{x'} = \boldsymbol{R}_x = [\rho_{jk}]$, with ρ_{jk} being the correlation coefficient between stochastic basic variables X_j and X_k.

To break the correlative relation among the stochastic basic variables, orthogonal transformation techniques can be applied (see Appendix 4C). As an example, through eigenvalue-eigenvector (or spectral) decomposition, a new vector of uncorrelated stochastic basic variables $\boldsymbol{U}$ can be obtained as

$$\boldsymbol{U} = \boldsymbol{V}^t \boldsymbol{X}' \tag{4.64}$$

in which $\boldsymbol{V}_x$ is the normalized eigenvector matrix of the correlation matrix $\boldsymbol{R}_x$ of the original random variables. The new random variables $\boldsymbol{U}$ have a mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{L}_x = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K)$, which is a diagonal matrix containing the eigenvalues of $\boldsymbol{R}_x$. Hence the standard deviation of each uncorrelated standardized stochastic basic variable U_k is the square root of the corresponding eigenvalue, that is, $\sqrt{\lambda_k}$. Further standardization of $\boldsymbol{U}$ leads to

$$\boldsymbol{Y} = \Lambda_x^{-1/2} \boldsymbol{U} \tag{4.65}$$

in which $\boldsymbol{Y}$ are uncorrelated random variables having a mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$ being an identity matrix.

Consider that the original stochastic basic variables are multivariate normal random variables. The orthogonal transformation by Eq. (4.64) is a linear transformation; the resulting transformed random variables $\boldsymbol{U}$ are individually normal but uncorrelated; that is, $\boldsymbol{U} \sim \text{N}(\boldsymbol{0}, \boldsymbol{L})$ and $\boldsymbol{Y} = \boldsymbol{Z}' \sim \text{N}(\boldsymbol{0}, \boldsymbol{I})$. Then the relationship between the original stochastic basic variables $\boldsymbol{X}$ and the uncorrelated standardized normal variables $\boldsymbol{Z}'$ can be written as

$$\boldsymbol{Z}' = \Lambda_x^{-1/2} \boldsymbol{V}_x^t \boldsymbol{D}_x^{-1/2} (\boldsymbol{X} - \mu_x) \tag{4.66a}$$

$$\boldsymbol{X} = \mu_x + \boldsymbol{D}_x^{1/2} \boldsymbol{V}_x \Lambda_x^{1/2} \boldsymbol{Z}' \tag{4.66b}$$

in which Λ_x and $\boldsymbol{V}_x$ are, respectively, the eigenvalue matrix and eigenvector matrix corresponding to the correlation matrix $\boldsymbol{R}_x$.

In the transformed domain as defined by $\boldsymbol{Z}'$, the directional derivatives of the performance function in $\boldsymbol{z}'$-space $\boldsymbol{\alpha}_{z'}$ can be computed, according to Eq. (4.37), as

$$\boldsymbol{\alpha}_{z'} = \frac{\nabla_{z'} \boldsymbol{W}'(\boldsymbol{z}')}{|\nabla_{z'} \boldsymbol{W}'(\boldsymbol{z}')|} \tag{4.67}$$

in which the vector of sensitivity coefficients in $\boldsymbol{Z}'$-space $\boldsymbol{s}_{z'} = \nabla_{z'} W'(\boldsymbol{z}')$ can be obtained from $\nabla_x W(\boldsymbol{x})$ using the chain rule of calculus, according to Eq. (4.66b), as

$$\boldsymbol{s}_{z'} = \nabla_{z'} W(\boldsymbol{z}') = \left(\frac{\partial X_k}{\partial Z_j} \right) \nabla_x W(\boldsymbol{x})$$

$$= \left(\boldsymbol{D}_x^{1/2} \boldsymbol{V}_x \Lambda_x^{1/2} \right)^t \nabla_x W(\boldsymbol{x}) = \Lambda_x^{1/2} \boldsymbol{D}_x^{1/2} \boldsymbol{V}_x^t \boldsymbol{s}_x \tag{4.68}$$

in which s_x is the vector of sensitivity coefficients of the performance function with respect to the original stochastic basic variables X.

After the design point is found, one also is interested in the sensitivity of the reliability index and failure probability with respect to changes in the involved stochastic basic variables. In the uncorrelated standardized normal Z'-space, the sensitivity of β and p_s with respect to Z' can be computed by Eqs. (4.49) and (4.50) with X' replaced by Z'. The sensitivity of β with respect to X in the original parameter space then can be obtained as

$$\nabla_x \beta = \left[\frac{\partial Z'_j}{\partial X_k}\right] \nabla_{z'}\beta = \left(\Lambda_x^{-1/2} V_x^t D_x^{-1/2}\right)^t \nabla_{z'}\beta = -D_x^{-1/2} V_x \Lambda_x^{-1/2} \alpha_{z'} \qquad (4.69)$$

from which the sensitivity for p_s can be computed by Eq. (4.50b). A flowchart using the Ang-Tang algorithm for problems involving correlated stochastic basic variables is shown in Fig. 4.12. Step-by-step procedures for the correlated normal case by the Hasofer-Lind and Ang-Tang algorithms are given as follows.

The Hasofer–Lind AFOSM algorithm for problems having correlated normal stochastic variables involves the following steps:

Step 1: Select an initial trial solution $x_{(r)}$.

Step 2: Compute $W(x_{(r)})$ and the corresponding sensitivity coefficient vector $s_{x,(r)}$.

Step 3: Revise solution point $x_{(r+1)}$ according to

$$x_{(r+1)} = \mu_x + C_x s_{x,(r)} \frac{(x_{(r)} - \mu_x)^t s_{x,(r)} - W(x_{(r)})}{s_{x,(r)}^t C_x s_{x,(r)}} \qquad (4.70)$$

Step 4: Check if $x_{(r)}$ and $x_{(r+1)}$ are sufficiently close. If yes, compute the reliability index $\beta_{(r)}$ according to

$$\beta_{\text{AFOSM}} = [(x_* - \mu_x)^t C_x^{-1}(x_* - \mu_x)]^{1/2} \qquad (4.71)$$

and the corresponding reliability $p_s = \Phi(\beta_{\text{AFOSM}})$; then, go to step 5. Otherwise, update the solution point by letting $x_{(r)} = x_{(r+1)}$ and return to step 2.

Step 5: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables at the design point x_* by Eqs. (4.49), (4.50), (4.69), and (4.58).

On the other hand, the Ang-Tang AFOSM algorithm for problems involving correlated, normal stochastic basic variables consists of following steps:

Step 1: Decompose the correlation matrix R_x to find its eigenvector matrix V_x and eigenvalues λ's, using appropriate techniques.

Step 2: Select an initial point $x_{(r)}$ in the original parameter space.

Step 3: At the selected point $x_{(r)}$ compute the mean and variance of the performance function $W(X)$ according to Eqs. (4.56) and (4.43), respectively.
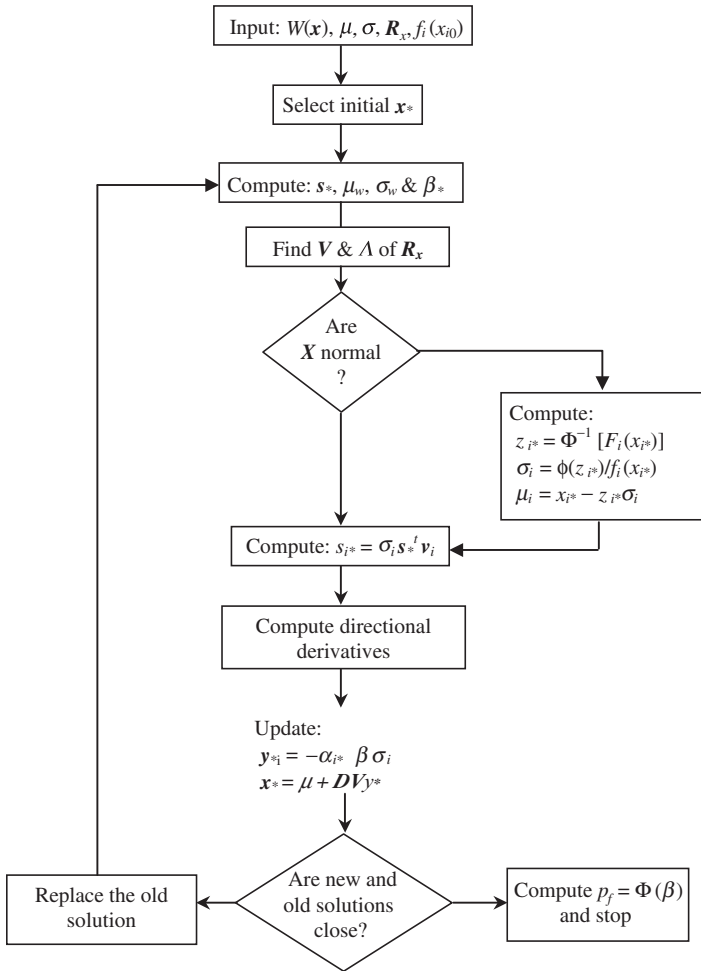
Figure 4.12 Flowchart for the Ang-Tang AFOSM reliability analysis involving correlated variables.

Step 4: Compute the corresponding reliability index $\beta_{(r)}$ according to Eq. (4.34).

Step 5: Compute sensitivity coefficients $\boldsymbol{s}_{z'}$ in the uncorrelated standard normal space according to Eq. (4.68) and the vector of directional derivatives $\boldsymbol{\alpha}_{z',(r)}$ according to Eq. (4.67).

Step 6: Using $\beta_{(r)}$ and $\boldsymbol{\alpha}_{z',(r)}$ obtained from steps 4 and 5, compute the location of expansion point $\boldsymbol{z}'_{(r+1)}$ in the uncorrelated standard normal space as

$$ z'_{k,(r+1)} = -\alpha_{k,(r)}\beta_{(r)} \qquad \text{for } k = 1, 2, \ldots, K \tag{4.72} $$

Step 7: Convert the obtained expansion point $\boldsymbol{z}'_{(r+1)}$ back to the original parameter space according to Eq. (4.66b).

Step 8: Check if the revised expansion point $\boldsymbol{x}_{(r+1)}$ differs significantly from the previous trial expansion point $\boldsymbol{x}_{(r)}$. If yes, use the revised expansion point as the trial point by letting $\boldsymbol{x}_{(r)} = \boldsymbol{x}_{(r+1)}$, and go to step 3 for another iteration. Otherwise, the iteration procedure is considered complete, and the latest reliability index $\beta_{(r)}$ is used to compute the reliability $p_s = \Phi(\beta_{(r)})$.

Step 9: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic basic variables at the design point $\boldsymbol{x}_*$ by Eqs. (4.49), (4.50), (4.69), and (4.68).

Example 4.11 (Correlated, normal)   Refer to the data in Example 4.9 for the storm sewer reliability analysis problem. Assume that Manning's roughness coefficient n and pipe diameter D are dependent normal random variables having a correlation coefficient of -0.75. Furthermore, the pipe slope S also is a normal random variable but is independent of Manning's roughness coefficient and pipe size. Compute the reliability that the sewer can convey an inflow discharge of 35 ft^3/s by the Hasofer-Lind algorithm.

Solution   The initial solution is taken to be the means of the three stochastic basic variables, namely, $\boldsymbol{x}_{(1)} = \boldsymbol{\mu}_x = (\mu_n, \mu_D, \mu_S)^t = (0.015, 3.0, 0.005)^t$. Since the stochastic basic variables are correlated normal random variables with a correlation matrix as follows:

$$\boldsymbol{R}_x = \begin{bmatrix} 1.0 & \rho_{n,D} & \rho_{n,S} \\ \rho_{n,D} & 1.0 & \rho_{D,S} \\ \rho_{n,S} & \rho_{D,S} & 1.0 \end{bmatrix} = \begin{bmatrix} 1.00 & -0.75 & 0.00 \\ -0.75 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

by the spectral decomposition, the eigenvalues matrix associated with the correlation matrix $\boldsymbol{R}_x$ is $\Lambda_x = \mathrm{diag}(1.75, 0.25, 1.00)$, and the corresponding eigenvector matrix $\boldsymbol{V}_x$ is

$$\boldsymbol{V}_x = \begin{bmatrix} 0.7071 & 0.7071 & 0.0000 \\ -0.7071 & 0.7071 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

At $\boldsymbol{x}_{(1)} = (0.015, 3.0, 0.005)^t$, the sensitivity vector for the performance function

$$W(n, D, S) = (Q_C - Q_L) = 0.463 n^1 D^{8/3} S^{1/2} - 35$$

is $$\boldsymbol{s}_{x,(1)} = (\partial W/\partial n, \partial W/\partial D, \partial W/\partial S)^t = (-2734, 36.50, 4101)^t$$

and the value of the performance function $W(\boldsymbol{x}_{(1)}) = 6.010$, is not equal to zero. This indicates that the solution point $\boldsymbol{x}_{(1)}$ does not lie on the limit-state surface. Applying Eq. (4.70), the new solution $\boldsymbol{x}_{(2)}$ can be obtained as $\boldsymbol{x}_{(2)} = (0.01569, 2.900, 0.004885)$. The difference between the two consecutive solutions is computed as

$$\delta = |\boldsymbol{x}_{(1)} - \boldsymbol{x}_{(2)}| = [(0.01569 - 0.015)^2 + (2.9 - 3.0)^2 + (0.004885 - 0.005)^2]^{0.5}$$
$$= 0.1002$$

which is considered large, and therefore, the iteration continues. The following table lists the solution point $x_{(r)}$, its corresponding sensitivity vector $s_{x,(r)}$, and the vector of directional derivatives $\alpha_{z',(r)}$, in each iteration. The iteration stops when the Euclidean distance between the two consecutive solution points is less than 0.001 and the value of the performance function is less than 0.001.

Iteration	Var.	$x_{(r)}$	$s_{(r)}$	$\alpha_{(r)}$	$x_{(r+1)}$
$r = 1$	n	0.1500×10^{-01}	$-0.2734 \times 10^{+04}$	$-0.9681 \times 10^{+00}$	0.1599×10^{-01}
	D	$0.3000 \times 10^{+01}$	$0.3650 \times 10^{+02}$	$0.2502 \times 10^{+00}$	$0.2920 \times 10^{+01}$
	S	0.5000×10^{-02}	$0.4101 \times 10^{+04}$	0.1203×10^{-01}	0.4908×10^{-02}
		$\delta = 0.8008 \times 10^{-01}$	$W = 0.6010 \times 10^{+01}$	$\beta = 0.000 \times 10^{+00}$	
$r = 2$	n	0.1599×10^{-01}	$-0.2217 \times 10^{+04}$	$-0.9656 \times 10^{+00}$	0.1607×10^{-01}
	D	$0.2920 \times 10^{+01}$	$0.3242 \times 10^{+02}$	$0.2583 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4908×10^{-02}	$0.3612 \times 10^{+04}$	0.2857×10^{-01}	0.4897×10^{-02}
		$\delta = 0.7453 \times 10 - 02$	$W = 0.4565 \times 10^{+00}$	$\beta = 0.1597 \times 10^{+01}$	
$r = 3$	n	0.1607×10^{-01}	$-0.2178 \times 10^{+04}$	$-0.9654 \times 10^{+00}$	0.1607×10^{-01}
	D	$0.2912 \times 10^{+01}$	$0.3209 \times 10^{+02}$	$0.2591 \times 10^{+00}$	$0.2912 \times 10^{+01}$
	S	0.4897×10^{-02}	$0.3574 \times 10^{+04}$	0.2991×10^{-01}	0.4896×10^{-02}
		$\delta = 0.7101 \times 10^{-04}$	$W = 0.2992 \times 10^{-02}$	$\beta = 0.1598 \times 10^{+01}$	

After four iterations, the solution converges to the design point $x_* = (n_*, D_*, S_*)^t = (0.01607, 2.912, 0.004896)^t$. At the design point x_*, $W = 0.5758 \times 10^{-07}$, and the mean and standard deviation of the performance function W can be estimated, by Eqs. (4.42) and (4.43), respectively, as

$$\mu_{w*} = 5.510 \quad \text{and} \quad \sigma_{w*} = 3.448$$

The reliability index then can be computed as $\beta_* = \mu_{w*}/\sigma_{w*} = 1.598$, and the corresponding reliability and failure probability can be computed, respectively, as

$$p_s = \Phi(\beta_*) = 0.9450 \quad p_f = 1 - p_s = 0.055$$

Finally, at the design point x_*, the sensitivity of the reliability index and reliability with respect to each of the three stochastic basic variables can be computed by Eqs. (4.49), (4.50), (4.56), and (4.57). The results are shown in the following table:

Variable (1)	x (2)	α_* (3)	$\partial\beta/\partial z$ (4)	$\partial p_s/\partial z$ (5)	$\partial\beta/\partial x$ (6)	$\partial p_s/\partial x$ (7)	$x\partial\beta/\beta\partial x$ (8)	$x\partial p_s/p_s\partial x$ (9)
n	0.01607	-0.9654	0.9654	0.1074	690.3	76.81	11.09	1.234
D	2.912	0.2591	-0.2591	-0.02883	-119.6	-13.31	-348.28	-38.76
S	0.004896	0.02991	-0.02991	-0.003328	$-1814.$	-201.9	-8.881	-0.9885

4.5.7 AFOSM reliability analysis for nonnormal correlated stochastic variables

For most practical engineering problems, parameters involved in load and resistance functions are correlated nonnormal random variables. Such distributional information has important implications for the results of reliability computations, especially on the tail part of the distribution for the performance

function. The procedures of the Rackwitz normal transformation and orthogonal decomposition described previously can be incorporated into AFOSM reliability analysis. The Ang-Tang algorithm, outlined below, first performs the orthogonal decomposition, followed by the normalization, for problems involving multivariate nonnormal stochastic variables (Fig. 4.12).

The Ang-Tang AFOSM algorithm for problems involving correlated nonnormal stochastic variables consists of the following steps:

Step 1: Decompose correlation matrix $\boldsymbol{R}_x$ to find its eigenvector matrix $\boldsymbol{V}_x$ and eigenvalues $\boldsymbol{\Lambda}_x$ using appropriate techniques.

Step 2: Select an initial point $\boldsymbol{x}_{(r)}$ in the original parameter space.

Step 3: At the selected point $\boldsymbol{x}_{(r)}$, compute the mean and variance of the performance function $W(\boldsymbol{X})$ according to Eqs. (4.56) and (4.43), respectively.

Step 4: Compute the corresponding reliability index $\beta_{(r)}$ according to Eq. (4.8).

Step 5: Compute the mean $\mu_{kN,(r)}$ and standard deviation $\sigma_{kN,(r)}$ of the normal equivalent using Eqs. (4.60) and (4.61) for the nonnormal stochastic variables.

Step 6: Compute the sensitivity coefficient vector with respect to the performance function $\boldsymbol{s}_{z',(r)}$ in the independent, standardized normal $\boldsymbol{z}'$-space, according to Eq. (4.68), with $\boldsymbol{D}_x$ replaced by $\boldsymbol{D}_{xN,(r)}$.

Step 7: Compute the vector of directional derivatives $\boldsymbol{\alpha}_{z',(r)}$ according to Eq. (4.67).

Step 8: Using $\beta_{(r)}$ and $\boldsymbol{\alpha}_{z',(r)}$ obtained from steps 4 and 7, compute the location of solution point $\boldsymbol{z}'_{(r+1)}$ in the transformed domain according to Eq. (4.70).

Step 9: Convert the obtained expansion point $\boldsymbol{z}'_{(r+1)}$ back to the original parameter space as

$$\boldsymbol{x}_{(r+1)} = \boldsymbol{\mu}_{x,N,(r)} + \boldsymbol{D}_{x,N,(r)}^{1/2} \boldsymbol{V}_x \boldsymbol{\Lambda}_x^{1/2} \boldsymbol{z}'_{(r+1)} \tag{4.73}$$

in which $\boldsymbol{\mu}_{x,N,(r)}$ is the vector of means of normal equivalent at solution point $\boldsymbol{x}_{(r)}$, and $\boldsymbol{D}_{x,N,(r)}$ is the diagonal matrix of normal equivalent variances.

Step 10: Check if the revised expansion point $\boldsymbol{x}_{(r+1)}$ differs significantly from the previous trial expansion point $\boldsymbol{x}_{(r)}$. If yes, use the revised expansion point as the trial point by letting $\boldsymbol{x}_{(r)} = \boldsymbol{x}_{(r+1)}$, and go to step 3 for another iteration. Otherwise, the iteration is considered complete, and the latest reliability index $\beta_{(r)}$ is used in Eq. (4.10) to compute the reliability p_s.

Step 11: Compute the sensitivity of the reliability index and reliability with respect to changes in stochastic variables according to Eqs. (4.48), (4.49), (4.51), (4.69), and (4.58), with $\boldsymbol{D}_x$ replaced by $\boldsymbol{D}_{x,N}$ at the design point $\boldsymbol{x}_*$.

One drawback of the Ang-Tang algorithm is the potential inconsistency between the orthogonally transformed variables $\boldsymbol{U}$ and the normal-transformed space in computing the directional derivatives in steps 6 and 7. This is so because the eigenvalues and eigenvectors associated with $\boldsymbol{R}_x$ will not be identical to those in the normal-transformed variables. To correct this inconsistency, Der Kiureghian and Liu (1985), and Liu and Der Kiureghian(1986) developed a

normal transformation that preserves the marginal probability contents and the correlation structure of the multivariate nonnormal random variables.

Suppose that the marginal PDFs of the two stochastic variables X_j and X_k are known to be $f_j(x_j)$ and $f_k(x_k)$, respectively, and their correlation coefficient is ρ_{jk}. For each individual random variable, a standard normal random variable that satisfies Eq. (4.59) is

$$\Phi(Z_j) = F_j(X_j) \qquad \Phi(Z_k) = F_k(X_k) \tag{4.74}$$

By definition, the correlation coefficient between the two stochastic variables X_j and X_k satisfies

$$\begin{aligned}
\rho_{jk} &= E\left[\left(\frac{X_j - \mu_j}{\sigma_j}\right)\left(\frac{X_k - \mu_k}{\sigma_k}\right)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_j - \mu_j}{\sigma_j}\right)\left(\frac{x_k - \mu_k}{\sigma_k}\right)f_{j,k}(x_j, x_k)\, dx_j\, dx_k
\end{aligned} \tag{4.75}$$

where μ_k and σ_k are, respectively, the mean and standard deviation of X_k. By the transformation of variable technique, the joint PDF $f_{j,k}(x_j, x_k)$ in Eq. (4.75) can be expressed in terms of a bivariate standard normal PDF as

$$f_{j,k}(x_j, x_k) = \phi\left(z_j, z_k | \rho_{jk}^*\right)\begin{vmatrix} \dfrac{\partial z_j}{\partial x_j} & \dfrac{\partial z_j}{\partial x_k} \\[2mm] \dfrac{\partial z_k}{\partial x_j} & \dfrac{\partial z_k}{\partial x_k} \end{vmatrix}$$

where $\phi(z_j, z_k \mid \rho_{jk}^*)$ is the bivariate standard normal PDF for Z_j and Z_k having zero means, unit standard deviations, and correlation coefficient ρ_{jk}^*, and the elements in *Jacobian matrix* can be evaluated as

$$\frac{\partial z_k}{\partial x_k} = \frac{\partial\,\Phi^{-1}[F_k(x_k)]}{\partial x_k} = \frac{f_k(x_k)}{\phi(z_k)}$$

$$\frac{\partial z_k}{\partial x_j} = 0 \qquad \text{for } j \neq k$$

Then the joint PDF of X_j and X_k can be simplified as

$$f_{j,k}(x_j, x_k) = \phi\left(z_j, z_k | \rho_{jk}^*\right)\frac{f_j(x_j)}{\phi(z_j)}\frac{f_k(x_k)}{\phi(z_k)} \tag{4.76}$$

Substituting Eq. (4.76) into Eq. (4.75) results in the *Nataf bivariate distribution model* (Nataf, 1962):

$$\rho_{jk} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_j - \mu_j}{\sigma_j}\right)\left(\frac{x_k - \mu_k}{\sigma_k}\right)\phi_{jk}(z_j, z_k | \rho_{jk}^*)\, dz_j\, dz_k \tag{4.77}$$

in which $x_k = F_k^{-1}[\Phi(z_k)]$.

Two conditions are inherently considered in the bivariate distribution model of Eq. (4.77):

1. According to Eq. (4.74), the normal transformation satisfies,

$$Z_k = \Phi^{-1}[F_k(X_k)] \qquad \text{for } k = 1, 2, \ldots, K \qquad (4.78)$$

   This condition preserves the probability content in both the original and the standard normal spaces.

2. The value of the correlation coefficient in the normal space lies between -1 and $+1$.

For a pair of nonnormal stochastic variables X_j and X_k with known means μ_j and μ_k, standard deviations σ_j and σ_k, and correlation coefficient ρ_{jk}, Eq. (4.77) can be applied to solve for ρ_{jk}^*. To avoid the required computation for solving ρ_{jk}^* in Eq. (4.74), Der Kiureghian and Liu (1985) developed a set of semiempirical formulas as

$$\rho_{jk}^* = T_{jk}\rho_{jk} \qquad (4.79)$$

in which T_{jk} is a transformation factor depending on the marginal distributions and correlation of the two random variables considered. In case both the random variables under consideration are normal, the transformation factor T_{jk} has a value of 1. Given the marginal distributions and correlation for a pair of random variables, the formulas of Der Kiureghian and Liu (1985) compute the corresponding transformation factor T_{jk} to obtain the equivalent correlation ρ_{jk}^* as if the two random variables were bivariate normal random variables. After all pairs of stochastic variables are treated, the correlation matrix in the correlated normal space $\boldsymbol{R}_z$ is obtained.

Ten different marginal distributions commonly used in reliability computations were considered by Der Kiureghian and Liu (1985) and are tabulated in Table 4.4. For each combination of two distributions, there is a corresponding formula. Therefore, a total of 54 formulas for 10 different distributions were developed, and they are divided into five categories, as shown in Fig. 4.13. The complete forms of these formulas are given in Table 4.5. Owing to the semiempirical nature of the equations in Table 4.5, it is a slight possibility that the resulting ρ_{jk}^* may violate its valid range when ρ_{jk} is close to -1 or $+1$.

Based on the normal transformation of Der Kiureghian and Liu, the AFOSM reliability analysis for problems involving multivariate nonnormal random variables can be conducted as follows:

Step 1: Apply Eq. (4.77) or Table 4.5 to construct the correlation matrix $\boldsymbol{R}_z$ for the equivalent random variables $\boldsymbol{Z}$ in the standardized normal space.

Step 2: Decompose correlation matrix $\boldsymbol{R}_z$ to find its eigenvector matrix V_z and eigenvalues λ_z's using appropriate orthogonal decomposition techniques. Therefore, $\boldsymbol{Z}' = \Lambda_z^{-1/2} \boldsymbol{V}_z^t \boldsymbol{Z}$ is a vector of independent standard normal random variables.

TABLE 4.4   Definitions of Distributions Used in Fig. 4.13 and Table 4.5

Distributions	PDF	Moments and parameters relations
Normal	$f_x(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2}$ for $-\infty < x < \infty$	—
Uniform	$f_x(x) = \frac{1}{b-a}$ for $a \le x \le b$	$\mu_x = \frac{a+b}{2}$ $\sigma_x^2 = \frac{(b-a)^2}{12}$
Shifted Exponential	$f_x(x) = \beta e^{-\beta(x-x_0)}$ for $x \ge x_0$	$\mu_x = \frac{1}{\beta} + x_0$ $\sigma_x^2 = \frac{1}{\beta^2}$
Shifted Rayleigh	$f_x(x) = \frac{(x-x_0)e^{-(x-x_0)^2/2\alpha^2}}{\alpha^2}$ for $x \ge x_0$	$\mu_x = 1.253\alpha + x_0$ $\sigma_x = 0.655136\alpha$
Type I, max	$f_x(x) = \frac{1}{\beta} e^{-\frac{(x-\xi)}{\beta} - e^{-\frac{(x-\xi)}{\beta}}}$ for $-\infty < x < \infty$	$\mu_x = \xi + 0.5772\beta$ $\sigma_x = \frac{\pi\beta}{\sqrt{6}}$ $\gamma_x = 1.29857$
Type I, min	$f_x(x) = \frac{1}{\beta} e^{(x-\xi)/\beta - e^{\frac{(x-\xi)}{\beta}}}$ for $-\infty < x < \infty$	$\mu_x = \xi - 0.5772\beta\beta$ $\sigma_x = \frac{\pi\beta}{\sqrt{6}}$ $\gamma_x = -1.29857$
Lognormal	$f_x(x) = \frac{1}{\sqrt{2\pi}\,x\sigma_{\ln x}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu_{\ln x}}{\sigma_{\ln x}}\right)^2}$ for $x \ge 0$	$\mu_{\ln x} = \ln(\mu_x) - \frac{1}{2}\sigma_{\ln x}^2$ $\sigma_{\ln x}^2 = \ln\left[1 + \left(\frac{\sigma_x}{\mu_x}\right)^2\right]$
Gamma	$f_x(x) = \frac{\beta^\alpha (x-\xi)^{\alpha-1} e^{-\beta(x-\xi)}}{\Gamma(\alpha)}$ for $x \ge \xi$	$\mu_x = \frac{\alpha}{\beta} + \xi$ $\sigma_x^2 = \frac{\alpha}{\beta^2}$
Type II, Largest	$f_x(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x}\right)^\alpha}$ for $x \ge 0$	$\mu_x = \beta\Gamma\left(1 - \frac{1}{\alpha}\right)$ $\sigma_x^2 = \beta^2\left[\Gamma\left(1 - \frac{2}{\alpha}\right) - \Gamma^2\left(1 - \frac{1}{\alpha}\right)\right]$
Type III, smallest (Weibull)	$f_x(x) = \frac{\alpha}{\beta}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\xi}{\beta}\right)^\alpha}$ for $x \ge \xi$	$\mu_x = \xi + \beta\Gamma\left(1 + \frac{1}{\alpha}\right)$ $\sigma_x^2 = \beta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$

Note:

N = Normal	T1S = Type 1 smallest
U = Uniform	L = Lognormal
E = Shifted exponential	G = Gamma
T1L = Type 1 largest	T2L = Type 2 largest
ρ_{jk} = Correlation coefficient	T3S = Type 3 smallest

Figure 4.13 Categories of the normal transformation factor T_{jk}. (*After Der Kiureghian and Liu, 1985*).

TABLE 4.5 Semiempirical Normal Transformation Formulas

(a) Category 1 of the transformation factor T_{jk} in Fig. 4.13

		U	E	R	T1L	T1S
N	$T_{jk} = $ constant	1.023	1.107	1.014	1.031	1.031
	Max. error	0.0%	0.0%	0.0%	0.0%	0.0%

NOTE: Distribution indices are N = normal; U = uniform; E = shifted exponential;
R = shifted Rayleigh; T1L = type 1, largest value; T1S = type 1, smallest value.

(b) Category 2 of the transformation factor T_{jk} in Fig. 4.13

		L	G	T2L	T3S
N	$T_{jk} = f(\Omega_k)$	$\dfrac{\Omega_k}{\sqrt{\ln\left(1+\Omega_k^2\right)}}$	$1.001 - 1.007\Omega_k + 0.118\Omega_k^2$	$1.030 + 0.238\Omega_k + 0.364\Omega_k^2$	$1.031 - 0.195\Omega_k + 0.328\Omega_k^2$
	Max. error	Exact	0.0%	0.1%	0.1%

NOTE: Ω_k is the coefficient of variation of the jth variable; distribution indices are N = normal; L = lognormal; G = gamma; T2L = type 2, largest value; T3S = type 3, smallest value.
SOURCE: After Der Kiureghian and Liu (1985).

(*Continued*)

TABLE 4.5   Semiempirical Normal Transformation Formulas (*Continued*)

(c) Category 3 of the transformation factor T_{jk} in Fig. 4.11

		U	E	R	T1L	T1S
U	$T_{jk} = f(\rho_{jk})$	$1.047 - 0.047\rho_{jk}^2$	$1.133 + 0.029\rho_{jk}^2$	$1.038 - 0.008\rho_{jk}^2$	$1.055 + 0.015\rho_{jk}^2$	$1.055 + 0.015\rho_{jk}^2$
	Max. error	0.0%	0.0%	0.0%	0.0%	0.0%
E	$T_{jk} = f(\rho_{jk})$		$1.229 - 0.367\rho_{jk}$ $+ 0.153\rho_{jk}^2$	$1.123 - 0.100\rho_{jk}$ $+ 0.021\rho_{jk}^2$	$1.142 - 0.154\rho_{jk}$ $+ 0.031\rho_{jk}^2$	$1.142 + 0.154\rho_{jk}$ $+ 0.031\rho_{jk}^2$
	Max. error		1.5%	0.1%	0.2%	0.2%
R	$T_{jk} = f(\rho_{jk})$			$1.028 - 0.029\rho_{jk}$	$1.046 - 0.045\rho_{jk}$ $+ 0.006\rho_{jk}^2$	$1.046 + 0.045\rho_{jk}$ $+ 0.006\rho_{jk}^2$
	Max. error			0.0%	0.0%	0.0%
T1L	$T_{jk} = f(\rho_{jk})$				$1.064 - 0.069\rho_{jk}$ $+ 0.005\rho_{jk}^2$	$1.064 + 0.069\rho_{jk}$ $+ 0.005\rho_{jk}^2$
	Max. error				0.0%	0.0%
T1S	$T_{jk} = f(\rho_{jk})$					$1.064 - 0.069\rho_{jk}$ $+ 0.005\rho_{jk}^2$
	Max. error					0.0%

NOTE: ρ_{jk} is the correlation coefficient between the jth variable and the kth variable; distribution indices are U = uniform; E = shifted exponential; R = shifted Rayleigh; T1L = type 1, largest value; T1S = type 1, smallest value.

(d) Category 4 of the transformation factor T_{jk} in Fig. 4.13

		L	G	T2L	T3S
U	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$1.019 + 0.014\Omega_k + 0.010\rho_{jk}^2$ $+ 0.249\Omega_k^2$	$1.023 - 0.007\Omega_k + 0.002\rho_{jk}^2$ $+ 0.127\Omega_k^2$	$1.033 + 0.305\Omega_k + 0.074\rho_{jk}^2$ $+ 0.405\Omega_k^2$	$1.061 - 0.237\Omega_k - 0.005\rho_{jk}^2$ $+ 0.379\Omega_k^2$
	Max. error	0.7%	0.1%	2.1%	0.5%
E	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$1.098 + 0.003\rho_{jk} + 0.019\Omega_k$ $+ 0.025\rho_{jk}^2 + 0.303\Omega_k^2$ $- 0.437\rho_{jk}\Omega_k$	$1.104 + 0.003\rho_{jk} - 0.008\Omega_k$ $+ 0.014\rho_{jk}^2 + 0.173\Omega_k^2$ $- 0.296\rho_{jk}\Omega_k$	$1.109 - 0.152\rho_{jk} + 0.361\Omega_k$ $+ 0.130\rho_{jk}^2 + 0.455\Omega_k^2$ $- 0.728\rho_{jk}\Omega_k$	$1.147 + 0.145\rho_{jk} - 0.271\Omega_k$ $+ 0.010\rho_{jk}^2 + 0.459\Omega_k^2$ $- 0.467\rho_{jk}\Omega_k$
	Max. error	1.6%	0.9%	0.9%	0.4%
R	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$1.011 + 0.001\rho_{jk} + 0.014\Omega_k$ $+ 0.004\rho_{jk}^2 + 0.231\Omega_k^2$ $- 0.130\rho_{jk}\Omega_k$	$1.014 + 0.001\rho_{jk} - 0.007\Omega_k$ $+ 0.002\rho_{jk}^2 + 0.126\Omega_k^2$ $- 0.090\rho_{jk}\Omega_k$	$1.036 - 0.038\rho_{jk} + 0.266\Omega_k$ $+ 0.028\rho_{jk}^2 + 0.383\Omega_k^2$ $- 0.229\rho_{jk}\Omega_k$	$1.047 + 0.042\rho_{jk} - 0.212\Omega_k$ $+ 0.353\Omega_k^2$ $- 0.136\rho_{jk}\Omega_k$
	Max. error	0.4%	0.9%	1.2%	0.2%
T1L	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$1.029 + 0.001\rho_{jk} + 0.014\Omega_k$ $+ 0.004\rho_{jk}^2 + 0.233\Omega_k^2$ $- 0.197\rho_{jk}\Omega_k$	$1.031 + 0.001\rho_{jk} - 0.007\Omega_k$ $+ 0.003\rho_{jk}^2 + 0.131\Omega_k^2$ $- 0.132\rho_{jk}\Omega_k$	$1.056 - 0.060\rho_{jk} + 0.263\Omega_k$ $+ 0.020\rho_{jk}^2 + 0.383\Omega_k^2$ $- 0.332\rho_{jk}\Omega_k$	$1.064 + 0.065\rho_{jk} - 0.210\Omega_k$ $+ 0.003\rho_{jk}^2 + 0.356\Omega_k^2$ $- 0.211\rho_{jk}\Omega_k$
	Max. error	0.3%	0.3%	1.0%	0.2%
T1S	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$1.029 + 0.001\rho_{jk} + 0.014\Omega_k$ $+ 0.004\rho_{jk}^2 + 0.233\Omega_k^2$ $+ 0.197\rho_{jk}\Omega_k$	$1.031 - 0.001\rho_{jk} - 0.007\Omega_k$ $+ 0.003\rho_{jk}^2 + 0.131\Omega_k^2$ $+ 0.132\rho_{jk}\Omega_k$	$1.056 + 0.060\rho_{jk} + 0.263\Omega_k$ $+ 0.020\rho_{jk}^2 + 0.383\Omega_k^2$ $+ 0.332\rho_{jk}\Omega_k$	$1.064 - 0.065\rho_{jk} - 0.210\Omega_k$ $+ 0.003\rho_{jk}^2 + 0.356\Omega_k^2$ $+ 0.211\rho_{jk}\Omega_k$
	Max. error	0.3%	0.3%	1.0%	0.2%

NOTE: ρ_{jk} is the correlation coefficient between the jth variable and the kth variable; Ω_k is the coefficient of variation of the kth variable; distribution indices are U = uniform; E = shifted exponential; R = shifted Rayleigh; T1L = type 1, largest value; T1S = type 1, smallest value; L = lognormal; G = gamma; T2L = type 2, largest value; T3S = type 3 smallest value.

(Continued)

TABLE 4.5 Semiempirical Normal Transformation Formulas (*Continued*)
(e) Category 5 of the transformation factor T_{jk} in Fig. 4.13

		L	G	T2L	T3S
L	$T_{jk} = f(\rho_{jk}, \Omega_k)$	$\dfrac{\ln(1+\rho_{jk}\Omega_j\Omega_k)}{\rho_{jk}\sqrt{\ln\left(1+\Omega_j^2\right)\ln\left(1+\Omega_k^2\right)}}$	$1.001 + 0.033\rho_{jk} + 0.004\Omega_j$ $- 0.016\Omega_k + 0.002\rho_{jk}^2$ $+ 0.223\Omega_j^2 + 0.130\Omega_k^2$ $- 0.104\rho_{jk}\Omega_j + 0.029\Omega_j\Omega_k$ $- 0.119\rho_{jk}\Omega_k$	$1.026 + 0.082\rho_{jk} - 0.019\Omega_j$ $- 0.222\Omega_k + 0.018\rho_{jk}^2 + 0.288\Omega_j^2$ $+ 0.379\Omega_k^2 - 0.104\rho_{jk}\Omega_j$ $+ 0.126\Omega_j\Omega_k - 0.277\rho_{jk}\Omega_k$	$1.031 + 0.052\rho_{jk} + 0.011\Omega_j$ $- 0.21\Omega_k + 0.002\rho_{jk}^2 + 0.22\Omega_j^2$ $+ 0.35\Omega_k^2 + 0.005\rho\Omega_j$ $+ 0.009\Omega_j\Omega_k - 0.174\rho\Omega_k$
	Max. error	Exact	4.0%	4.3%	2.4%
G	$T_{jk} = f(\rho_{jk}, \Omega_j, \Omega_k)$		$1.002 + 0.022\rho_{jk}$ $- 0.012(\Omega_j + \Omega_k) + 0.001\rho_{jk}^2$ $+ 0.125(\Omega_j^2 + \Omega_k^2)$ $- 0.077\rho_{jk}(\Omega_j + \Omega_k)$ $+ 0.014\Omega_j\Omega_k$	$1.029 + 0.056\rho_{jk} - 0.030\Omega_j$ $+ 0.225\Omega_k + 0.012\rho_{jk}^2$ $+ 0.174\Omega_j^2 + 0.379\Omega_k^2$ $- 0.313\rho_{jk}\Omega_j + 0.075\Omega_j\Omega_k$ $- 0.182\rho_{jk}\Omega_k$	$1.032 + 0.034\rho_{jk} - 0.007\Omega_j$ $- 0.202\Omega_k + 0.121\Omega_j^2$ $+ 0.339\Omega_k^2 - 0.006\rho\Omega_j$ $+ 0.003\Omega_j\Omega_k - 0.111\rho\Omega_k$
	Max. error		4.0%	4.2%	4.0%
T2L	$T_{jk} = f(\rho_{jk}, \Omega_j, \Omega_k)$			$1.086 + 0.054\rho_{jk} + 0.104(\Omega_j + \Omega_k)$ $+ 0.055\rho_{jk}^2 + 0.662(\Omega_j^2 + \Omega_k^2)$ $- 0.570\rho_{jk}(\Omega_j + \Omega_k) + 0.203\Omega_j\Omega_k$ $- 0.020\rho_{jk}^3 \quad - 0.218(\Omega_j^3 + \Omega_k^3)$ $- 0.371\rho_{jk}(\Omega_j^2 + \Omega_k^2)$ $+ 0.257\rho_{jk}^2(\Omega_j + \Omega_k)$ $+ 0.141\Omega_j\Omega_k(\Omega_j + \Omega_k)$	$1.065 + 0.146\rho_{jk} + 0.241\Omega_j$ $- 0.259\Omega_k + 0.013\rho_{jk}^2 + 0.372\Omega_j^2$ $+ 0.435\Omega_k^2 + 0.005\rho\Omega_j$ $+ 0.034\Omega_j\Omega_k - 0.481\rho\Omega_k$
	Max. error			4.3%	3.8%
T3S	$T_{jk} = f(\rho_{jk}, \Omega_j, \Omega_k)$				$1.063 - 0.004\rho_{jk} - 0.200(\Omega_j + \Omega_k)$ $- 0.001\rho_{jk}^2 + 0.337(\Omega_j^2 + \Omega_k^2)$ $+ 0.007\rho(\Omega_j + \Omega_k) - 0.007\Omega_j\Omega_k$
	Max. error				2.62%

NOTE: ρ_{jk} is the correlation coefficient between the jth variable and the kth variable; Ω_j is the coefficient of variation of the jth variable; Ω_k is the coefficient of variation of the kth variable; distribution indices are L = lognormal; G = gamma; T2L = type 2, largest value; T3S = type 3, smallest value.

Step 3: Select an initial point $\boldsymbol{x}_{(r)}$ in the original parameter space $\boldsymbol{X}$, and compute the sensitivity vector for the performance function $\boldsymbol{s}_{x,(r)} = \boldsymbol{\nabla}_x W(\boldsymbol{x}_{(r)})$.

Step 4: At the selected point $\boldsymbol{x}_{(r)}$, compute the means $\boldsymbol{\mu}_{N,(r)} = (\mu_{1N}, \mu_{2N}, \ldots, \mu_{KN})^t$ and standard deviations $\boldsymbol{\sigma}_{N,(r)} = (\sigma_{1N}, \sigma_{2N}, \ldots, \sigma_{KN})^t$ of the normal equivalent using Eqs. (4.59) and (4.60) for the nonnormal stochastic variables. Compute the corresponding point $\boldsymbol{z}'_{(r)}$ in the independent standardized normal space as

$$\boldsymbol{z}'_{(r)} = \boldsymbol{\Lambda}_z^{1/2} \boldsymbol{V}_z^t \boldsymbol{D}_{x,N,(r)}^{1/2}(\boldsymbol{x}_{(r)} - \boldsymbol{\mu}_{x,N,(r)}) \tag{4.80}$$

in which $\boldsymbol{D}_{x,N,(r)} = \mathrm{diag}(\sigma_{1N}^2, \sigma_{2N}^2, \ldots, \sigma_{KN}^2)$, a diagonal matrix containing the variance of normal equivalent at the selected point $\boldsymbol{x}_{(r)}$. The corresponding reliability index can be computed as $\beta_{(r)} = \mathrm{sign}[W'(\boldsymbol{0})]|\boldsymbol{z}'_{(r)}|$.

Step 5: Compute the vector of sensitivity coefficients for the performance function in $\boldsymbol{Z}'$-space $\boldsymbol{s}_{z',(r)} = \boldsymbol{\nabla}_{z'} W(\boldsymbol{z}'_{(r)})$, by Eq. (4.68), with $\boldsymbol{D}_x$ replaced by $\boldsymbol{D}_{x,N,(r)}$, and $\boldsymbol{V}_x$ and $\boldsymbol{\Lambda}_x$ replaced by $\boldsymbol{V}_z$ and $\boldsymbol{\Lambda}_z$, respectively. Then the vector of directional derivatives in the independent standard normal space $\boldsymbol{\alpha}_{z',(r)}$ can be computed by Eq. (4.67).

Step 6: Apply Eq. (4.51) of the Hasofer-Lind algorithm or Eq. (4.70) of the Ang-Tang algorithm to obtain a new solution $\boldsymbol{z}'_{(r+1)}$.

Step 7: Convert the new solution $\boldsymbol{z}'_{(r+1)}$ back to the original parameter space by Eq. (4.66a), and check for convergence. If the new solution does not satisfy convergence criteria, go to step 3; otherwise, go to step 8.

Step 8: Compute the reliability, failure probability, and their sensitivity vectors with respect to change in stochastic variables.

Note that the previously described normal transformation of Der Kiureghian and Liu (1985) preserves only the marginal distributions and the second-order correlation structure of the correlated random variables, which are partial statistical features of the complete information represented by the joint distribution function. Regardless of its approximate nature, the normal transformation of Der Kiureghian and Liu, in most practical engineering problems, represents the best approach to treat the available statistical information about the correlated random variables. This is so because, in reality, the choices of multivariate distribution functions for correlated random variables are few as compared with univariate distribution functions. Furthermore, the derivation of a reasonable joint probability distribution for a mixture of correlated nonnormal random variables is difficult, if not impossible. When the joint PDF for the correlated nonnormal random variables is available, a practical normal transformation proposed by Rosenblatt (1952) can be viewed as the generalization of the normal transformation described in Sec. 4.5.5 for the case involving independent variables. Notice that the correlations among each pair of random variables are implicitly embedded in the joint PDF, and determination of correlation coefficients can be made according to Eqs. (2.47) and (2.48).

The Rosenblatt method transforms the correlated nonnormal random variables X to independent standard normal random variables Z' in a manner similar to Eq. (4.78) as

$$
\begin{aligned}
z'_1 &= \Phi^{-1}[F_1(x_1)] \\
z'_2 &= \Phi^{-1}[F_2(x_2|x_1)] \\
&\vdots \\
z'_k &= \Phi^{-1}[F_k(x_k|x_1, x_2, \ldots, x_{k-1})] \\
&\vdots \\
z'_K &= \Phi^{-1}[F_k(x_K|x_1, x_2, \ldots, x_{K-1})]
\end{aligned}
\tag{4.81}
$$

in which $F_k(x_k|x_1, x_2, \ldots, x_{k-1}) = P(X_k \leq x_k|x_1, x_2, \ldots, x_{k-1})$ is the conditional CDF for the random variable X_k conditional on $X_1 = x_1, X_2 = x_2, \ldots,$ $X_{k-1} = x_{k-1}$. Based on Eq. (2.17), the conditional PDF $f_k(x_k|x_1, x_2, \ldots, x_{k-1})$ for the random variable X_k can be obtained as

$$
f_k(x_k|x_1, x_2, \ldots, x_{k-1}) = \frac{f(x_1, x_2, \ldots, x_{k-1}, x_k)}{f(x_1, x_2, \ldots, x_{k-1})}
$$

with $f(x_1, x_2, \ldots, x_{k-1}, x_k)$ being the marginal PDF for $X_1, X_2, \ldots, X_{k-1}, X_k$; the conditional CDF $F_k(x_k|x_1, x_2, \ldots, x_{k-1})$ then can be computed by

$$
F_k(x_k|x_1, x_2, \ldots, x_{k-1}) = \frac{\int_{-\infty}^{x_k} f(x_1, x_2, \ldots, x_{k-1}, t)\,dt}{f(x_1, x_2, \ldots, x_{k-1})}
\tag{4.82}
$$

To incorporate the Rosenblatt normal transformation in the AFOSM algorithms described in Sec. 4.5.5, the marginal PDFs $f_k(x_k)$ and the conditional CDFs $F_k(x_k|x_1, x_2, \ldots, x_{k-1})$, for $k = 1, 2, \ldots, K$, first must be derived. Then Eq. (4.81) can be implemented in a straightforward manner in each iteration, within which the elements of the trial solution point $x_{(r)}$ are selected successively to compute the corresponding point in the equivalent independent standard normal space $z'_{(r)}$ and the means and variances by Eqs. (4.80) and (4.81), respectively. It should be pointed out that the order of selection of the stochastic basic variables in Eq. (4.81) can be arbitrary. Madsen et al. (1986, pp. 78–80) show that the order of selection may affect the calculated failure probability, and their numerical example does not show a significant difference in resulting failure probabilities.

4.5.8   Overall summary of AFOSM reliability method

Convergence criteria for locating the design point. The previously described Hasofer-Lind and Ang-Tang iterative algorithms to determine the design point indicate that the iterations may end when $x_{(r)}$ and $x_{(r+1)}$ are sufficiently close. The key question then becomes what constitutes sufficiently close. In the examples given previously in this section, the iterations were stopped when the

difference between the current and previous design point was less than 0.001. Whereas such a tight tolerance worked for the pipe-capacity examples in this book, it might not be appropriate for other cases, particularly for practical problems. Thus alternative convergence criteria often have been used.

In some cases, the solution has been considered to have converged when the values of $\beta_{(r)}$ and $\beta_{(r+1)}$ are sufficiently close. For example, Ang and Tang (1984, pp. 361–383) presented eight example applications of the AFOSM method to civil engineering systems, and the convergence criteria for differences in β ranged from 0.025 to 0.001. The Construction Industry Research and Information Association (CIRIA, 1977) developed an iterative approach similar to that of Ang and Tang (1984), only their convergence criterion was that the performance function should equal zero within some tolerance. The CIRIA procedure was applied in the uncertainty analysis of backwater computations using the HEC-2 water surface profiles model done by Singh and Melching (1993).

In order for iterative algorithms to locate the design point to achieve convergence, the performance function must be locally differentiable, and the original density functions of X_k must be continuous and monotonic, at least for $X_k \leq x_{k^*}$ (Yen et al., 1986). If the performance function is discontinuous, it must be treated as a series of continuous functions.

The search for the design point may become numerically more complex if the performance function has several local minima or if the original density functions of the X_k are discontinuous and bounded. It has been found that some of the following problems occasionally may result for the iteration algorithms to locate the design point (Yen et al., 1986):

1. The iteration may diverge or it may give different β values because of local minima in the performance function.

2. The iteration may converge very slowly when the probability of failure is very small, for example, $p_f < 10^{-4}$.

3. In the case of bounded random variables, the iteration may yield some x_{k^*} values outside the bounded range of the original density function. However, if the bounds are strictly enforced, the iterations may diverge.

Yen et al. (1986) recommended use of the generalized reduced gradient (GRG) optimization method proposed by Cheng et al. (1982) to determine the design point to reduce these numerical problems. However, the GRG-based method may not work well when complex computer models are needed to determine the system performance function.

Melching (1992) applied the AFOSM method using the Rackwitz iterative algorithm (Rackwitz and Fiessler, 1978), which is similar to the Ang-Tang algorithm, to determine the design point for estimation of the probability of flooding for 16 storms on an example watershed using two rainfall-runoff models. In this application, problems with performance function discontinuities, slow convergence for small values of p_f, and divergence in the estimated

β values were experienced for some of the cases. In the case of discontinuity in the performance function (resulting from the use of a simple initial loss-continuing loss rate abstraction scheme), in some cases the iterations went back and forth between one side of the discontinuity and the other, and convergence in the values of the x_ks could not be achieved. Generally, in such cases, the value of β had converged to the second decimal place, and thus a good approximation of β_* corresponding to the design point was obtained.

For extreme probability cases ($\beta > 2.5$), the iterations often diverged. The difference in β values for performance function values near zero typically was on the order of 0.2 to 0.4. The iteration of which the β value was smallest was selected as a reasonable estimate of the true β_* corresponding to the design point. In Melching (1992), the p_f values so approximated were on the order of 0.006 to 0.00004. Thus, from a practical viewpoint of whether or not a flood is likely, such approximations of β_* do not greatly change the estimated flood risk for the event in question. However, if various flood-mitigation alternatives were being compared in this way, one would have to be very careful that consistent results were obtained when comparing the alternatives.

A shortcoming of the afosm reliability index. As shown previously, use of the AFOSM reliability index removes the problem of lack of invariance associated with the MFOSM reliability index. This allows one to place different designs on the same common ground for comparing their relative reliabilities using β_{AFOSM}. A design with higher value of β_{AFOSM} would be associated with a higher reliability and lower failure probability. Referring to Fig. 4.14, in which failure surfaces of four different designs are depicted in the uncorrelated standardized parameter space, an erroneous conclusion would be made if one assesses the relative reliability on the basis of the reliability index. Note that in Fig. 4.14 the designs A, B, and C have identical values of the reliability index, but the size of their safe regions S_A, S_B, and S_C are not the same, and in fact, they satisfy $S_A \subset S_B \subset S_C$. The actual reliability relationship among the three designs should be $p_s(A) < p_s(B) < p_s(C)$, which is not reflected by the reliability index. One could observe that if the curvatures of different failure surfaces at the design point are similar, such as those with designs A and B, relative reliabilities between different designs could be indicated accurately by the value of reliability index. On the other hand, when the curvatures of failure surfaces are significantly different, such as those for designs C and D, β_{AFOSM} alone could not be used as the basis for comparison.

For this reason, Ditlevsen (1979) proposed a generalized reliability index $\beta_G = \Phi(\gamma)$, with γ being a reliability measure obtained from integrating a weight function over the safe region S, that is,

$$\gamma = \int_{\boldsymbol{x} \in s} \psi(\boldsymbol{x}) \, d\boldsymbol{x} \qquad (4.83)$$

in which $\psi(\boldsymbol{x})$ is the weight function, which is rotationally symmetric and positive (Ditlevsen, 1979). One such function that is mathematically tractable is
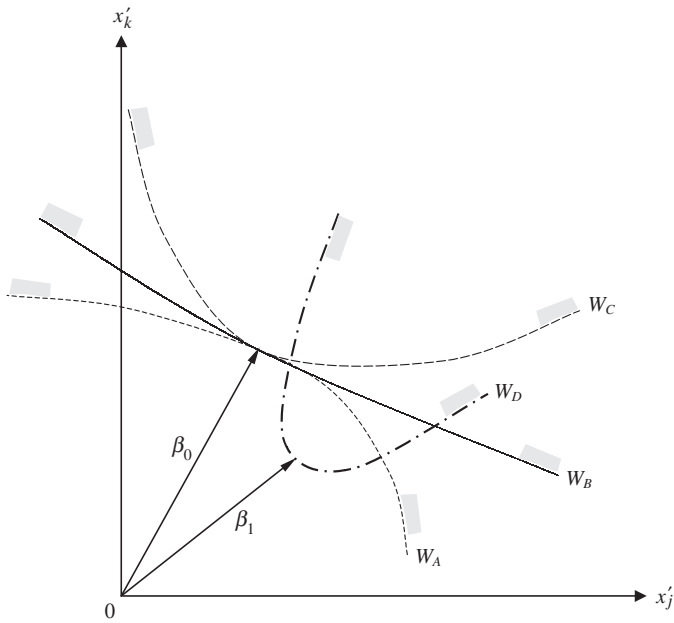
Figure 4.14   Nonunique reliability associated with an identical reliability index.

the K-dimensional standardized independent normal PDF. Although the generalized reliability index provides a more consistent and selective measure of reliability than β_{AFOSM} for a nonlinear failure surface, it is, however, more computationally difficult to obtain. From a practical viewpoint, most engineering applications result in the general reliability index whose value is close to β_{AFOSM}. Only in cases where the curvature of the failure surface at the design point is large and there are several design points on the failure surface would the two reliability indices deviate significantly.

4.6   Second-Order Reliability Methods

By the AFOSM reliability method, the design point on the failure surface is identified. This design point has the shortest distance to the mean point of the stochastic basic variables in the original space or to the origin of standardized normal parameter space. In the AFOSM method, the failure surface is locally approximated by a hyperplane tangent to the design point using the first-order terms of the Taylor series expansion. As shown in Fig. 4.14, second-order reliability methods (SORMs) can improve the accuracy of calculated reliability under a nonlinear limit-state function by which the failure surface is approximated locally at the design point by a quadratic surface. Literature on the SORMs can be found elsewhere (Fiessler et al., 1979; Shinozuka, 1983;

Breitung, 1984; Ditlevsen, 1984; Naess, 1987; Wen, 1987; Der Kiureghian et al., 1987; Der Kiureghian and De Stefano, 1991). Tvedt (1983) and Naess (1987) developed techniques to compute the bounds of the failure probability. Wen (1987), Der Kiureghian et al. (1987), and others demonstrated that the second-order methods yield an improved estimation of failure probability at the expense of an increased amount of computation. Applications of second-order reliability analysis to hydrosystem engineering problems are relatively few as compared with the first-order methods.

In the following presentations of the second-order reliability methods, it is assumed that the original stochastic variables $\boldsymbol{X}$ in the performance function $W(\boldsymbol{X})$ have been transformed to the independent standardized normal space by $\boldsymbol{Z}' = T(\boldsymbol{X})$, in which $\boldsymbol{Z}' = (Z_1', Z_2', \ldots, Z_K')^t$ is a column vector of independent standard normal random variables. Realizing that the first-order methods do not account for the curvature of the failure surface, the first-order failure probability could over- or underestimate the true p_f depending on the curvilinear nature of $W(\boldsymbol{Z}')$ at $\boldsymbol{z}_*'$. Referring to Fig. 4.15a, in which the failure surface is convex toward the safe region, the first-order method would overestimate the failure probability p_f, and, in the case of Fig. 4.15b, the opposite effect would result. When the failure region is a convex set, a bound of the failure probability is (Lind, 1977)

$$\Phi(-\beta_*) \leq p_f \leq 1 - F_{\chi_K^2}(\beta_*) \tag{4.84}$$

in which β_* is the reliability index corresponding to the design point $\boldsymbol{z}_*'$, and $F_{\chi_K^2}(\beta_*)$ is the value of the χ_K^2 CDF with K degrees of freedom. Note that the upper bound in Eq. (4.84) is based on the use of a hypersphere to approximate the failure surface at the design point and, consequently, is generally much more conservative than the lower bound. To improve the accuracy of the failure-probability estimation, a better quadratic approximation of the failure surface is needed.

4.6.1 Quadratic approximations of the performance function

At the design point $\boldsymbol{z}_*'$ in the independent standard normal space, the performance function can be approximated by a quadratic form as

$$W(\boldsymbol{Z}') \approx \boldsymbol{s}_{z_*'}^t (\boldsymbol{Z}' - \boldsymbol{z}_*') + \frac{1}{2}(\boldsymbol{Z}' - \boldsymbol{z}_*')^t \boldsymbol{G}_{z_*'}(\boldsymbol{Z}' - \boldsymbol{z}_*')$$

$$= \sum_{k=1}^{K} \left(\frac{\partial W(\boldsymbol{Z}')}{\partial Z_k'} \right)_{\boldsymbol{z}_*'} (Z_k' - z_{*,k}')$$

$$+ \frac{1}{2} \sum_{j=1}^{K} \sum_{k=1}^{K} \left[\frac{\partial^2 W(\boldsymbol{Z}')}{\partial Z_j' \partial Z_k'} \right]_{\boldsymbol{z}_*'} (Z_j' - z_{*,j}')(Z_k' - z_{*,k}') = 0 \tag{4.85}$$

(a)

(b)
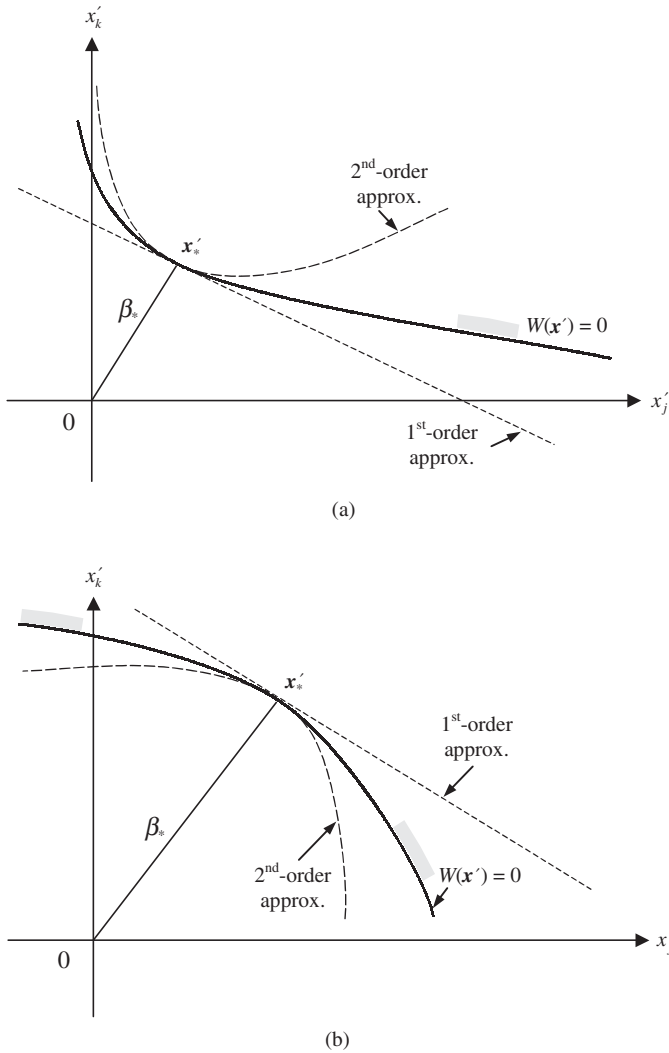
Figure 4.15 Schematic sketch of nonlinear performance functions: (*a*) convex performance function (positive curvature); (*b*) concave performance function (negative curvature).

in which $\boldsymbol{s}_{z'_*} = \nabla_{z'} W(\boldsymbol{z}'_*)$ and $\boldsymbol{G}_{z'_*} = \nabla^2_{z'} W(\boldsymbol{z}'_*)$ are, respectively, the gradient vector containing the sensitivity coefficients and the Hessian matrix of the performance function $W(\boldsymbol{Z}')$ evaluated at the design point $\boldsymbol{z}'_*$. The quadratic approximation by Eq. (4.85) involves cross-product of the random variables. To eliminate the cross-product interaction terms in the quadratic approximation, an orthogonal transform is accomplished by utilizing the symmetric square

nature of the Hessian matrix:

$$\boldsymbol{G}_{z'_*} = \Delta^2_{z'} W(\boldsymbol{z}'_*) = \left[\frac{\partial^2 W(\boldsymbol{z}')}{\partial z'_j \, \partial z'_k} \right]_{\boldsymbol{z}'_*}$$

By way of spectral decomposition, $\boldsymbol{G}_{z'_*} = \boldsymbol{V}^t_{G_*} \Lambda_{G_*} \boldsymbol{V}_{G_*}$, with $\boldsymbol{V}_{G_*}$ and Λ_{G_*} being, respectively, the eigenvector matrix and the diagonal eigenvalue matrix of the Hessian matrix $\boldsymbol{G}_{z'_*}$. Consider the orthogonal transformation $\boldsymbol{Z}'' = \boldsymbol{V}^t_{G_*} \boldsymbol{Z}'$ by which the new random vector $\boldsymbol{Z}''$ is also a normal random vector because it is a linear combination of the independent standard normal random variables $\boldsymbol{Z}'$. Furthermore, it can be shown that

$$E(\boldsymbol{Z}'') = \boldsymbol{0}$$

$$\text{Cov}(\boldsymbol{Z}'') = \boldsymbol{C}_{z''} = E(\boldsymbol{Z}'' \boldsymbol{Z}''^t) = \boldsymbol{V}^t_{G_*} \boldsymbol{C}_{Z''} \boldsymbol{V}_{G_*} = \boldsymbol{V}^t_{G_*} \boldsymbol{V}_{G_*} = \boldsymbol{I}$$

This indicates that $\boldsymbol{Z}''$ is also an independent standard normal random vector. In terms of $\boldsymbol{Z}''$, Eq. (4.85) can be expressed as

$$W(\boldsymbol{Z}'') \approx \boldsymbol{s}^t_{z''_*}(\boldsymbol{Z}'' - \boldsymbol{z}''_*) + \frac{1}{2}(\boldsymbol{Z}'' - \boldsymbol{z}''_*)^t \Lambda_{G_*} (\boldsymbol{Z}'' - \boldsymbol{z}''_*)$$

$$= \sum_{k=1}^{K} \boldsymbol{s}_{z'',k}(\boldsymbol{Z}''_k - \boldsymbol{z}''_{*,k}) + \frac{1}{2} \sum_{k=1}^{K} \lambda''_k (\boldsymbol{Z}''_k - \boldsymbol{z}''_{*,k})^2 = 0 \qquad (4.86)$$

in which $\boldsymbol{s}_{z''_{*,k}}$ is the kth element of sensitivity vector $\boldsymbol{s}_{z''_*} = \boldsymbol{V}^t_{G_*} \boldsymbol{s}_{z'_*}$ in $\boldsymbol{z}''$-space, and λ''_k is the kth eigenvalue of the Hessian matrix $\boldsymbol{G}_{z'_*}$.

In addition to Eqs. (4.85) and (4.86), the quadratic approximation of the performance function in the second-order reliability analysis can be expressed in a simpler form through other types of orthogonal transformation. Referring to Eq. (4.85), consider a $K \times K$ matrix $\boldsymbol{H}$ with its last column defined by the negativity of the unit directional derivatives vector $\boldsymbol{d}_* = -\boldsymbol{\alpha}_* = -\boldsymbol{s}_{z'_*}/|\boldsymbol{s}_{z'_*}|$ evaluated at the design point $\boldsymbol{z}'_*$, namely, $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_{K-1}, \boldsymbol{d}_*]$, with $\boldsymbol{h}_k$ being the kth column vector in $\boldsymbol{H}$. The matrix $\boldsymbol{H}$ is an *orthonormal matrix* because all column vectors are orthogonal to each other; that is, $\boldsymbol{h}^t_j \boldsymbol{h}_k = 0$, for $j \neq k$, $\boldsymbol{h}^t_k \boldsymbol{d}_* = 0$, and all of them have unit length $\boldsymbol{H}^t \boldsymbol{H} = \boldsymbol{H} \boldsymbol{H}^t = \boldsymbol{I}$. One simple way to find such an orthonormal matrix $\boldsymbol{H}$ is the *Gram-Schmid orthogonal transformation*, as described in Appendix 4D. Using the orthonormal matrix as defined above, a new random vector $\boldsymbol{U}$ can be obtained as $\boldsymbol{U} = \boldsymbol{H}^t \boldsymbol{Z}'$. As shown in Fig. 4.16, the orthonormal matrix $\boldsymbol{H}$ geometrically rotates the coordinates in the $\boldsymbol{z}'$-space to a new $\boldsymbol{u}$-space with its last $\boldsymbol{u}_K$ axis pointing in the direction of the design point $\boldsymbol{z}'_*$. It can be shown easily that the elements of the new random vector $\boldsymbol{U} = (U_1, U_2, \ldots, U_K)^t$ remain to be independent standard normal random variables as $\boldsymbol{Z}'$.

(a)

(b)
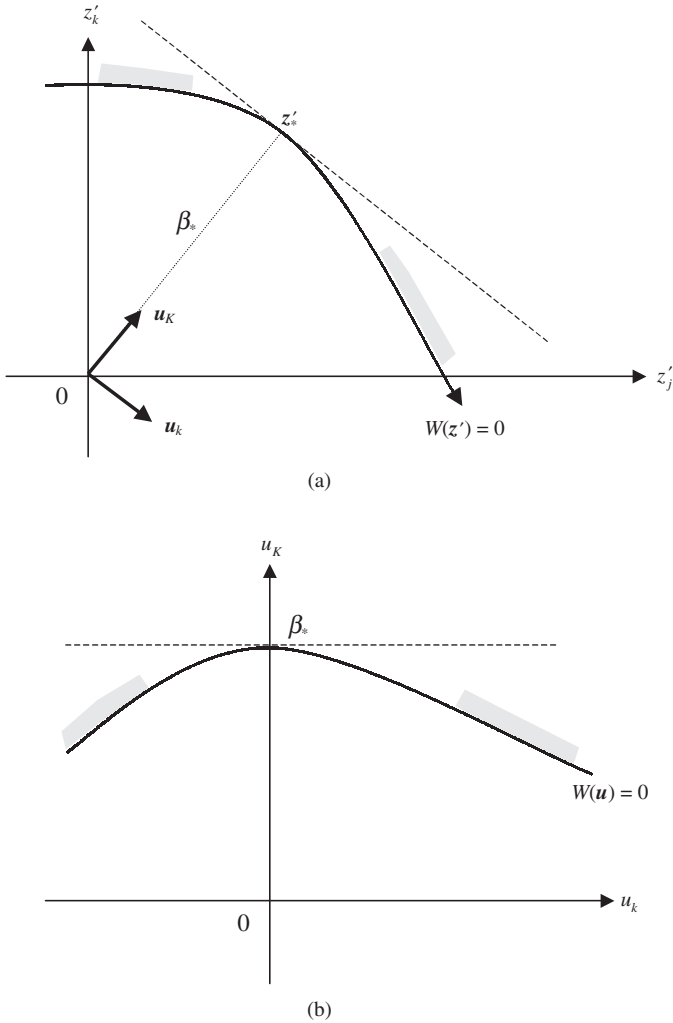
Figure 4.16 Geometric illustration of orthonormal rotation. (a) Before rotation (b) After rotation.

Knowing $\boldsymbol{z}'_* = \beta_* \boldsymbol{d}_*$, the orthogonal transformation using $\boldsymbol{H}$ results in

$$\boldsymbol{u}_* = \boldsymbol{H}^t \boldsymbol{z}'_* = \boldsymbol{H}^t (\beta_* \boldsymbol{d}_*) = \beta_* \boldsymbol{H}^t \boldsymbol{d}_* = \beta_* (0, 0, \dots, 1)^t$$

indicating that the coordinate of the design point in the transformed $\boldsymbol{u}$-space is $(0, 0, \dots, 0, \beta_*)$. In terms of the new $\boldsymbol{u}$-coordinate system, Eq. (4.2) can be expressed as

$$W(\boldsymbol{U}) \approx \boldsymbol{s}^t_{u_*}(\boldsymbol{U} - \boldsymbol{u}_*) + \frac{1}{2}(\boldsymbol{U} - \boldsymbol{u}_*)^t \boldsymbol{H}^t \boldsymbol{G}_{z'_*} \boldsymbol{H}(\boldsymbol{U} - \boldsymbol{u}_*) = 0 \qquad (4.87)$$

where $\boldsymbol{s}_{u_*} = \boldsymbol{H}^t \boldsymbol{s}_{z'_*}$, which simply is

$$
\begin{aligned}
\boldsymbol{s}^t_{u_*} &= \left(\boldsymbol{s}^t_{z'_*} \boldsymbol{h}_1, \boldsymbol{s}^t_{z'_*} \boldsymbol{h}_2, \ldots, \boldsymbol{s}^t_{z'_*} \boldsymbol{h}_{K-1}, \boldsymbol{s}^t_{z'_*} \boldsymbol{d}_* \right) \\
&= \left(-|\boldsymbol{s}_{z'_*}| \boldsymbol{d}^t_* \boldsymbol{h}_1, -|\boldsymbol{s}_{z'_*}| \boldsymbol{d}^t_* \boldsymbol{h}_2, \ldots, -|\boldsymbol{s}_{z'_*}| \boldsymbol{d}^t_* \boldsymbol{h}_{K-1}, -|\boldsymbol{s}_{z'_*}| \boldsymbol{d}^t_* \boldsymbol{d}_* \right) \\
&= (0, 0, \ldots, 0, -|\boldsymbol{s}_{z'_*}|)
\end{aligned}
\tag{4.88}
$$

After dividing $|\boldsymbol{s}_{z'_*}|$ on both sides of Eq. (4.4), it can be rewritten as

$$
W(\boldsymbol{U}) \approx \beta_* \boldsymbol{U}_k + \frac{1}{2}(\boldsymbol{U} - \boldsymbol{u}_*)^t \boldsymbol{A}_*(\boldsymbol{U} - \boldsymbol{u}_*) = 0
\tag{4.89}
$$

in which $\boldsymbol{A}_* = \boldsymbol{H}^t \boldsymbol{G}_{z'_*} \boldsymbol{H} / |\boldsymbol{s}_{z'_*}|$. Equation (4.6) can further be reduced to a parabolic form as

$$
W(\boldsymbol{U}) \approx \beta_* - U_K + \frac{1}{2}\tilde{\boldsymbol{U}}^t \tilde{\boldsymbol{A}}_* \tilde{\boldsymbol{U}} = 0
\tag{4.90}
$$

where $\tilde{\boldsymbol{U}} = (U_1, U_2, \ldots, U_{K-1})^t$ and $\tilde{\boldsymbol{A}}_*$ is the $(K-1)$th order leading principal submatrix of $\boldsymbol{A}_*$ obtained by deleting the last row and last column of matrix $\boldsymbol{A}_*$.

To further simplify the mathematical expression for Eq. (4.7), an orthogonal transformation is once more applied to $\tilde{\boldsymbol{U}}$ as $\tilde{\boldsymbol{U}}' = \boldsymbol{V}^t_{\tilde{A}_*} \tilde{\boldsymbol{U}}$ with $\boldsymbol{V}_{\tilde{A}_*}$ being the eigenvector matrix of $\tilde{\boldsymbol{A}}_*$ satisfying $\tilde{\boldsymbol{A}}_* = \boldsymbol{V}^t_{\tilde{A}_*} \Lambda_{\tilde{A}_*} \boldsymbol{V}_{\tilde{A}_*}$, in which $\Lambda_{\tilde{A}_*}$ is the diagonal eigenvalues matrix of $\tilde{\boldsymbol{A}}_*$. It can easily be shown that the elements of the new random vector $\tilde{\boldsymbol{U}}'$s are independent standard normal random variables. In terms of the new random vector $\tilde{\boldsymbol{U}}'$, the quadratic term in Eq. (4.7) can be rewritten as

$$
\begin{aligned}
W(\tilde{\boldsymbol{U}}', U_K) &\approx \beta_* - U_K + \frac{1}{2}\tilde{\boldsymbol{U}}'^t \Lambda_{\tilde{A}_*} \tilde{\boldsymbol{U}}' \\
&= \beta_* - U_K + \frac{1}{2}\sum_{k=1}^{K-1} \kappa_k \tilde{U}_k^2 = 0
\end{aligned}
\tag{4.91}
$$

where κ's are the main curvatures, which are equal to the elements of the diagonal eigenvalue matrix $\Lambda_{\tilde{A}_*}$ of matrix $\tilde{\boldsymbol{A}}_*$. Note that the eigenvalues of $\boldsymbol{A}_*$ are identical to those of $\boldsymbol{G}_{z'_*}$ defined in Eq. (4.2). This is so because $\boldsymbol{A}_* = \boldsymbol{H}^t \boldsymbol{G}_{z'_*} \boldsymbol{H}$ is a similarity transform. Therefore, the main curvatures of the hyperparabolic approximation of $W(\boldsymbol{Z}') = 0$ are equal to the eigenvalues of $\tilde{\boldsymbol{A}}_*$.

4.6.2 Breitung's formula

For a problem involving K independent standard normal random variables $\boldsymbol{Z}'$, the computation of failure probability involves multiple integration as

$$
p_f = \int_{W(\boldsymbol{z}')<0} \phi_K(\boldsymbol{z}') \, d\boldsymbol{z}'
\tag{4.92}
$$

where $\phi_K(\boldsymbol{z}')$ is the joint PDF of K independent standard normal random variables. This type of integration is called the *Laplace integral*, and its asymptotic characteristics have been investigated recently by Breitung (1993).

Once the design point $\boldsymbol{z}'_*$ is found and the corresponding reliability index $\beta_* = |\boldsymbol{z}'_*|$ is computed, Breitung (1984) shows that the failure probability based on a hyperparabolic approximation of $W(\boldsymbol{Z}')$, Eq. (4.92), can be estimated asymptotically (that is, $\beta_* \to \infty$) as

$$p_f \approx \Phi(-\beta_*) \prod_{k=1}^{K-1} (1 + \beta_* \kappa_k)^{-1/2} \tag{4.93}$$

where $\kappa_k, k = 1, 2, \ldots, K-1$, are the main curvatures of the performance function $W(\boldsymbol{Z}')$ at $\boldsymbol{z}'_*$, which is equal to the eigenvalues of the $(K-1)$ leading principal submatrix of $\tilde{\boldsymbol{A}}_*$ defined in Eq. (4.90). It should be pointed out that owing to the asymptotic nature of Eq. (4.93), the accuracy of estimating p_f by it may not be satisfactory when the value of $|\beta_*|$ is not large.

Equation (4.93) reduces to $p_f = \Phi(-\beta_*)$ if the curvature of the performance function is zero. A near-zero curvature of $W(\boldsymbol{Z}')$ in all directions at the design point implies that the performance function behaves like a hyperplane around $\boldsymbol{z}'_*$. In this case, $W(\boldsymbol{Z}')$ at $\boldsymbol{z}'_*$ can be described accurately by the first-order expansion terms, and reliability corresponds to the first-order failure probability. Figure 4.17 shows the ratio of second-order failure probability by Eq. (4.93) to the first-order failure probability as a function of main curvature and number of stochastic variables in the performance function. It is clearly shown in Fig. 4.17a that when the limit-state surface is convex toward the failure region with a constant positive curvature (see Fig. 4.15a), the failure probability estimated by the first-order method is larger than that by the second-order methods. This magnitude of the overestimation increases with the curvature
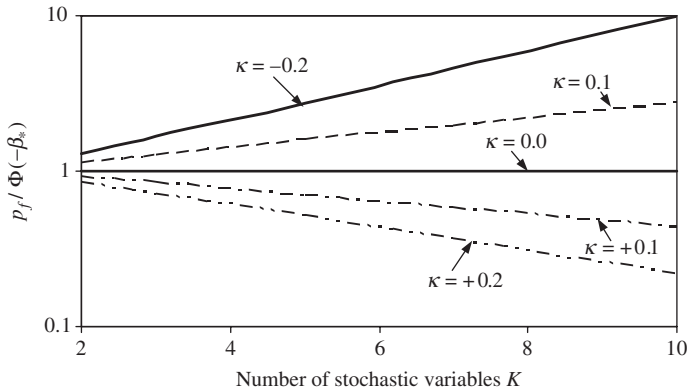
Figure 4.17 Comparison of the second-order and first-order failure probabilities for performance function with different curvatures.

and the number of stochastic basic variables involved. On the other hand, the first-order methods yield a smaller value of failure probability than the second-order methods when the limit-state surface is concaved toward the safe region (see Fig. 4.15b), which corresponds to a negative curvature. Hohenbichler and Rackwitz (1988) suggested further improvement on Breitung's results using the importance sampling technique (see Sec. 6.7.1).

In case there exists multiple design points yielding the same minimum distance β_*, Eq. (4.93) for estimating the failure probability p_f can be extended as

$$p_f \approx \Phi(-\beta_*) \left\{ \sum_{j=1}^{J} \left[\prod_{k=1}^{K-1} \left(1 + \beta_* \kappa_{k,j} \right)^{-1/2} \right] \right\} \qquad (4.94)$$

in which J is the number of design points with $\beta_* = |\mathbf{z}'_{*1}| = |\mathbf{z}'_{*2}| = \cdots = |\mathbf{z}'_{*J}|$, and $\kappa_{k,j}$ is the main curvature for the kth stochastic variables at the jth design point.

The second-order reliability formulas described earlier are based on fitting a paraboloid to the failure surface at the design points on the basis of curvatures. The computation of failure probability requires knowledge of the main curvatures at the design point, which are related to the eigenvalues of the Hessian matrix of the performance function. Der Kiureghian et al. (1987) pointed out several computational disadvantages of the paraboloid-fitting procedure:

1. When the performance function is not continuous and twice differentiable in the neighborhood of the design point, numerical differencing would have to be used to compute the Hessian matrix. In this case, the procedure may be computational intensive, especially when the number of stochastic variables is large and the performance function involves complicated numerical algorithms.

2. When using numerical differencing techniques for computing the Hessian, errors are introduced into the failure surface. This could result in error in computing the curvatures.

3. In some cases, the curvatures do not provide a realistic representation of the failure surface in the neighborhood of design point, as shown in Fig. 4.18.

   To circumvent these disadvantages of curvature-fitting procedure, Der Kiureghian et al. (1987) proposed an approximation using a point-fitted paraboloid (see Fig. 4.18) by which two semiparabolas are used to fit the failure surface in such a manner that both semiparabolas are tangent to the failure surface at the design point. Der Kiureghian et al. (1987) showed that one important advantage of the point-fitted paraboloid is that it requires less computation when the number of stochastic variables is large.
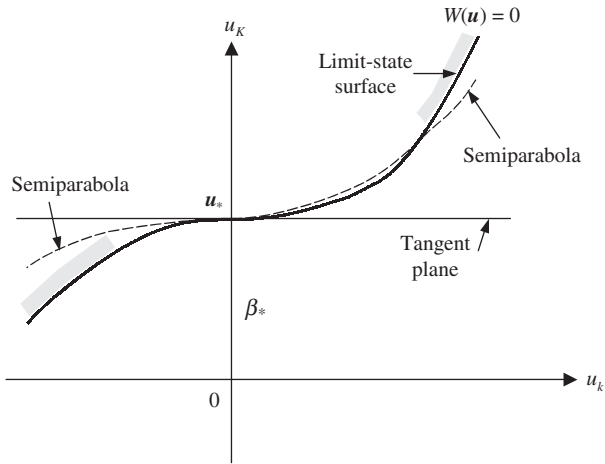
Figure 4.18 Fitting of paraboloid in rotated standard space. (*Der Kiurighian et al., 1987.*)

4.7 Time-Dependent Reliability Models

The development of hydrosystems engineering projects often includes the design of various types of hydraulic structures, such as pipe networks for water supply, storm sewer systems for runoff collection, levee and dike systems for flood control and protection, and others. Generally, the system, once designed and constructed, is expected to serve its intended objectives over a period of several years, during which the system behavior and environmental factors could change with respect to time. In such circumstances, engineers often are interested in evaluating the reliability of the hydraulic structure with respect to a specified time framework. For example, one might be interested in the risk of overflow of an urban storm water detention basin in the summer when convective thunderstorms prevail. Loads to most hydrosystems are caused by the occurrence of hydrologic events such as floods, storms, or droughts that are random by nature. Time-dependent reliability analysis considers repeated applications of loads and also can consider the change of the distribution of resistance with time.

In preceding sections, emphasis was placed on static reliability analysis, which does not consider the time dependency of the load and resistance. This section considers the time-dependent random variables in reliability analysis. As a result, the reliability is a function of time, i.e., time dependent or time variant. The difference between the time-to-failure analysis described in Chap. 6 and the time-dependent reliability analysis should be pointed out. The commonality between the two reliability analyses is that both attempt to assess the variation of reliability with respect to time. The difference lies in the manner in which the reliability is computed. Time-to-failure analysis is concerned

only with the time history of the performance of the system as a whole without giving explicit consideration to the load-resistance interference as done by time-dependent reliability analysis. The objective of time-dependent reliability models is to determine the system reliability over a specified time interval in which the number of occurrences of loads is a random variable.

When both loading and resistance are functions of time, the performance function $W(t) = R(t) - L(t)$ is time-dependent. Consequently, the reliability $p_s(t) = P[W(t) > 0]$ would vary with respect to time. Figure 4.19 shows schematically the key feature of the time-dependent reliability problem in which the PDFs of load and resistance change with time. In Fig. 4.19, the mean of resistance has a downward trend with time, whereas that of the load increases with time. As the standard deviations of both resistance and load increase with time, the area of interference increases, and this results in an increase in the failure probability with time. The static reliability analysis described in preceding sections considers neither load nor resistance being functions of time.

If the load is to be applied many times, it is often the largest load that is considered in reliability analysis. Then this maximum load can be described by an extreme-value distribution such as the Gumbel distribution described in Sec. 2.6.4. In doing so, the effect of time is ignored in reliability analysis, which may not be appropriate, especially when more than one load is involved or the resistance changes with time. A comprehensive treatment of time-dependent reliability issues can be found in Melchers (1999).
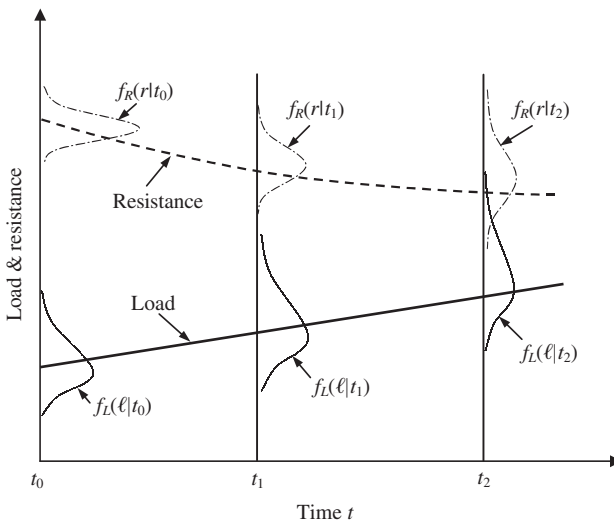
Figure 4.19 Time-dependence of load and resistance probability distribution functions.

4.7.1 Time-dependent resistance

For a hydraulic structure placed in a natural environment over a period of time, its operational characteristics could change over time owing to deterioration, aging, fatigue, and lack of maintenance. Consequently, the structural capacity (or resistance) would vary with respect to time. Examples of time-dependent characteristics of resistance in hydrosystems are change in flow-carrying capacity of storm sewers owing to sediment deposition and settlement, decrease in flow-carrying capacity in water distribution pipe networks owing to aging, seasonal variation in waste assimilative capacity of natural streams, etc.

Modeling time-dependent features of the resistance of a hydrosystem requires descriptions of the time-varying nature of statistical properties of the resistance. This would require monitoring resistance of the system over time, which, in general, is not practical. Alternatively, since the resistance of a hydrosystem may depend on several stochastic basic parameters, the time-dependent features of resistance of hydraulic structures or hydrosystems can be deduced, through appropriate engineering analysis, from the time-varying behavior of the stochastic parameters affecting the resistance of the systems. For example, the flow-carrying capacity of a storm sewer depends on pipe slope, roughness coefficient, and pipe size. Therefore, the time-dependent behavior of storm sewer capacity may be derived from the time-varying features of pipe slope, roughness coefficient, and pipe size by using appropriate hydraulic models.

Although simplistic in idea, information about the time-dependent nature of stochastic basic parameters in the resistance function of a hydrosystem is generally lacking. Only in a few cases and systems is partial information available. Table 4.6 shows the value of Hazen-Williams coefficient of cast iron pipe types

TABLE 4.6 Typical Hazen-Williams Pipe Roughness Coefficients for Cast Iron Pipes

Age (years) new	Pipe diameter all sizes	Roughness coefficient C_{HW} 130
5	>380 mm (15 in)	120
	>100 mm (4 in)	118
10	>600 mm (24 in)	113
	>300 mm (12 in)	111
	>100 mm (4 in)	107
20	>600 mm (24 in)	100
	>300 mm (12 in)	96
	>100 mm (4 in)	89
30	>760 mm (30 in)	90
	>400 mm (16 in)	87
	>100 mm (4 in)	75
40	>760 mm (30 in)	83
	>400 mm (16 in)	80
	>100 mm (4 in)	64

SOURCE: After Wood (1991).

as affected by pipe age. Owing to a lack of sufficient information to accurately define the time-dependent features of resistance or its stochastic basic parameters, it has been the general practice to treat them as time-invariant quantities by which statistical properties of resistance and its stochastic parameters do not change with time.

The preceding discussions consider the relationship between resistance and time only, namely, the aging effect. In some situations, resistance also could be affected by the number of occurrences of loadings and/or the associated intensity. If the resistance is affected only by the load occurrences, the effect is called *cyclic damage*, whereas if both load occurrence and its intensity affect the resistance, it is called *cumulative damage* (Kapur and Lamberson, 1977).

4.7.2   Time-dependent load

In time-dependent reliability analysis, one is concerned with system reliability over a specified time period during which external loads can occur more than once. Therefore, not only the intensity or magnitude of load is important but also the number or frequency of load occurrences is an important parameter.

Over an anticipated service period, the characteristics of load to be imposed on the system could change. For example, when a watershed undergoes a progressive change, it could induce time dependence in load. More specifically, the magnitude of floods could increase as urbanization progresses, and sediment discharge from overland erosion and non-point-source pollution could decrease over time if the farming and irrigation practices in the watershed involve pollution control measures. Again, characterization of the time-varying nature of load intensity requires extensive monitoring, data collection, and engineering analysis.

The occurrence of load over an anticipated service period can be classified into two cases (Kapur and Lamberson, 1977): (1) The number and time of occurrence are known, and (2) the number and time of occurrences are random. Section 4.7.4 presents probabilistic models for describing the occurrence and intensity of load.

4.7.3   Classification of time-dependent
reliability models

Repeated loadings on a hydrosystem are characterized by the time each load is applied and the behavior of time intervals between load applications. From a reliability theory viewpoint, the uncertainty about the loading and resistance variables may be classified into three categories: deterministic, random fixed, and random independent (Kapur and Lamberson, 1977). For the deterministic category, the loadings assume values that are exactly known a priori. For the random-fixed case, the randomness of loadings varies in time in a known manner. For the random-independent case, the loading is not only random, but the successive values assumed by the loading are statistically independent.

Deterministic. A variable that is deterministic can be quantified as a constant without uncertainty. A system with deterministic resistance and load implies that the behavior of the system is completely controllable, which is an idealized case. However, in some situations, a random variable can be treated as deterministic if its uncertainty is small and can be ignored.

Random fixed. A random-fixed variable is one whose initial condition is random in nature, and after its realization, the variable value is a known function of time. This can be expressed as

$$X_\tau = X_0\, g(\tau) \qquad \text{for } \tau > 0 \tag{4.95}$$

where X_0 and X_τ are, respectively, the random variable X at times $t = 0$ and $t = \tau$, and $g(\tau)$ is a known function involving time. Although X_t is a random variable, its PDF, however, is completely dependent on that of X_0. Therefore, once the value of the random initial condition X_0 is realized or observed, the value of subsequent time can be uniquely determined. For this case, given the PDF of X_0, the PDF and statistical moments of X_t can be obtained easily. For instance, the mean and variance of X_t can be obtained, in terms of those of X_0, as

$$E(X_t) = E(X_0)g(t) \qquad \text{for } t > 0 \tag{4.96a}$$

$$\text{Var}(X_t) = \text{Var}(X_0)g^2(t) \qquad \text{for } t > 0 \tag{4.96b}$$

in which $E(X_0)$ and $E(X_t)$ are the means of X_0 and X_t, respectively, and $\text{Var}(X_0)$ and $\text{Var}(X_t)$ are the variances of X_0 and X_t, respectively.

Random independent. A random-independent variable, unlike the random-fixed variable, whose values occurred at different times are not only random but also independent each other. There is no known relationship between the values of X_0 and X_t.

4.7.4 Modeling intensity and occurrence of loads

A hydraulic structure placed in a natural environment over an expected service period is subject to repeated application of loads of varying intensities. The magnitude of load intensity and the number of occurrences of load are, in general, random by nature. Therefore, probabilistic models that properly describe the stochastic mechanisms of load intensity and load occurrence are essential for accurate evaluation of the time-dependent reliability of hydrosystems.

Probability models for load intensity. In the great majority of situations in hydrosystems reliability analysis, the magnitudes of load to be imposed on the system are continuous random variables. Therefore, univariate probability distributions described in Sec. 2.6 potentially can be used to model the intensity of

a single random load. In a case in which more than one type of load is considered in the analysis, multivariate distributions should be used. Some commonly used multivariate distribution models are described in Sec. 2.7.

The selection of an appropriate probability model for load intensity depends on the availability of information. In a case for which sample data about the load intensity are available, formal statistical goodness-of-fit tests (see Sec. 3.7) can be applied to identify the best-fit distribution. On the other hand, when data on load intensity are not available, selection of the probability distribution for modeling load intensity has to rely on the analyst's logical judgment on the basis of the physical processes that produce the load.

Probability models for load occurrence.   In time-dependent reliability analysis, the time domain is customarily divided into a number of intervals such as days, months, or years, and the random nature of the load occurrence in each time interval should be considered explicitly. The occurrences of load are discrete by nature, which can be treated as a point random process. In Sec. 2.5, basic features of two types of discrete distributions, namely, binomial and Poisson distributions, for point process were described. This section briefly summarizes two distributions in the context of modeling the load-occurrences. Other load-occurrence models (e.g., renewal process, Polya process) can be found elsewhere (Melchers, 1999; Wen, 1987).

Bernoulli process.   A *Bernoulli process* is characterized by three features: (1) binary outcomes in each trial, (2) constant probability of occurrence of outcome in each time interval, and (3) the outcomes are independent between trials. In the context of load-occurrence modeling, each time interval represents a trial in which the outcome is either the occurrence or nonoccurrence of the load (with a constant probability) causing failure or nonfailure of the system. Hence the number of occurrences of load follows a binomial distribution, Eq. (2.51), with parameters p (the probability of occurrence of load in each time interval) and n (the number of time intervals). It is interesting to note that the number of intervals until the first occurrence T (the waiting time) in a Bernoulli process follows a *geometric distribution* with the PMF

$$g(T = t) = (1 - p)^{t-1} p \qquad (4.97)$$

The expected value of waiting time T is $1/p$, which is the mean occurrence period. It should be noted that the parameter p depends on the time interval used.

Poisson process.   In the Bernoulli process, as the time interval shrinks to zero and the number of time intervals increases to infinity, the occurrence of events reduces to a *Poisson process*. The conditions under which a Poisson process applies are (1) the occurrence of an event is equally likely at any time instant, (2) the occurrences of events are independent, and (3) only one event occurs at

a given time instant. The PMF describing the number of occurrences of loading in a specified time period $(0, t]$ is given by Eq. (2.55) and is repeated here:

$$P_x(x|\lambda, t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} \qquad \text{for } x = 0, 1, \ldots$$

in which λ is the average time rate of occurrence of the event of interest. The interarrival time between two successive occurrences is described by an exponential distribution with the PDF

$$f_t(t|\lambda) = \lambda e^{-\lambda t} \qquad \text{for } t > 0 \tag{4.98}$$

Although condition (1) implies that the Poisson process is stationary, it can be generalized to a nonstationary Poisson process, in which the rate of occurrence is a function of time $\lambda(t)$. Then the Poisson PMF for a nonstationary process can be written as

$$P(X = x) = \frac{\left[\int_0^t \lambda(\tau)\, d\tau\right]^x \exp\left[-\int_0^t \lambda(\tau)\, d\tau\right]}{x!} \tag{4.99}$$

Equation (4.99) allows one to incorporate the seasonality of many hydrologic events.

4.7.5 Time-dependent reliability models

Reliability computations for time-dependent models can be made for deterministic and random cycle times. The development of a model for deterministic cycles is given first, which naturally leads to the model for random cycle times.

Number of occurrences is deterministic. Consider a hydrosystem with a fixed resistance (or capacity) $R = r$ subject to n repeated loads $L_1, L_2, \ldots, L_n$. When the number of loads n and system capacity r are fixed, the reliability of the system after n loadings $p_s(n, r)$ can be expressed as

$$p_s(n, r) = P[(L_1 < r) \cap (L_2 < r) \cap \cdots \cap (L_n < r)] = P(L_{\max} < r) \tag{4.100}$$

where $L_{\max} = \max\{L_1, L_2, \ldots, L_n\}$, which also is a random variable. If all random loadings L are independent with their own distributions, Eq. (4.100) can be written as

$$p_s(n, r) = \prod_{i=1}^{n} [F_{L_i}(r)] \tag{4.101}$$

where $F_{L_i}(r)$ is the CDF of the ith load. In the case that all loadings are generated by the same statistical process, that is, all L's are identically distributed with $F_{L_i}(r) = F_L(r)$, for $i = 1, 2, \ldots, n$, Eq. (4.101) can further be reduced to

$$p_s(n, r) = [F_L(r)]^n \tag{4.102}$$

If the resistance of the system also is a random variable, the system reliability under the fixed number of loads n can be expressed as

$$p_s(n) = \int_0^\infty p_s(n, r) \, f_R(r) \, dr \qquad (4.103)$$

Number of occurrences is random. Since the loadings to hydrosystems are related to hydrologic events, the occurrence of the number of loads, in general, is uncertain. The reliability of the system under random loading in the specified time interval $[0, t]$ can be expressed as

$$p_s(t) = \sum_{n=0}^\infty \pi\,(t|n) \, p_s(n) \qquad (4.104)$$

in which $\pi(t|n)$ is the probability of n loadings occurring in the time interval $[0, t]$. A Poisson distribution can be used to describe the probability of the number of events occurring in a given time interval. In fact, the Poisson distribution has been found to be an appropriate model for the number of occurrences of hydrologic events (Clark, 1998; Todorovic and Yevjevich, 1969; Zelenhasic, 1970). Referring to Eq. (2.55), $\pi(t|n)$ can be expressed as

$$\pi(t|n) = \frac{e^{-\lambda t}\,(\lambda t)^n}{n!} \qquad (4.105)$$

where λ is the mean rate of occurrence of the loading in $[0, t]$, which can be estimated from historical data.

Substituting Eq. (4.105) in Eq. (4.104), the time-dependent reliability for the random independent load and random-fixed resistance can be expressed as

$$p_s(t) = \sum_{n=0}^\infty \left[\frac{e^{-\lambda t}\,(\lambda t)^n}{n!}\right] \left[\int_0^\infty p_s(n, r) \, f_R(r) \, dr\right] \qquad (4.106)$$

Under the condition that random loads are independently and identically distributed, Eq. (4.106) can be simplified as

$$p_s(t) = \int_0^\infty e^{-\lambda t\,[1 - F_L(r)]} \, f_R(r) \, dr \qquad (4.107)$$

4.7.6 Time-dependent reliability models for hydrosystems

Considering only inherent hydrologic uncertainty. Traditionally, the risk associated with the natural hydrologic randomness of flow or rainfall is explicitly considered in terms of a return period. By setting the resistance equal to the load with a return period of T years (that is, $r_* = \ell_T$), the annual reliability, without considering the uncertainty associated with ℓ_T, is $1 - 1/T$, that is,

$P(L < r_*|r_* = \ell_T) = 1 - 1/T$. Correspondingly, the reliability that the random loads would not exceed $r_* = \ell_T$ in a period of t years can be calculated as (Yen, 1970)

$$p_s(t, T) = \left(1 - \frac{1}{T}\right)^t \tag{4.108}$$

For large T, Eq. (4.108) reduces to

$$p_s(t, T) = \exp(-t/T) \tag{4.109}$$

If $T > t$, Eq. (4.108) can further be approximated simply as $p_s(t, T) = 1 - t/T$.

Considering both inherent hydrologic uncertainty and hydraulic uncertainty. In the case where the uncertainty of the resistance is not negligible and is to be considered, the annual reliability of a hydrosystem infrastructure then has to be evaluated through load-resistance interference on an annual basis. That is, the annual reliability will be calculated by evaluating $P(L \leq R)$ as Eq. (4.1), with $f_L(\ell)$ being the probability distribution function of annual maximum load. Hence the reliability of a hydrosystem over a service period of t years can be calculated by replacing the term $1/T$ in Eqs. (4.108) and (4.109) by $1 - P(L \leq R)$. Then the results are

$$p_s(t, L, R) = [P(L \leq R)]^t \tag{4.110}$$

$$p_s(t, L, R) = \exp\{-t \times [1 - P(L \leq R)]\} \tag{4.111}$$

in which the evaluation of annual reliability $P(L \leq R)$ can be made through the reliability methods described in preceding sections.

Incorporation of a design event. In the design of hydraulic structures, the common practice is to determine the design capacity based on a preselected design return period ℓ_T and safety factor SF. Under such a condition, the magnitude of the future annual maximum hydrologic load can be partitioned into two complementary subsets, that is, $\ell \leq \ell_T$ and $\ell \geq \ell_T$, with each representing different recurrence intervals of the hydrologic process. The reliability of the hydrosystem subject to the ith hydrologic load occurring in the future can be expressed by using the total probability theorem (Sec. 2.2.4) as

$$
\begin{aligned}
p_{s,i} &= P(L_i \leq r) \\
&= P(L_i \leq r \,|\, L_i \geq \ell_T) \, P(L_i \geq \ell_T) + P(L_i \leq r \,|\, L_i \leq \ell_T) \, P(L_i \leq \ell_T) \\
&= P(\ell_T \leq L_i \leq r) + P(L_i \leq r, L_i \leq \ell_T) \\
&= P_1 + P_2
\end{aligned}
\tag{4.112}
$$

where P_1 and P_2 can be written explicitly as

$$P_1 = P(\ell_T \leq L \leq R) = \int_{\ell_T}^{\infty} \left[\int_{\ell_T}^{r} f_{R,L}(r, \ell)\, d\ell \right] dr \tag{4.113}$$

$$P_2 = P(L \leq R, L < \ell_T)$$

$$= \int_{0}^{\ell_T} \left[\int_{0}^{r} f_{R,L}(r, \ell)\, d\ell \right] dr + \int_{\ell_T}^{\infty} \left[\int_{0}^{\ell_T} f_{R,L}(r, \ell)\, d\ell \right] dr \tag{4.114}$$

where ℓ_T is the magnitude of the design hydrologic event associated with a return period of T years. Based on this partition of the load domain, Tung (1985) presented two generalized time-dependent reliability models as follows:

$$p_s(t, T, SF) = \sum_{x=0}^{t} C_{t,x} P_1^x P_2^{t-x} \tag{4.115}$$

and

$$p_s(t, T, SF) = \sum_{n=0}^{\infty} \frac{e^{-t} t^n}{n!} \left(\sum_{x=0}^{n} C_{n,x} P_1^x P_2^{n-x} \right) \tag{4.116}$$

in which $C_{n,x} = n!/[(n-x)!x!]$ is a binomial coefficient, t is the expected service life (in years), T is the design return period (in years), SF is the safety factor, and n is the number of occurrences of load within the service life. From the design viewpoint, the selected design load ℓ_T and safety factor SF are reflected in the determination of the mean resistance of the structure μ_r as

$$\mu_r = SF\,\ell_T \tag{4.117}$$

Equation (4.115) is based on the binomial distribution for random occurrence of the loads, whereas Eq. (4.116) is based on the Poisson distribution. When hydraulic uncertainty is negligible, Eqs. (4.115) and (4.116) reduce, respectively, to

$$p_s(t, T, SF) = \left[1 - \frac{1}{T(SF)} \right]^t \tag{4.118}$$

and

$$p_s(t, T, SF) = \exp[-t/T(SF)] \tag{4.119}$$

In Eqs. (4.118) and (4.119), the return period $T(SF)$ can be determined by $1/(1 - p_s)$, in which p_s is computed by Eq. (4.35) with $r = SF\,\ell_T$. Equations (4.118) and (4.119) are used frequently by engineers in hydrologic designs that correspond to Eqs. (4.108) and (4.109) under $SF = 1$ (Yen, 1970). On the other hand, when both hydrologic and hydrologic-inherent uncertainties are considered, Eq. (4.117) can be explicitly incorporated in calculating the annual reliability $P[L \leq R(\ell_T, SF)]$ through load-resistance interference and then use of Eqs. (4.110) or (4.111) for calculating the reliability over a specified service period (Gui et al., 1998).

Figure 4.20 indicates that the time-dependent models using the Poisson distribution yield slightly lower failure-probability values. The values of failure
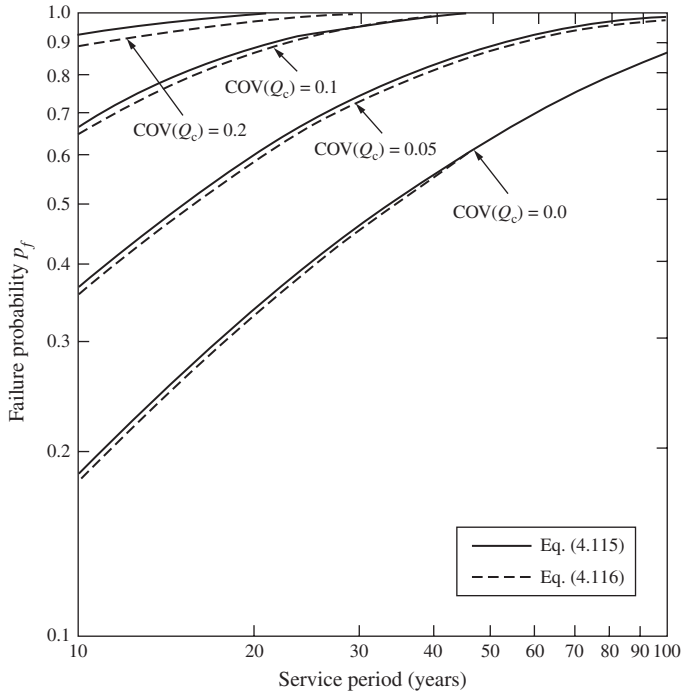
Figure 4.20 Comparison of two generalized time-dependent reliability models under $T = 50$ years, $SF = 1.0$, and $\Omega_L = 0.1$. (*After Tung, 1985.*)

probability computed by the two models converge as the service life increases. Without considering hydraulic uncertainty [i.e., $\text{Cov}(Q_c) = 0$], the failure probability is significantly underestimated. Computationally, the time-dependent model based on the binomial distribution, i.e., Eq. (4.39a), is much simpler than that based on the Poisson distribution.

Appendix 4A: Some One-Dimensional Numerical Integration Formulas

This appendix summarizes some commonly used numerical formulas for evaluating the following integral:

$$I = \int_a^b f(x)\, dx \tag{4A.1}$$

Detailed descriptions of these and other numerical integration procedures can be found in any numerical analysis textbook.

4A.1 Trapezoidal rule

For a closed integral, Eq. (4A.1) can be approximated as

$$I = \frac{h}{2}\left(f_1 + 2\sum_{i=1}^{n-1} f_i + f_n \right) \tag{4A.2a}$$

where h is a constant space increment for discretization, n is the number of discretization points over the interval (a, b), including the two end points, and f_i is the function values at discretized point, x_i.

For open and semiopen integrals, Eq. (4A.1) can be computed numerically as

$$I = \frac{h}{2}\left(3f_2 + 2\sum_{i=3}^{n-2} f_i + 3f_{n-1} \right) \tag{4A.2b}$$

4A.2 Simpson's rule

For closed integrals, one has

$$I = \frac{h}{3}[f_1 + 4(f_2 + f_4 + f_6 + \cdots) + 2(f_3 + f_5 + f_7 + \cdots) + f_n] \tag{4A.3a}$$

For open and semiopen integrals, one has

$$I = \frac{h}{12}[27f_2 + 13(f_4 + f_6 + \cdots) + 16(f_5 + f_7 + \cdots) + 27f_{n-1}] \tag{4A.3b}$$

4A.3 Gaussian quadratures

Equation (4A.1) can be expressed as

$$I = \sum_{i=1}^{n} w_i f(x_i) \tag{4A.4}$$

where w_i is the weight associated with the ith abscissa x_i in the discretization. The weight w_i is related to orthogonal polynomials. Table 4A.1 lists some commonly used orthogonal polynomials and their applied integral range, abscissas, and weights. Definitions of those polynomials and tables of abscissas and weights for different Gaussian quadratures are given by Abramowitz and Stegun (1972).

TABLE 4A.1 Some Commonly Used Gaussian Quadratures

Gauss	Range (a, b)	Abscissas x_i	Weight w_i
Legendre	$(-1, 1)$	ith root of $P_n(x)$	$\dfrac{2}{\left(1 - x_i^2\right)[P_{n'}(x_i)]^2}$
Chebyshev	$(-1, 1)$	$\cos\left[\dfrac{(2i - 1)\pi}{2n}\right]$	$\dfrac{\pi}{n}$
Laguerre	$(0, \infty)$	ith root of $L_n(x)$	$\dfrac{(n!)^2 x_i}{(n + 1)^2[L_{n+1}(x_i)]^2}$
Hermite	$(-\infty, \infty)$	ith root of $H_n(x)$	$\dfrac{2^{n-1} n! \sqrt{\pi}}{n^2[H_{n-1}(x_i)]^2}$

NOTE: $P_n(x)$ = Legendre polynomial of order n; $L_n(x)$ = Laguerre polynomial of order n; $H_n(x)$ = Hermite polynomial of order n.

Appendix 4B: Cholesky Decomposition

For any nonsingular square matrix $\boldsymbol{A}$, it can be decomposed as

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} \tag{4B.1}$$

where $\boldsymbol{L}$ is a lower triangular matrix as

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ l_{K1} & l_{K2} & l_{K3} & \dots & l_{KK} \end{bmatrix}$$

and $\boldsymbol{U}$ is an upper triangular matrix. In general, the matrices $\boldsymbol{L}$ and $\boldsymbol{U}$ are not unique. However, Young and Gregory (1973) show that if the diagonal elements of $\boldsymbol{L}$ or $\boldsymbol{U}$ are specified, the decomposition will be unique.

When the matrix $\boldsymbol{A}$ is real, symmetric, and positive-definite, then $\boldsymbol{U} = \boldsymbol{L}^t$, which means that $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^t$. This is called the *Cholesky decomposition*. Writing out $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^t$ in components, one readily obtains the following relationships between the elements in matrices $\boldsymbol{L}$ and $\boldsymbol{A}$ as

$$l_{kk}^2 + \sum_{j=1}^{k-1} l_{kj}^2 = a_{kk} \qquad \text{for } k = 1, 2, \dots, K \tag{4B.2}$$

$$l_{kk}l_{jj} + \sum_{i=1}^{j-1} l_{ki}l_{ji} = a_{kj} \qquad \text{for } k = j + 1, \dots, K \tag{4B.3}$$

in which l_{kj} and a_{kj} are elements in matrices $\boldsymbol{L}$ and $\boldsymbol{A}$, respectively, and K is the size of the matrices. In terms of a_{kj}'s, l_{kj}'s can be expressed as

$$l_{kk} = \left(a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2 \right) \tag{4B.4}$$

$$l_{kj} = \frac{1}{l_{kk}} \left(a_{kj} - \sum_{i=1}^{j-1} l_{ki} l_{ji} \right) \qquad \text{for } k = j+1, \ldots, K \tag{4B.5}$$

Computationally, the values of l_{kj}'s can be obtained by solving Eqs. (4B.4) and (4B.5) sequentially following the order $k = 1, 2, \ldots, K$. Numerical examples can be found in Wilkinson (1965, p. 71). A simple computer program for the Cholesky decomposition is available from Press et al. (1992, p. 90). Note that the requirement of positive definite for matrix $\boldsymbol{A}$ is to ensure that the quantity in the square root of Eq. (4B.4) always will be positive throughout the computation. If $\boldsymbol{A}$ is not a positive-definite matrix, the algorithm will fail.

For a real, symmetric, positive-definite matrix $\boldsymbol{A}$, the Cholesky decomposition is sometimes expressed as

$$\boldsymbol{A} = \tilde{\boldsymbol{L}} \boldsymbol{\Lambda} \tilde{\boldsymbol{L}}^t \tag{4B.6}$$

in which $\boldsymbol{L}$ is a unit lower triangular matrix with all its diagonal elements having values of ones, and $\boldsymbol{\Lambda}$ is a diagonal eigenvalue matrix. Therefore, the eigenvalues associated with matrix $\boldsymbol{A}$ are the square roots of the diagonal elements in matrix $\boldsymbol{L}$. If a matrix is positive-definite, all its eigenvalues will be positive, and vice versa.

In theory, the covariance and correlation matrices in any multivariate problems should be positive-definite. In practice, sample correlation and sample covariance often are used in the analysis. Owing to the sampling errors, the resulting sample correlation matrix may not be positive-definite, and in such cases, the Cholesky decomposition may fail, whereas the spectral decomposition described in Appendix 4C can be applicable.

Appendix 4C: Orthogonal Transformation Techniques

The *orthogonal transformation* is an important tool for treating problems with correlated stochastic basic variables. The main objective of the transformation is to map correlated stochastic basic variables from their original space to a new domain in which they become uncorrelated. Hence the analysis is greatly simplified.

Consider K multivariate stochastic basic variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$ having a mean vector $\boldsymbol{\mu}_x = (\mu_1, \mu_2 \ldots, \mu_K)^t$ and covariance matrix $\boldsymbol{C}_x$ as

$$\boldsymbol{C}_x = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2K} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ \sigma_{K1} & \sigma_{K2} & \sigma_{K3} & \cdots & \sigma_{KK} \end{bmatrix}$$

in which $\sigma_{ij} = \text{Cov}(X_i, X_j)$, the covariance between stochastic basic variables X_i and X_j. The vector of correlated standardized stochastic basic variables $\boldsymbol{X}' = \boldsymbol{D}_x^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}_x)$, that is, $\boldsymbol{X}' = (X_1', X_2', \ldots, X_K')^t$ with $X_k' = (X_k - \mu_k)/\sigma_k$, for $k = 1, 2, \ldots, K$, and $\boldsymbol{D}_x$ being an $K \times K$ diagonal matrix of variances of stochastic basic variables, that is, $\boldsymbol{D}_x = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$, would have a mean vector of $\boldsymbol{0}$ and the covariance matrix equal to the correlation matrix $\boldsymbol{R}_x$:

$$\boldsymbol{C}_{x'} = \boldsymbol{R}_x = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1K} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2K} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ \rho_{K1} & \rho_{K2} & \rho_{K3} & \cdots & 1 \end{bmatrix}$$

Note that from Sec. 2.4.5, the covariance matrix and correlation matrix are symmetric matrices, that is, $\sigma_{ij} = \sigma_{ji}$ and $\rho_{ij} = \rho_{ji}$, for $i \neq j$. Furthermore, both matrices theoretically should be positive-definite.

In the orthogonal transformation, a $K \times K$ square matrix $\boldsymbol{T}$ (called the *transformation matrix*) is used to transform the standardized correlated stochastic basic variables $\boldsymbol{X}'$ into a set of uncorrelated standardized stochastic basic variables $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = \boldsymbol{T}^{-1}\boldsymbol{X}' \tag{4C.1}$$

where $\boldsymbol{Y}$ is a vector with the mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$, a $K \times K$ identity matrix. Stochastic variables $\boldsymbol{Y}$ are uncorrelated because the off-diagonal elements of the covariance matrix are all zeros. If the original stochastic basic variables $\boldsymbol{X}$ are multivariate normal variables, then $\boldsymbol{Y}$ is a vector of uncorrelated standardized normal variables specifically designated as $\boldsymbol{Z}'$ because the right-hand side of Eq. (4C.1) is a linear transformation of the normal random vector.

It can be shown that from Eq. (4C.1), the transformation matrix $\boldsymbol{T}$ must satisfy

$$\boldsymbol{R}_x = \boldsymbol{T}\boldsymbol{T}^t \tag{4C.2}$$

There are several methods that allow one to determine the transformation matrix in Eq. (4C.2). Owing to the fact that $\boldsymbol{R}_x$ is a symmetric and positive-definite matrix, it can be decomposed into

$$\boldsymbol{R}_x = \boldsymbol{L}\boldsymbol{L}^t \tag{4C.3}$$

in which $\boldsymbol{L}$ is a $K \times K$ *lower triangular matrix* (Young and Gregory, 1973; Golub and Van Loan, 1989):

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & 0 & 0 & \ldots & 0 \\ l_{21} & l_{22} & 0 & \ldots & 0 \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ l_{K1} & l_{K2} & l_{K3} & \ldots & l_{KK} \end{bmatrix}$$

which is unique. Comparing Eqs.(4C.2) and (4C.3), the transformation matrix $\boldsymbol{T}$ is the lower triangular matrix $\boldsymbol{L}$. An efficient algorithm to obtain such a lower triangular matrix for a symmetric and positive-definite matrix is the *Cholesky decomposition* (or *Cholesky factorization*) method (see Appendix 4B).

The orthogonal transformation alternatively can be made using the *eigenvalue-eigenvector decomposition* or *spectral decomposition* by which $\boldsymbol{R}_x$ is decomposed as

$$\boldsymbol{R}_x = \boldsymbol{C}_{x'} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^t \tag{4C.4}$$

where $\boldsymbol{V}$ is a $K \times K$ eigenvector matrix consisting of K eigenvectors as $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_K)$, with $\boldsymbol{v}_k$ being the kth eigenvector of the correlation matrix $\boldsymbol{R}_x$, and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K)$ being a diagonal eigenvalues matrix. Frequently, the eigenvectors $\boldsymbol{v}$'s are normalized such that the norm is equal to unity, that is, $\boldsymbol{v}^t\boldsymbol{v} = 1$. Furthermore, it also should be noted that the eigenvectors are orthogonal, that is, $\boldsymbol{v}_i^t\boldsymbol{v}_j = 0$, for $i \neq j$, and therefore, the eigenvector matrix $\boldsymbol{V}$ obtained from Eq. (4C.4) is an orthogonal matrix satisfying $\boldsymbol{V}\boldsymbol{V}^t = \boldsymbol{V}^t\boldsymbol{V} = \boldsymbol{I}$ where $\boldsymbol{I}$ is an identity matrix (Graybill, 1983). The preceding orthogonal transform satisfies

$$\boldsymbol{V}^t\boldsymbol{R}_x\boldsymbol{V} = \boldsymbol{\Lambda} \tag{4C.5}$$

To achieve the objective of breaking the correlation among the standardized stochastic basic variables $\boldsymbol{X}'$, the following transformation based on the eigenvector matrix can be made:

$$\boldsymbol{U} = \boldsymbol{V}^t\boldsymbol{X}' \tag{4C.6}$$

The resulting transformed stochastic variables $\boldsymbol{U}$ has the mean and covariance matrix as

$$E(\boldsymbol{U}) = \boldsymbol{V}^t E(\boldsymbol{X}') = \boldsymbol{0} \tag{4C.7a}$$

and $\qquad\qquad C(\boldsymbol{U}) = \boldsymbol{V}^t\boldsymbol{C}_{x'}\boldsymbol{V} = \boldsymbol{V}^t\boldsymbol{R}_x\boldsymbol{V} = \boldsymbol{\Lambda} \tag{4C.7b}$

As can be seen, the new vector of stochastic basic variables U obtained by Eq. (4C.6) is uncorrelated because its covariance matrix C_u is a diagonal matrix Λ. Hence, each new stochastic basic variable U_k has the standard deviation equal to $\sqrt{\lambda_k}$, for all $k = 1, 2, \ldots, K$.

The vector U can be standardized further as

$$Y = \Lambda^{-1/2}U \tag{4C.8}$$

Based on the definitions of the stochastic basic variable vectors $X \sim (\mu_x, C_x)$, $X' \sim (0, R_x)$, $U \sim (0, \Lambda)$, and $Y \sim (0, I)$ given earlier, relationships between them can be summarized as the following:

$$Y = \Lambda^{-1/2}U = \Lambda^{-1/2}V^t X' \tag{4C.9}$$

Comparing Eqs.(4C.1) and (4C.9), it is clear that

$$T^{-1} = \Lambda^{-1/2}V^t$$

Applying an inverse operator on both sides of the equality sign, the transformation matrix T alternatively, as opposed to Eq. (4C.3), can be obtained as

$$T = V\Lambda^{1/2} \tag{4C.10}$$

Using the transformation matrix T as given above, Eq. (4C.1) can be expressed as

$$X' = TY = V\Lambda^{1/2}Y \tag{4C.11a}$$

and the random vector in the original parameter space is

$$X = \mu_x + D^{1/2}V\Lambda^{1/2}Y = \mu_x + D^{1/2}LY \tag{4C.11b}$$

Geometrically, the stages involved in orthogonal transformation from the originally correlated parameter space to the standardized uncorrelated parameter space are shown in Fig. 4C.1 for a two-dimensional case.

From Eq. (4C.1), the transformed variables are linear combinations of the standardized original stochastic basic variables. Therefore, if all the original stochastic basic variables X are normally distributed, then the transformed stochastic basic variables, by the reproductive property of the normal random variable described in Sec. 2.6.1, are also independent normal variables. More specifically,

$$X \sim \mathrm{N}(\mu_x, C_x) \qquad X' \sim \mathrm{N}(0, R_x) \qquad U \sim \mathrm{N}(0, \Lambda) \text{ and } Y = Z' \sim \mathrm{N}(0, I)$$

The advantage of the orthogonal transformation is to transform the correlated stochastic basic variables into uncorrelated ones so that the analysis can be made easier.
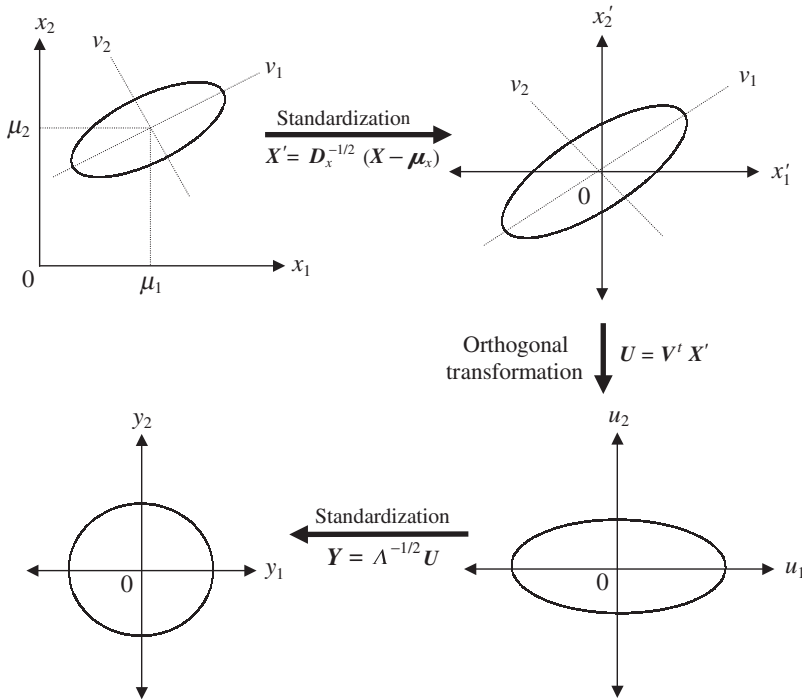
Figure 4C.1 Geometric diagrams of various stages of transformations in spectral decomposition. (*Tung and Yen, 2005.*)

The orthogonal transformations described earlier are applied to the standardized parameter space in which the lower triangular matrix and eigenvector matrix of the correlation matrix are computed. In fact, the orthogonal transformation can be applied directly to the variance-covariance matrix $\boldsymbol{C}_x$. The lower triangular matrix of $\boldsymbol{C}_x$, $\tilde{\boldsymbol{L}}$, can be obtained from that of the correlation matrix $\boldsymbol{L}$ by

$$\tilde{\boldsymbol{L}} = \boldsymbol{D}_x^{1/2}\boldsymbol{L} \tag{4C.12}$$

Following a similar procedure to that described for spectral decomposition, the uncorrelated standardized random vector $\boldsymbol{Y}$ can be obtained as

$$\boldsymbol{Y} = \tilde{\boldsymbol{\Lambda}}^{-1/2}\tilde{\boldsymbol{V}}^t(\boldsymbol{X} - \boldsymbol{\mu}_x) = \tilde{\boldsymbol{\Lambda}}^{-1/2}\tilde{\boldsymbol{U}} \tag{4C.13}$$

where $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\Lambda}}$ are the eigenvector matrix and diagonal eigenvalue matrix of the covariance matrix $\boldsymbol{C}_x$ satisfying

$$\boldsymbol{C}_x = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{V}}^t$$

and $\check{U}$ is an uncorrelated vector of the random variables in the eigenspace having a zero mean $\mathbf{0}$ and covariance matrix $\tilde{\Lambda}$. Then the original random vector X can be expressed in terms of Y and $\check{L}$:

$$X = \mu_x + \tilde{V}\tilde{\Lambda}^{1/2}Y = \mu_x + \check{L}Y \qquad (4C.14)$$

One should be aware that the eigenvectors and eigenvalues associated with the covariance matrix C_x will not be identical to those of the correlation matrix R_x.

Appendix 4D: Gram-Schmid Ortho-normalization

Consider a vector x_1 in an K-dimensional space to be used as one of the basis vectors. It is desirable to find the additional vectors, along with x_1, so that they would form K orthonormal basis vectors for the K-dimensional space. To do that, one can arbitrarily select $K-1$ vectors in the K-dimensional space as x_2, $x_3, \ldots, x_K$.

The first basis vector can be obtained as $u_1 = x_1/|x_1|$. Referring to Fig. 4D.1, a second basis vector (not necessarily normalized) that will be orthogonal to the first basis vector u_1 can be derived as

$$y_2 = x_2 - \hat{y}_2 = x_2 - \left(x_2^t u_1\right)u_1$$

Therefore, the second normalized basis vector u_2, that is perpendicular to u_2 can be determined as $u_2 = y_2/|y_2|$.

Note that the third basis vector must be orthogonal to the previously determined basis vectors (u_1, u_2) or (y_1, y_2). Referring to Fig. 4D.2, the projection of x_3 onto the plane defined by y_1 and y_2 is

$$\hat{y}_3 = \left(x_3^t u_1\right)u_1 + \left(x_3^t u_2\right)u_2$$
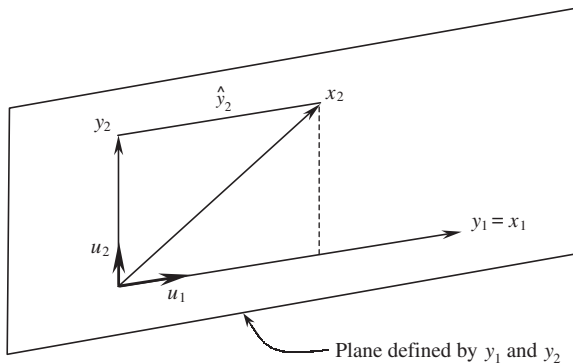
Figure 4D.1   Determination of the second basis vector.
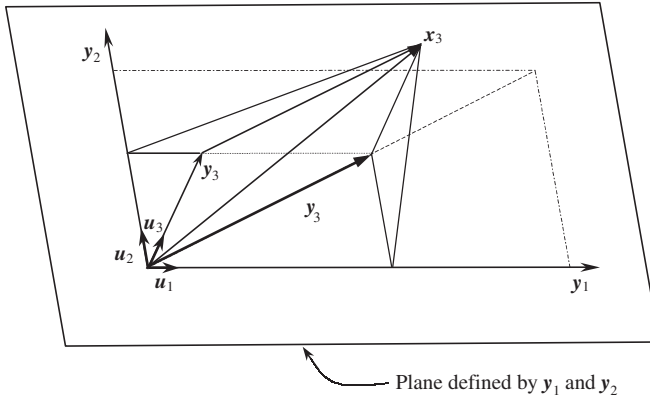
Figure 4D.2 Determination of the third basis vector.

Therefore, the third basis vector $\boldsymbol{y}_3$ that is orthogonal to both $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ can be determined as

$$\boldsymbol{y}_3 = \boldsymbol{x}_3 - \hat{\boldsymbol{y}}_3 = \boldsymbol{x}_3 - \left[(\boldsymbol{x}_3^t \boldsymbol{u}_1) \boldsymbol{u}_1 + (\boldsymbol{x}_3^t \boldsymbol{u}_2) \boldsymbol{u}_2 \right]$$

and the corresponding normalized basis vector $\boldsymbol{u}_3$ can be determined as $\boldsymbol{u}_3 = \boldsymbol{y}_3/|\boldsymbol{y}_3|$.

From the preceding derivation, the kth basis vector $\boldsymbol{y}_k$ can be computed as

$$\boldsymbol{y}_k = \boldsymbol{x}_k - \left[\sum_{i=1}^{k} (\boldsymbol{x}_k^t \boldsymbol{\mu}_i) \boldsymbol{\mu}_i \right] \qquad \text{for } k = 2, 3, \ldots, k \qquad (4\text{D}.1)$$

In the case that $\boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_K$ are unit vectors, the basis vectors $\boldsymbol{y}_2, \boldsymbol{y}_3, \ldots, \boldsymbol{y}_K$ obtained by Eq. (4D.1) are orthonormal vectors. It should be noted that the results Gram-Schmid orthogonalization is dependent on the order of vectors $\boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_K$ selected in the computation. Therefore, the orthonormal basis from the Gram-Schmid method is not unique.

The preceding Gram-Schmid method has poor numerical properties in that there is a severe loss of orthogonality among the generated $\boldsymbol{y}_k$ (Golub and Van Loan, 1989). The modified Gram-Schmid algorithm has the following steps:

1. $k = 0$.

2. Let $k = k + 1$ and $\boldsymbol{y}_k = \boldsymbol{x}_k$, for $k = 1$. Normalize vector $\boldsymbol{y}_k$ as $\boldsymbol{u}_k = \boldsymbol{y}_k/|\boldsymbol{y}_k|$.

3. For $k + 1 \leq j \leq K$, compute the vector of $\boldsymbol{x}_j$ projected on $\boldsymbol{u}_k$:

$$\tilde{\boldsymbol{y}}_j = (\boldsymbol{x}_j^t \boldsymbol{u}_k) \boldsymbol{u}_k$$

and the component of $\boldsymbol{x}_j$ orthogonal to $\boldsymbol{u}_i$ as

$$\boldsymbol{y}_j = \boldsymbol{x}_j - \tilde{\boldsymbol{y}}_j = \boldsymbol{x}_j - (\boldsymbol{x}_j^t \boldsymbol{u}_k) \boldsymbol{u}_k$$

4. Go to step 2 until $k = K$.

Problems

4.1 Refer to Sec. 1.6 for the central safety factor. Assuming that both R and L are independent normal random variables, show that the reliability index β is related to the central safety factor as

$$\beta = \frac{\mu_{SF} - 1}{\sqrt{\mu_{SF}^2 \, \Omega_R^2 + \Omega_L^2}}$$

in which Ω_x represents the coefficient of variation of random variable X.

4.2 Referring to Problem 4.1, the central safety factor can be expressed in terms of reliability index β as

$$\mu_{SF} = \frac{1 + \beta \sqrt{\Omega_R^2 + \Omega_L^2 - \beta^2 \Omega_R^2 \Omega_L^2}}{1 - \beta^2 \Omega_R^2}$$

4.3 Referring to Problem 4.1, how should the equation be modified if the resistance and load are correlated?

4.4 Refer to Sec. 1.6 for the characteristic safety factor. Let R_o be defined on the lower side of resistance distribution as $R_o = r_p$ with $P(R < r_p) = p$ (see Fig. 4P.1). Similarly, let L_o be defined on the upper side of load distribution with $L_o = \ell_{1-q}$. Consider that R and L are independent normal random variables. Show that characteristic safety factor SF_c is related to the central safety factor as

$$SF_c = \left(\frac{1 + z_p \Omega_R}{1 - z_q \Omega_L} \right) \mu_{SF}$$

in which $z_p = \Phi^{-1}(p)$.

4.5 Define the characteristic safety factor as the ratio of the median resistance to the median load as

$$\widetilde{SF} = \frac{r_{0.5}}{\ell_{0.5}} = \frac{\tilde{r}}{\tilde{\ell}}$$

where $\tilde{r} = r_{0.5} = F_R^{-1}(0.5)$ and $\tilde{\ell} = \ell_{0.5} = F_L^{-1}(0.5)$, with $F_R(\cdot)$ and $F_L(\cdot)$ being the CDFs of the resistance and load, respectively. Suppose that the resistance R
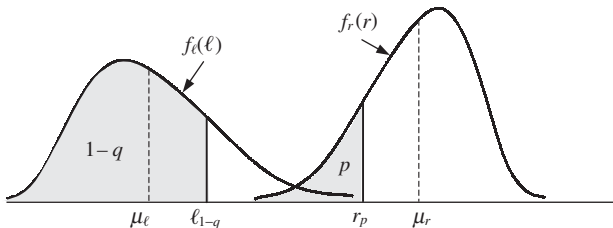
Figure 4P.1

and load L are independent lognormal random variables. Show that the central safety factor $\mu_{SF} = \mu_R/\mu_L$ is related to $\widetilde{SF}$ as

$$\mu_{SF} = \widetilde{SF} \times \sqrt{\frac{1 + \Omega_R^2}{1 + \Omega_L^2}}$$

4.6   Referring to Problem 4.4, show that for independent lognormal resistance and load, the following relation holds:

$$SF_c = \mu_{SF} \times \sqrt{\frac{1 + \Omega_R^2}{1 + \Omega_L^2}} \times \exp(z_p \Omega_R + z_q \Omega_L)$$

(Note: For small Ω_x, $\sigma_{\ln x} \approx \Omega_x$.)

4.7   Let $W(\boldsymbol{X}) = X_1 + X_2 - c$, in which X_1 and X_2 are independent stochastic variables with PDFs, $f_1(x_1)$ and $f_2(x_2)$, respectively. Show that the reliability can be computed as

$$p_s = \int_{-\infty}^{\infty} f_1(x_1)[1 - F_2(c - x_1)]dx_1$$

or

$$= \int_{-\infty}^{\infty} f_2(x_2)[1 - F_1(c - x_2)]dx_2$$

4.8   Suppose that the load and resistance are independent random variables and that each has an exponential PDF as

$$f_x(x) = \lambda_x \exp(-\lambda_x x) \qquad \text{for } x > 0$$

in which x can be the resistance R and load L. Show that the reliability is

$$p_s = \frac{\lambda_L}{\lambda_L + \lambda_R} = \frac{\mu_R}{\mu_R + \mu_L}$$

4.9   Show that the reliability for independently normally distributed resistance (with mean μ_R and standard deviation σ_R) and exponentially distributed load (with the mean $1/\lambda_L$) is

$$p_s = 1 - \Phi\left(-\frac{\mu_R}{\sigma_R}\right) - \exp\left[-\frac{1}{2}(2\mu_R\lambda_L - \lambda_L^2\sigma_R^2)\right] \times \left[1 - \Phi\left(-\frac{\mu_R - \lambda_L\sigma_R^2}{\sigma_R}\right)\right]$$

4.10   Suppose that the annual maximum flood in a given river reach has Gumbel distribution [Eq. (2.85a)] with mean μ_L and coefficient of variation Ω_L. Let the levee system be designed to have the mean capacity of $\mu_R = SF_c \times \ell_T$, with SF_c being the characteristic safety factor and T-year flow, respectively. For simplicity, assume that the levee conveyance capacity has a symmetric PDF, as shown in Fig. 4P.2. Derive the expression for the levee reliability assuming that flood magnitude and levee capacity are independent random variables.

4.11   Numerically solve Problem 4.10 using the following data:

$$\mu_L = 6000 \text{ ft}^3/\text{s} \qquad \Omega_L = 0.5 \qquad T = 100 \text{ years} \qquad \alpha = 0.15$$

for $SF_c = 1.0$ and 1.5.
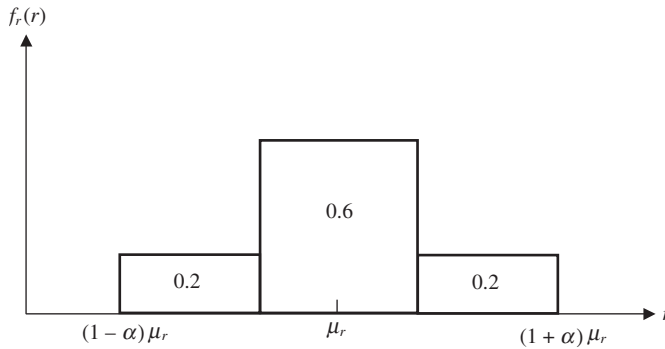
Figure 4P.2

4.12 Consider that load and resistance are independent uniform random variables with PDFs as

Load: $\qquad f_L(\ell) = 1/(\ell_2 - \ell_1) \qquad \ell_1 \le \ell \le \ell_2$
Resistance: $\qquad f_R(r) = 1/(r_2 - r_1) \qquad r_1 \le r \le r_2$

Furthermore, $\ell_1 < r_1 < \ell_2 < r_2$, as shown in Fig. 4P.3. Derive the expression for the failure probability.

4.13 Consider that load and resistance are independent random variables. The load has an extreme type I (max) distribution [Eq. (2.85a)], with the mean 1.0 and standard deviation of 0.3, whereas the resistance has a Weibull distribution [Eq. (2.89)], with mean 1.5 and standard deviation 0.5. Compute the failure probability using appropriate numerical integration technique.

4.14 Consider that the annual maximum flood has an extreme type I (max) distribution with the mean 1000 m^3/s and coefficient of 0.3. On the other hand, the levee capacity has a lognormal distribution with a mean of 1500 m^3/s and coefficient of variation of 0.2. Assume that flood and levee capacity are two independent random variables. Compute the failure probability that the levee will be overtopped using appropriate numerical integration technique.

4.15 Resolve Example 4.6 taking into account the fact that stochastic variables n and D are correlated with a correlation coefficient -0.75.
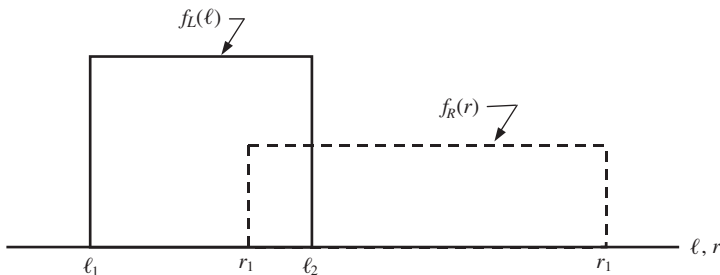
Figure 4P.3

4.16 The annual benefit and cost of a small hydropower project are random variables, and each has a Weibull distribution [see Eq. (2.89)] with the following distributional parameter values:

	α	ξ	β
Benefit	4.5422	60,000	266,000
Cost	3.7138	100,000	110,000

(a) Compute the mean and standard deviation of the annual benefit and cost.

(b) Assume that the annual benefit and cost are statistically independent. Find out the probability that the project is economically feasible, i.e., the annual benefit exceeds the annual cost.

4.17 Suppose that at a given dam site the flood flows and the spillway capacity follow triangular distributions, as shown in Fig. 4P.4. Use the direct integration method to calculate the reliability of the spillway to convey the flood flow (Mays and Tung, 1992).

4.18 The Hazen-Williams equation is used commonly to compute the head losses in a water distribution system, and it is written as

$$h_L = 4.728 \frac{L}{D^{4.87}} \left(\frac{Q}{C_{\text{HW}}} \right)^{1.852}$$

in which h_L is the head loss (in feet), L is the pipe length (in feet), D is the pipe diameter (in feet), Q is the flow rate (in ft³/s), and C_{HW} is the Hazen-Williams roughness coefficient.

Consider a water distribution system (see Fig. 4P.5) consisting of a storage tank serving as the source and a 1-ft-diameter cast iron pipe of 1 mile length leading to a user. The head elevation at the source is maintained at a constant level of 100 ft above the user. It is also known that at the user end the required pressure head is fixed at 20 psi (pounds per square inch) with variable demand on flow rate. Assume that the demand in flow rate is random, having a lognormal distribution with a mean of 3 ft³/s and a standard deviation of 0.3 ft³/s. Because of

Figure 4P.4

(2 ft diameter - 1 mile long)
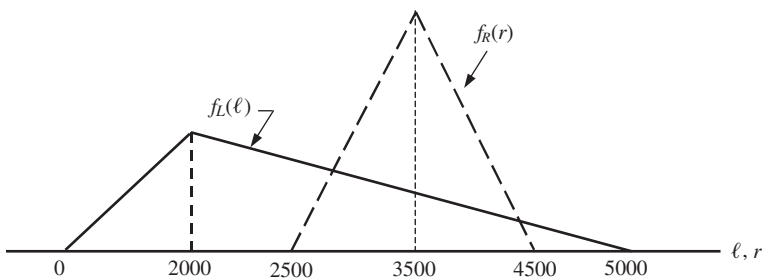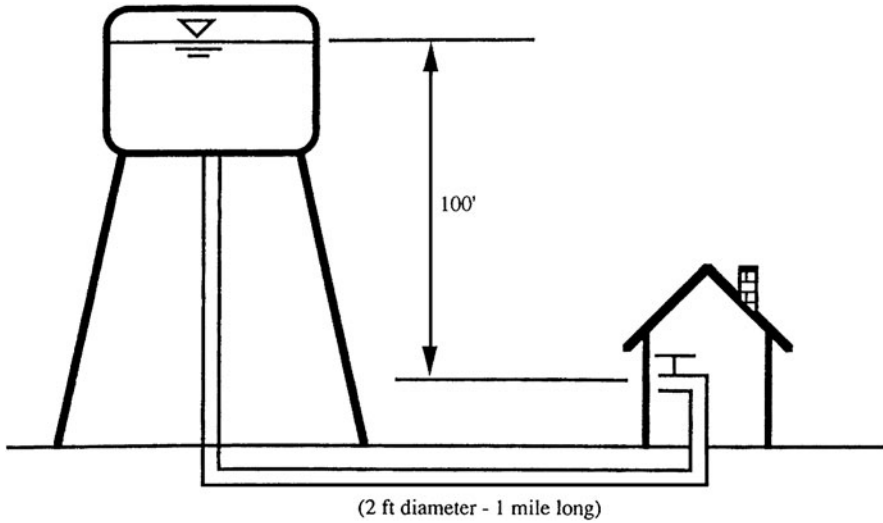
Figure 4P.5   (*After Mays and Tung, 1992*).

the uncertainty in pipe roughness and pipe diameter, the supply to the user is not certain. We know that the pipe has been installed for about 3 years. Therefore, our estimation of the pipe roughness in the Hazen-Williams equation is about 130 with some error of ± 20. Furthermore, knowing the manufacturing tolerance, the 1-ft pipe has an error of ± 0.05 ft. Assume that both the pipe diameter and Hazen-Williams' C_{HW} coefficient have lognormal distributions with means of 1 ft and 130 and standard deviations of 0.05 ft and 20, respectively. Using the MFOSM method, determine the reliability that the demand of the user can be satisfied (Mays and Tung, 1992).

4.19   In the design of storm sewer systems, the rational formula

$$Q_L = CiA$$

is used frequently, in which Q_L is the surface inflow resulting from a rainfall event of intensity i falling on the contributing drainage area of A, and C is the runoff coefficient. On the other hand, Manning's formula for full pipe flow, that is,

$$Q_C = 0.463n^{-1}S^{1/2}D^{8/3}$$

is used commonly to compute the flow-carrying capacity of storm sewers, in which D is the pipe diameter, n is the Manning's roughness, and S is pipe slope.

Consider that all the parameters in the rational formula and Manning's equation are independent random variables with their mean and standard deviation given below. Compute the reliability of a 36-in pipe using the MFOSM method (Mays and Tung, 1992).

Parameter	Mean	Std. Dev.	Distribution
C	0.825	0.057575	Uniform
i (in/h)	4.000	0.6	Gumbel
A (acres)	10.000	0.5	Normal
n	0.015	0.00083	Lognormal
D (ft)	3.000	0.03	Normal
S (ft/ft)	0.005	0.00082	Lognormal

4.20 In most locations, the point rainfall intensity can be expressed by the following empirical rainfall intensity-duration-frequency (IDF) formula:

$$i = \frac{aT^m}{b + t^c}$$

where i is the rainfall intensity (in in/h or mm/h), t is the storm duration (in minutes), T is the return period (in years), and a, m, b, and c are constants.

At Urbana, Illinois, the data analysis results in the following information about the coefficients in the preceding rainfall IDF equation:

Variable	Mean, μ	Coef. of Var. Ω	Distribution
a	120	0.10	Normal
b	27	0.10	Normal
c	1.00	0.05	Normal
m	0.175	0.08	Normal

Assuming independence among the IDF coefficients, analyze the uncertainty of the rainfall intensity for a 10-year, 24-minute storm. Furthermore, incorporate the derived information herein to Problem 4.19 to evaluate the sewer reliability.

4.21 The storm duration used in the IDF equation (see Problem 4.20) in general is equal to the time of concentration. One of the most commonly used in the Kirpich (Chow, 1964):

$$t_c = c_1 \left(\frac{L}{S^{0.5}} \right)^{c_2}$$

where t_c is the time of concentration (in minutes), L is the length of travel (in feet) from the most remote point on the drainage basin along the drainage channel to the basin outlet, S is the slope (in ft/ft) determined by the difference in elevation of the most remote point and that of the outlet divided by L, and c_1 and c_2 are coefficients.

Assume that c_1 and c_2 are the only random variables in the Kirpich formula with the following statistical features:

Parameter	Mean	Coeff. of Var.	Distribution
c_1	0.0078	0.3	Normal
c_1	0.77	0.2	Normal

(a) Determine the mean and standard deviation of t_c for the basin with $L = 1080$ ft and $S = 0.001$.
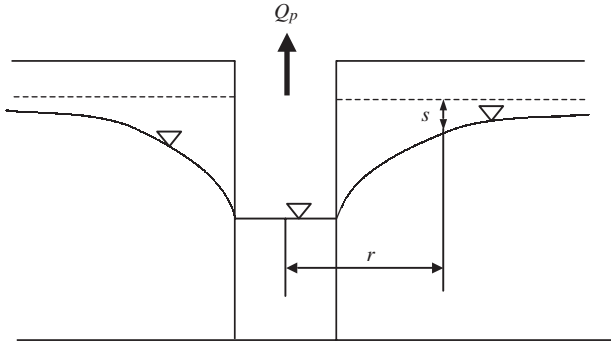
Figure 4P.6

    (b) Incorporate the uncertainty feature of t_c obtained in (a), and resolve the sewer reliability as Problem 4.20.

    (c) Compare the computed reliability with those from Problems 4.19 and 4.20.

4.22 Referring to Fig. 4P.6, the drawdown of a confined aquifer table owing to pumping can be estimated by the well-known Copper-Jacob equation:

$$s = \xi \frac{Q_p}{4\pi T}\left[-0.5772 - \ln\left(\frac{r^2 S}{4Tt}\right)\right]$$

in which ξ is the model correction factor accounting for the error of approximation, s is the drawdown (in meters), S is the storage coefficient, T is the transmissivity (in m²/day), Q_p is the pumping rate (in m³/day), and t is the elapse time (in days). Owing to the nonhomogeneity of geologic formation, the storage coefficient and transmissivity are in fact random variables. Furthermore, the model correction factor can be treated as a random variable. Given the following information about the stochastic variables in the Copper-Jacob equation, estimate the probability that the total drawdown will exceed 1.5 m under the condition of $Q_p = 1000$ m³/day, $r = 200$ m, and $t = 7$ days by the MFOSM method.

Variable	Mean μ	Coeff. of Var. Ω	Distribution
ξ	1.000	0.10	Normal
T (m²/day)	1000.0	0.15	Lognormal
S	0.0001	0.10	Lognormal

NOTE: $\rho(T, S) = -0.70$; $\rho(\xi, T) = 0.0$; $\rho(\xi, S) = 0.0$.

4.23 Referring to Fig. 4P.7, the time required for the original phreatic surface at h_o to have a drawdown s at a distance L from the toe of a cut slope can be approximated by (Nguyen and Chowdhury, 1985)

$$\frac{s}{h_o} = 1 - \text{erf}\left(\frac{L}{2\sqrt{Kh_o t/S}}\right)$$

where erf(x) is the *error function*, which is related to the standard normal CDF as erf(x) = $2\sqrt{2}[\Phi(x) - 0.5]$, K is the conductivity of the aquifer, S is the storage coefficient, and t is the drawdown time. From the slope stability viewpoint, it is
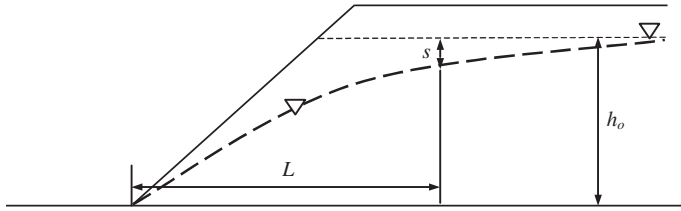
Figure 4P.7

required that further excavation can be made safely only when the drawdowns reach at least half the original phreatic head. Therefore, the drawdown time to reach $s/h_o = 0.5$ can be determined from the preceding equation as

$$t_d = \left(\frac{L}{2\xi}\right)^2 \frac{S}{Kh_o}$$

where $\xi = \text{erf}^{-1}(0.5) = 0.477$.

Consider that K and S are random variables having the following statistical properties:

Variable	Mean μ	Std. Dev. σ	Distribution
K (m/day)	0.1	0.01	Lognormal
S	0.05	0.005	Lognormal

NOTE: $\rho(K, S) = 0.5$.

Estimate the probability by the MFOSM method that the drawdown time t_d will be less than 40 days under the condition $L = 50$ m and $h_o = 30$ m.

4.24 The one-dimensional convective contaminant transport in steady flow through porous media can be expressed as (Ogata, 1970):

$$\frac{C(x,t)}{C_o} = \frac{1}{2}\text{erfc}\left[\frac{x - (q/n)t}{2\sqrt{a_\ell(q/n)t}}\right]$$

in which $C(x,t)$ is the concentration at point x and time t, C_o is the concentration of the incoming solute, x is the location along a one-dimensional line, q is the specific discharge, n is the porosity, a_ℓ is the *longitudinal dispersivity,* erfc is the *complimentary error function*, $\text{erfc}(x) = 1 - \text{erf}(x)$, and t is the time.

Assume that the specific discharge q, longitudinal dispersivity a_ℓ, and porosity n are random variables with the following statistical properties:

Variable	Mean μ	Std. Dev. σ	Distribution
q (m/day)	1.0	0.10	Lognormal
n	0.2	0.02	Normal
a_ℓ(m)	10.0	1.00	Lognormal

NOTE: $\rho(n, a_\ell) = 0.75$; zero for other pairs.

Estimate $P[C(x,t)/C_o > 0.5]$ for $x = 525$ m and $t = 100$ days by the MFOSM method.

4.25 Referring to the following Streeter-Phelps equation:

$$D_x = \frac{K_d L_0}{K_a - K_d} \left(e^{-K_d x/U} - e^{-K_a x/U}\right) + D_0 e^{-K_a x/U}$$

consider that the deoxygenation coefficient K_d, the reaeration coefficient K_a, the average stream velocity U, the initial dissolved oxygen DO, deficit concentrations D_0, and the initial in-stream BOD concentration L_0 are random variables. Assuming a saturated DO concentration of 8.48 mg/L, use the MFOSM method to estimate the probability that the in-stream DO concentration will be less than 4.0 mg/L at $x = 10$ miles downstream of the waste discharge point by adopting a lognormal distribution for the DO concentration with the following statistical properties for the involved random variables:

Variable	Mean μ	Std. Dev. σ	Distribution
K_d	0.60 L/day	0.060 L/day	Lognormal
K_a	0.76 L/day	0.076 L/day	Lognormal
U	1.2 ft/sec	0.012 ft/sec	Normal
D_0	1.60 mg/L	0.160 mg/L	Normal
L_0	6.75 mg/L	0.0675 mg/L	Normal

NOTE: $\rho(K_a, U) = 0.8$ and zero for all other pairs.

4.26 Referring to the Steeter-Phelps equation in Problem 4.25, determine the critical location associated with the maximum probability that the DO concentration is less than 4.0 mg/L using the statistical properties of involved random variables given in Problem 4.25. At any trial location, use the MFOSM method, along with the lognormal distribution for the random DO concentration, to compute the probability.

4.27 Develop a computer program for the Hasofer-Lind algorithm that can be used for problems involving correlated nonnormal random variables.

4.28 Develop a computer program for the Ang-Tang algorithm that can be used for problems involving correlated nonnormal random variables.

4.29 Solve Problem 4.18 by the AFOSM method. Also compute the sensitivity of the failure probability with respect to the stochastic variables. Compare the results with those obtained in Problem 4.18.

4.30 Solve Problem 4.21 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.21.

4.31 Solve Problem 4.22 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.22.

4.32 Solve Problem 4.23 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.23.

4.33 Solve Problem 4.24 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.24.

4.34   Solve Problem 4.25 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.25.

4.35   Solve Problem 4.26 by the AFOSM method considering all stochastic basic variables involved, and compare the results with those obtained in Problem 4.26.

4.36   Prove that Eq. (4.107) is true.

4.37   Show that under the condition of independent resistance and load, P_1 in Eq. (4.113) can be written as

$$P_1 = p_s - \int_0^{l_T} F_L(r) f_R(r) dr - \left(1 - \frac{1}{T}\right)[1 - F_R(l_T)]$$

4.38   Show that under the condition of independent resistance and load, P_2 in Eq. (4.114) can be written as

$$P_2 = \int_0^{l_T} F_L(r) f_R(r) dr + \left(1 - \frac{1}{T}\right)[1 - F_R(l_T)]$$

4.39   Assume that the annual maximum load and resistance are statistically independent normal random variables with the following properties:

Variable	Mean	Coefficient of variation
Load	1.0	0.25
Resistance	$SF \times \ell_{T=10-yr}$	0.15

Derive the reliability–safety factor–service life curves based on Eqs. (4.115) and (4.116).

4.40   Repeat Problem 4.39 by assuming that the annual maximum load and resistance are independent lognormal random variables.

4.41   Resolve Problem 4.39 by assuming that the resistance is a constant, that is, $r_* = SF \times \ell_{T=10-\text{yr}}$. Compare the reliability–safety factor–service life curves with those obtained in Problem 4.39.

References

Abramowitz, M., and Stegun, I. A. (eds.) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed., Dover Publications, New York.

Ang, A. H. S. (1973). Structural risk analysis and reliability-based design, *Journal of Structural Engineering*, ASCE, 99(9):1891–1910.

Ang, A. H. S., and Cornell, C. A. (1974). Reliability bases of structural safety and design, *Journal of Structural Engineering*, ASCE, 100(9):1755–1769.

Ang, A. H. S., and Tang, W. H. (1984). *Probability Concepts in Engineering Planning and Design,* Vol. II: *Decision, Risk, and Reliability*, John Wiley & Sons, New York.

Bechteler, W., and Maurer, M. (1992). Reliability theory applied to sediment transport formulae, in *Proceedings*, 5th International Symposium on River Sedimentation, Karlsruhe, Germany, pp. 311–317.

Berthouex, P. M. (1975). Modeling concepts considering process performance variability, and uncertainty, in *Mathematical Modeling for Water Pollution Control Process*, ed. by T. M. Keinath and M. P. Wanielista, Ann Arbor Science, Ann Arbor, MI, 405–439.

Bodo, B., and Unny, T. E. (1976). Model uncertainty in flood frequency-analysis and frequency-based design, *Water Resources Research*, AGU, 12(6):1109–1117.

Breitung, K. (1984). Asymptotic approximations for multinormal integrals, *Journal of Engineering Mechanics*, ASCE. 110(3):357–366.

Breitung, K. (1993). *Asymptotic Approximations for Probability Integrals*, Springer-Verlag, New York.

Cesare, M. A. (1991). First-order analysis of open-channel flow, *Journal of Hydraulic Engineering*, ASCE, 117(2):242–247.

Chang, C. H. (1994). Incorporating information incorporating non-normal marginal distributions in uncertainty analysis of hydrosystems, Ph.D. Thesis, Civil Enginerring Department, National Chiao-Tung University, Hsinchu, Taiwan.

Cheng, S. T., Yen, B. C., and Tang, W. H. (1982). Overtopping risk for an existing dam, *Hydraulic Engineering Series*, No. 37, Department of Civil Engineering, University of Illinois at Urbana-Champaign, IL.

Cheng, S. T., Yen, B. C., and Tang, W. H. (1986a). Wind-induced overtopping risk of dams, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 48–58.

Cheng, S. T., Yen, B. C., and Tang, W. H. (1986b). Sensitivity of risk evaluation to coefficient of variation, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 266–273.

Cheng, S. T., Yen, B. C., and Tang, W. H. (1993). Stochastic risk modeling of dam overtopping, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, ASCE, New York, pp. 123–132.

Chow, V. T. (ed.) (1964). *Handbook of Applied Hydrology*, McGraw-Hill, New York.

CIRIA (Construction Industry Research and Information Association) 1997. "Rationalization of safety and serviceability factors in structural codes," *CIRIA Report* No. 63, London.

Clark, R. T. (1998). *Stochastic Processes for Water Scientists: Development and Application*, John Wiley and Sons, New York.

Cornell, C. A. (1969). A probability-based structural code, *Journal of American Concrete Institute*, 66(12):974–985.

Der Kiureghian, A. (1989). Measures of structural safety under imperfect states of knowledge, *Journal of Structural Engineering*, ASCE, 115(5):1119–1140.

Der Kiureghian, A., and Liu, P. L. (1985). Structural reliability under incomplete probability information, *Journal of Engineering Mechanics*, ASCE, 112(1):85–104.

Der Kiureghian, A., Lin, H. Z., and Hwang, S. J. (1987). Second-order reliability approximations, *Journal of Engineering Mechanics*, ASCE, 113(8):1208–1225.

Der Kiureghian, A., and De Stefano. M. (1991). Efficient algorithm for second-order reliability analysis, *Journal of Engineering Mechanics*, ASCE, 117(12):2904–2923.

Ditlevsen, O. (1973). Structural reliability and the invariance problem, Research Report No. 22, Solid Mechanics Division, University of Waterloo, Waterloo, Canada.

Ditlevsen, O. (1979). Generalized second-order reliability index, *Journal of Structural Mechanics*, 7(4):435–451.

Ditlevsen, O. (1981). Principle of normal tail approximation, *Journal of Engineering Mechanics*, 107(6):1191–1208.

Ditlevsen, O. (1984). Taylor expansion of series system reliability, *Journal of Engineering Mechanics*, ASCE, 110(2):293–307.

Dolinski, K. (1983). First-order second-moment approximation in reliability of system: Critical review and alternative approach, *Structural Safety*, 1:211–213.

Draper, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society, Series B*, 57(1):45–70.

Easa, S. M. (1992). Probabilistic design of open drainage channels, *Journal of Irrigation Engineering*, ASCE, 118(6):868–881.

Farebrother, R. W. (1980). Algorithm AS 153: Pan's procedure for the tail probabilities of the durbin-Watson statistic, *Applied Statistics*, 29:224–227.

Farebrother, R. W. (1984a). Algorithm AS 204: The distribution of a positive linear combination of χ^2 random variables, *Applied Statistics*, 33:332–339.

Farebrother, R. W. (1984b). A remark on algorithms AS 106, AS 153, and AS 155: The distribution of a linear combination of χ^2 random variables, *Applied Statistics*, 33:366–369.

Fiessler, B., Neumann, H. J., and Rackwitz, R. (1979). Quadratic limit states in structural reliability, *Journal of Engineering Mechanics*, ASCE, 105(4):661–676.

Golub, G. H., and Van Loan, C. F. (1989). *Matrix Computations*, 2d ed., Johns Hopkins University Press, Baltimore.

Graybill, F. A. (1983). Matrices with Application in Statistics, Wadsworth Publishing Company, Inc., Belmond, CA.

Green, P. E. (1976). *Mathematical Tools for Applied Multivariate Analysis*, Academic Press, New York.

Gui, S. X., Zhang, R., and Wu, J. Q. (1998). Simplified dynamic reliability models for hydraulic structures, *Journal of Hydraulic Engineering*, ASCE, 124(3):329–333.

Han, K-Y, Kim, S-H and Bae, D-H (2001). Stochastic water quality analysis using reliability method. Journal of the American Water Resources Association, 37(3):695–708.

Hasofer, A. M., and Lind, N. C. (1974). Exact and invariant second-moment code format, *Journal of Engineering Mechanics*, ASCE, 100(1):111–121.

Hohenbichler, M., and Rackwitz, R. (1988). Improvement of second-order reliability estimates by importance sampling, *Journal of Engineering Mechanics*, ASCE, 114(12):2195–2199.

Huang, K. Z. (1986). Reliability analysis of hydraulic design of open channel, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 59–65

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48(3):419–426.

Jang, Y. S, Sitar, N., and Der Kiureghian, A. (1990). Reliability approach to probabilistic modelling of contaminant transport, Report No. UCB/GT-90/3, Department of Civil Engineering, University of California, Berkeley, CA.

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions,* 2d ed., John Wiley and Sons, New York.

Kapur, K. C., and Lamberson, L. R. (1977). *Reliability in Engineering Designs*, John Wiley and Sons, New York.

Lasdon, L. S., Waren, A. D., and Ratner, M. W. (1982). *GRG2 User's Guide*, Department of General Business, University of Texas, Austin.

Lee, H. L., and Mays, L. W. (1986). Hydraulic uncertainties in flood levee capacity, *Journal of Hydraulic Engineering*, ASCE, 112(10):928–934.

Lind, N. C. (1977). "Formulation of probabilistic design, *Journal of Engineering Mechanics*, ASCE, 103(2):273–284.

Liu, P. L., and Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances, *Probabilistic Engineering Mechanics*, 1(2):105–112.

Liu, P. L., and Der Kiureghian, A. (1991). Optimization algorithms for structural reliability, *Structural Safety*, 9:161–177.

Low, B. K., and Tang, W. H. (1997). Efficient reliability evaluation using spreadsheet, *Journal of Engineering Mechanics*, ASCE, 123(7):350–362.

Madsen, H. O., Krenk, S., and Lind, N. C. (1986). *Methods of Structural Safety*, Prentice-Hall, Englewood Cliffs, NJ.

Maier, H. R., Lence, B. J., Tolson, B. A., and Foschi, R. O. (2001). First-order reliability method for estimating reliability, vulnerability, and resilience, *Water Resources Research*, 37(3):779–790.

Mays, L. W., and Tung, Y. K. (1992). *Hydrosystems Engineering and Management*, McGraw-Hill, New York.

McBean, E. A., Penel, J., and Siu, K. L. (1984). Uncertainty analysis of a delineated floodplain, *Canadian Journal of Civil Engineers*, 11:387–395.

Melchers, R. E. (1999) *Structural Reliability: Analysis and Prediction*, 2d ed., John Wiley and Sons, New York.

Melching, C. S. (1992). An improved first-order reliability approach for assessing uncertainties in hydrologic modeling, *Journal of Hydrology*, 132:157–177.

Melching, C. S., and Yen, B. C. (1986). Slope influence on storm sewer risk, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 66–78.

Melching, C. S., Wenzel, H. G., Jr., and Yen, B. C. (1990). A reliability estimation in modeling watershed runoff with uncertainties, *Water Resources Research*, 26:2275–2286.

Melching, C. S., and Anmangandla, S. (1992). Improved first-order uncertainty method for water quality modeling, *Journal of Environmental Engineering*, ASCE, 118(5):791–805.

Melching, C. S., and Yoon, C. G. (1996). "Key sources of uncertainty in QUALE model of Passaic River," Journal of Water Resources Planning and Management, ASCE, 122(2):105–113.

Nataf, A. (1962). Dtermination des distributions de probabilits dont les marges sont donnes, *Computes Rendus de l'Acadmie des Sciences*, Paris, 255:42–43.

Naess, A. (1987). Bounding approximations to some quadratic limit states, *Journal of Engineering Mechanics*, ASCE, 113(10):1474–1492.

Nguyen, V. U., and Chowdhury, R. N. (1985). Simulation for risk analysis with correlated variables, *Geotechnique*, 35(1):47–58.

Ogata, A. (1970). Theory of dispersion in granular medium, U.S. Geological Survey, Professional Paper, 411.

Press, S. J. (1966). Linear combinations of noncentral χ^2 variates, *Annals of Mathematical Statistics*, 37:480–487.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989). *Numerical Recipes: The Art of Scientific Computing (Fortran Version)*, Cambridge University Press, New York.

Rackwitz, R. (1976). Practical probabilistic approach to design, *Bulletin 112*, Comite European du Beton, Paris, France.

Rackwitz, R., and Fiessler, B. (1978). Structural reliability under combined random load sequence, *Computers and Structures*, 9:489–494.

Rice, S. O. (1980). Distribution of quadratic forms in normal random variables: Evaluation by numerical integration, *Journal of Scientific Statistical Computation*, SIAM, 1(4):438–448.

Rosenblatt, M. (1952). Remarks on a multivariate transformation, *Annals of Mathematical Statistics*, 23:470–472.

Ruben, H. (1962). Probability content of regions under spherical normal distribution, part IV, *Annals of Mathematical Statistics*, 33:542–570.

Ruben, H. (1963). A new result of the distribution of quadratic forms, *Annals of Mathematical Statistics*, 34:1582–1584.

Shah, B. K. (1986). The distribution of positive definite quadratic forms, in *Selected Tables in Mathematical Statistics*, Vol. 10, American Mathematical Society, Providence, RD.

Shinozuka, M. (1983). Basic analysis of structural safety, *Journal of Structural Engineering*, ASCE, 109(3):721–740.

Singh, S., and Melching, C. S. (1993). "Importance of hydraulic model uncertainty in flood-stage estimation," In: Hydraulic Engineering '93, Proceedings, 1993 ASCE National Conference on Hydraulic Engineering, edited by H.-W. Shen, S.-T. Su, and F. Wen, Vol. 2:1939–1944.

Sitar, N., Cawfiled, J. D., and Der Kiureghian, A. (1987). First-order reliability approach to stochastic analysis of subsurface flow and contaminant transport, *Water Resources Research*, AGU, 23(5):794–804.

Tang, W. H., and Yen, B. C. (1972). Hydrologic and hydraulic design under uncertainties, in *Proceedings*, International Symposium on Uncertainties in Hydrologic and Water Resources Systems, Tucson, AZ, 2, 868–882; 3, 1640–1641.

Tang, W. H., Mays, L. W., and Yen, B. C. (1975). Optimal risk-based design of storm sewer networks, *Journal of Environmental Engineering*, ASCE, 103(3):381–398.

Todorovic, P., and Yevjevich. V. (1969) Stochastic process of precipitation, Hydrology Paper No. 35, Colorado State University, Fort Collins, CO.

Tung Y. K. (1985). Models for evaluating flow conveyance reliability of hydraulic structures, *Water Resources Research*, AGU, 21(10):1463–1468.

Tung, Y. K. (1990). Evaluating the probability of violating dissolved oxygen standard, *Journal of Ecological Modeling*, 51:193–204.

Tung, Y. K., and Mays. L. W. (1980). Optimal risk-based design of hydraulic structures, Technical Report, CRWR-171, Center for Research in Water Resources, University of Texas, Austin.

Tung, Y. K., and Mays, L. W. (1981). Risk models for levee design, *Water Resources Research*, AGU, 17(4):833–841.

Tung, Y. K., and Yen, B. C. (2005). Hydrosystems Engineering Uncertainty Analysis. Mc-Graw Hill Book Company, New York.

Tvedt, L. (1983). Two second-order approximations to the failure probability, Veritas Report RDIV/20-004-83, Det norske Veritas, Oslo, Norway.

Tvedt. L. (1990). Distribution of quadratic forms in normal space: Application to structural reliability, *Journal of Engineering Mechanics*, ASCE, 116(6):1183–1197.

Vrijling, J. K. (1987). Probabilistic design of water retaining structures, in *Engineering Reliability and Risk in Water Resources*, ed. by L. Duckstein and E. J. Plate, Martinus Nijhoff, Dordrecht, The Netherlands, pp. 115–134.

Vrijling, J. K. (1993). Development in probabilistic design of flood defenses in the Netherlands, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, ASCE, New York, pp. 133–178.

Wen, Y. K. (1987). Approximate methods for nonlinear time-variant reliability analysis, *Journal of Engineering Mechanics*, ASCE, 113(12):1826–1839.

Wood, D. J. (1991). *Comprehensive Computer Modeling of Pipe Distribution Networks*, Civil Engineering Software Center, College of Engineering, University of Kentucky, Lexington, KY.

Wood, E. F., and Rodriguez-Iturbe, I. (1975a). Bayesian inference and decision making for extreme hydrologic events, *Water Resources Research*, AGU, 11(4):533–543.

Wood, E. F., and Rodriguez-Iturbe, I. (1975b). A Bayesian approach to analyzing uncertainty among flood frequency models, *Water Resources Research*, AGU, 11(6):839–843.

Yen, B. C. (1970). Risks in hydrologic design of engineering projects, *Journal of Hydraulic Division*, ASCE., 96(HY4):959–966.

Yen, B. C., and Ang, A. H. S. (1971). Risk analysis in design of hydraulic projects, in *Stochastic Hydraulics*, ed. by C. L. Chiu, University of Pittsburgh, Pittsburgh, PA, pp. 697–701.

Yen, B. C., and Tang, W. H. (1976). Risk–safety factor relation for storm sewer design, *Journal of Environmental Engineering*, ASCE, 102(2):509–516.

Yen, B. C., Wenzel, H. G., Jr., Mays, L. W., and Tang, W. H. (1976). Advanced methodologies for design of storm sewer systems, Research Report, No. 112, Water Resources Center, University of Illinois at Urbana-Champaign, IL.

Yen, B. C., Cheng, S. T., and Tang, W. H. (1980). Reliability of hydraulic design of culverts, *Proceedings*, International Conference on Water Resources Development, IAHR Asian Pacific Division Second Congress, Taipei, Taiwan, 2:991–1001.

Yen, B. C., Cheng, S. T., and Melching, C. S. (1986). First-order reliability analysis, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, CO, pp. 1–36.

Yen, B. C., and Melching, C. S. (1991). Reliability analysis methods for sediment problems, in *Proceedings*, 5th Federal Interagency Sediment Conference, Las Vegas, 2:9.1–9.8.

Young, D. M. and Gregory, R. T. (1973). A Survey of Numerical Mathematics—Vol. 2, Dover Publications, New York.

Zelenhasic, E. (1970). Theoretical probability distribution for flood peaks, Hydrology Paper, No. 42, Colorado State University, Fort Collins, CO.

Time-to-Failure Analysis

5.1 Basic Concept

In preceding chapters, evaluations of reliability were based on analysis of the interaction between loads on the system and the resistance of the system. A system would perform its intended function satisfactorily within a specified time period if its capacity exceeds the load. Instead of considering detailed interactions of resistance and load over time, in a *time-to-failure* (TTF) *analysis*, a system or its components can be treated as a black box or a lumped-parameter system, and their performances are observed over time. This reduces the reliability analysis to a one-dimensional problem involving time as the only variable describable by the TTF of a system or a component of the system. The *time-to-failure* is an important parameter in reliability analysis, representing the length of time during which a component or system under consideration remains operational. The TTF generally is affected by inherent, environmental, and operational factors. The *inherent factors* involve the strength of the materials, manufacturing process, and the quality control. The *environmental factors* include such things as temperature, humidity, air quality, and others. The *operational factors* include external load conditions, intensity and frequency of use, and technical capability of users. In a real-life setting, the elements of the factors affecting the TTF of a component are often subject to uncertainty. Therefore, the TTF is a random variable.

In some situations, other physical scale measures, such as distance or length, may be appropriate for system performance evaluation. For example, the reliability of an automobile could be evaluated over its traveling distance, or the pipe break probability owing to the internal pressure or external loads from gravity or soil could be evaluated based on the length of the pipe. Therefore, the notion of "time" should be regarded in a more general sense.

TTF analysis is particularly suitable for assessing the reliability of systems and/or components that are repairable. The primary objectives of the reliability analysis techniques described in the preceding chapters were the probability of

the first failure of a system subject to external loads. In case the system fails, how and when the system is repaired or restored are of little importance. Hence such techniques are often used to evaluate the reliability of nonrepairable systems or the failure probability when systems are subject to extraordinary events. For a system that is repairable after its failure, the time period it would take to have it repaired back to the operational state, called the *time-to-repair* or *restore* (TTR), is uncertain.

Several factors affect the value of the TTR and include personal, conditional, and environmental factors (Knezevic, 1993). *Personal factors* are those represented by the skill, experience, training, physical ability, responsibility, and other similar characteristics of the personnel involved in the repair. The *conditional factors* include the operating environment and the extent of the failure. The *environmental factors* are humidity, temperature, lighting, noise, time of day, and similar factors affecting the maintenance crew during the repair. Again, owing to the inherently uncertain nature of the many elements, the TTR is a random variable.

For a repairable system or component, its service life can be extended indefinitely if repair work can restore the system like new. Intuitively, the probability of a repairable system available for service is greater than that of a nonrepairable system. Consider two identical systems: One is to be repaired after its failure, and the other is not to be repaired. The difference in probability that a system would be found in operating condition at a given instance would become wider as the age of the two systems increased. This chapter focuses on the characteristics of failure, repair, and availability of repairable systems by TTF analysis.

5.2   Failure Characteristics

Any system will fail eventually; it is just a matter of time. Owing to the presence of many uncertainties that affect the operation of a physical system, the time the system fails to perform its intended function satisfactorily is random.

5.2.1   Failure density function

The probability distribution governing the time occurrence of failure is called the *failure density function*. This failure density function serves as the common thread in the reliability assessments by TTF analysis. Referring to Fig. 5.1, the reliability of a system or a component within a specified time interval $(0, t]$, can be expressed, assuming that the system is operational initially at $t = 0$, as

$$p_s(t) = P(\text{TTF} > t) = \int_t^\infty f_t(\tau)\, d\tau \qquad (5.1a)$$

in which the TTF is a random variable having $f_t(t)$ as the failure density function. The reliability $p_s(t)$ represents the probability that the system experiences
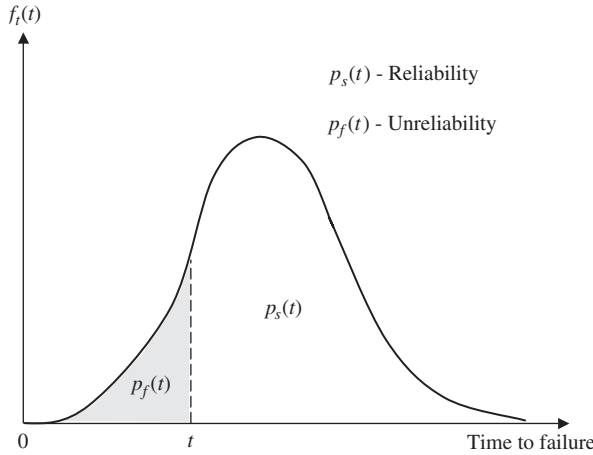
Figure 5.1 Schematic diagram of reliability and unreliability in the time-to-failure analysis.

no failure within $(0, t]$. The failure probability, or *unreliability*, can be expressed as

$$p_f(t) = P(\text{TTF} \leq t) = 1 - p_s(t) = \int_0^t f_t(\tau)\, d\tau \qquad (5.1b)$$

Note that unreliability $p_f(t)$ is the probability that a component or a system would experience its first failure within the time interval $(0, t]$. As can be seen from Fig. 5.1, as the age of system t increases, the reliability $p_s(t)$ decreases, whereas the unreliability $p_f(t)$ increases. Conversely, the failure density function can be obtained from the reliability or unreliability as

$$f_t(t) = -\frac{d\,[p_s(t)]}{dt} = \frac{d\,[p_f(t)]}{dt} \qquad (5.2)$$

The TTF is a continuous, nonnegative random variable by nature. Many continuous univariate distribution functions described in Sec. 2.6 are appropriate for modeling the stochastic nature of the TTF. Among them, the exponential distribution, Eq. (2.79), perhaps is the most widely used. Besides its mathematical simplicity, the exponential distribution has been found, both phenomenologically and empirically, to describe the TTF distribution adequately for components, equipment, and systems involving components with a mixture of life distributions. Table 5.1 lists some frequently used failure density functions and their distributional properties.

5.2.2 Failure rate and hazard function

The *failure rate* is defined as the number of failures occurring per unit time in a time interval $(t, t + \Delta t]$ per unit of the remaining population in operation at

TABLE 5.1   Selected Time-to-Failure Probability Distributions and Their Properties

Distribution	Failure density function $f_t(t)$	Reliability $p_s(t)$	Hazard function $h(t)$	Mean time to failure
Normal	$\dfrac{1}{\sqrt{2\pi}\,\sigma_t}\exp\left[-\dfrac{1}{2}\left(\dfrac{t-\mu_t}{\sigma_t}\right)^2\right]$	$\int_t^\infty f_t(\tau)\,d\tau = 1 - \Phi\left(\dfrac{t-\mu_t}{\sigma_t}\right)$	$\dfrac{f_t(t)}{p_s(t)}$	μ_t
Lognormal	$\dfrac{1}{\sqrt{2\pi}\,t\sigma_{t'}}\exp\left[-\dfrac{1}{2}\left(\dfrac{t'-\mu_{t'}}{\sigma_{t'}}\right)^2\right]$	$\int_t^\infty f_t(\tau)\,d\tau = 1 - \Phi\left(\dfrac{t'-\mu_{t'}}{\sigma_{t'}}\right)$	$\dfrac{f_t(t)}{p_s(t)}$	$\exp[\mu_{t'} + 0.5\sigma_{x'}]$ where $t' = \ln(t)$
Exponential	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ	$1/\lambda$
Rayleigh	$\dfrac{t}{\beta^2}\exp\left[-\dfrac{1}{2}\left(\dfrac{t}{\beta}\right)^2\right]$	$\exp\left[-\dfrac{1}{2}\left(\dfrac{t}{\beta}\right)^2\right]$	$\dfrac{t}{\beta^2}$	1.253β
Gamma (two-parameter)	$\dfrac{\beta}{\Gamma(\alpha)}(\beta t)^{\alpha-1}e^{-\beta t}$	$\int_t^\infty f_t(\tau)\,d\tau$	$\dfrac{f_t(t)}{p_s(t)}$	α/β
Gumbel (max)	$\dfrac{1}{\beta}\exp\left\{-\left(\dfrac{t-t_o}{\beta}\right) - \exp\left[-\left(\dfrac{t-t_o}{\beta}\right)\right]\right\}$	$1 - \exp\left\{-\exp\left[-\left(\dfrac{t-t_o}{\beta}\right)\right]\right\}$	$\dfrac{f_t(t)}{p_s(t)}$	$t_o + 0.577\beta$
Weibull	$\dfrac{\alpha}{\beta}\left(\dfrac{t-t_o}{\beta}\right)^{\alpha-1}\exp\left[-\left(\dfrac{t-t_o}{\beta}\right)^\alpha\right]$	$\exp\left[-\left(\dfrac{t-t_o}{\beta}\right)^\alpha\right]$	$\dfrac{\alpha(t-t_o)^{\alpha-1}}{\beta^\alpha}$	$t_o + \beta\Gamma\left(1+\dfrac{1}{\alpha}\right)$
Uniform	$\dfrac{1}{b-a},\; a \le t \le b$	$\dfrac{b-t}{b-a}$	$\dfrac{1}{b-t}$	$\dfrac{a+b}{2}$

time t. Consider that a system consists of N identical components. The number of failed components in $(t, t + \Delta t]$, $N_F(\Delta t)$, is

$$N_F(\Delta t) = N \times p_f(t + \Delta t) - N \times p_f(t) = N\left[p_f(t + \Delta t) - p_f(t)\right]$$

and the remaining number of operational components at time t is

$$N(t) = N \times p_s(t)$$

Then, according to the preceding definition of the failure rate, the *instantaneous failure rate* $h(t)$ can be obtained as

$$
\begin{aligned}
h(t) &= \lim_{\Delta t \to 0}\left[\frac{N_F(\Delta t)/\Delta t}{N(t)}\right] = \lim_{\Delta t \to 0}\left[\frac{N \times p_f(t + \Delta t) - N \times p_f(t)}{N(t) \times \Delta t}\right] \\
&= \frac{1}{p_s(t)}\lim_{\Delta t \to 0}\left[\frac{p_f(t + \Delta t) - p_f(t)}{\Delta t}\right] \\
&= \frac{1}{p_s(t)}\frac{d\left[p_f(t)\right]}{dt} \\
&= \frac{f_t(t)}{p_s(t)}
\end{aligned}
\tag{5.3}
$$

This instantaneous failure rate is also called the *hazard function* or *force-of-mortality function* (Pieruschka, 1963). Therefore, the hazard function indicates the change in the failure rate over the operating life of a component. The
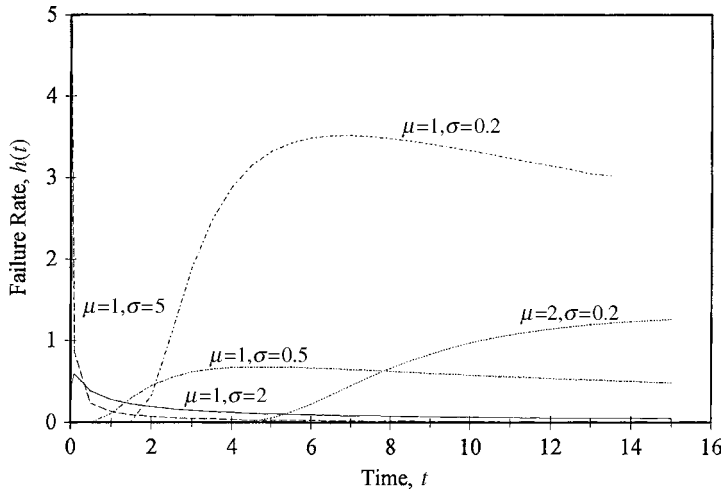
Figure 5.2   Failure rate for lognormal failure density function.

hazard functions for some commonly used failure density functions are given in Table 5.1. Figures 5.2 through 5.6 show the failure rates with respect to time for various failure density functions.

Alternatively, the meaning of the hazard function can be seen from

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{1}{\Delta t} \times \frac{p_f(t + \Delta t) - p_f(t)}{p_s(t)} \right] \qquad (5.4)$$

in which the term $[p_f(t + \Delta t) - p_f(t)]/p_s(t)$ is the *conditional failure probability* in $(t, t + \Delta t]$, given that the system has survived up to time t. Hence the
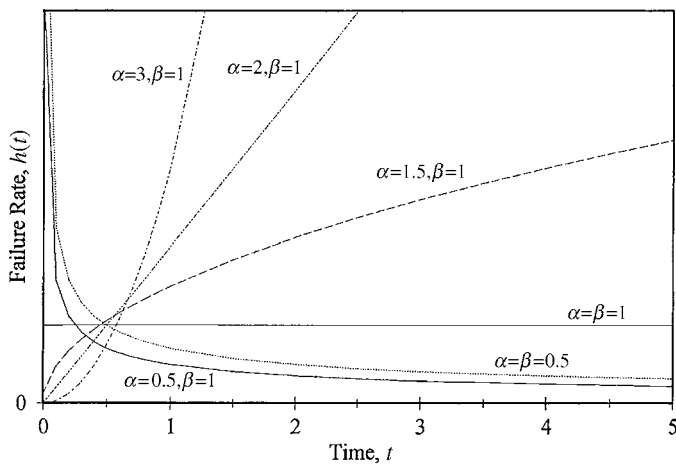
Figure 5.3   Failure rate for Weibull failure density function with $t_o = 0$.
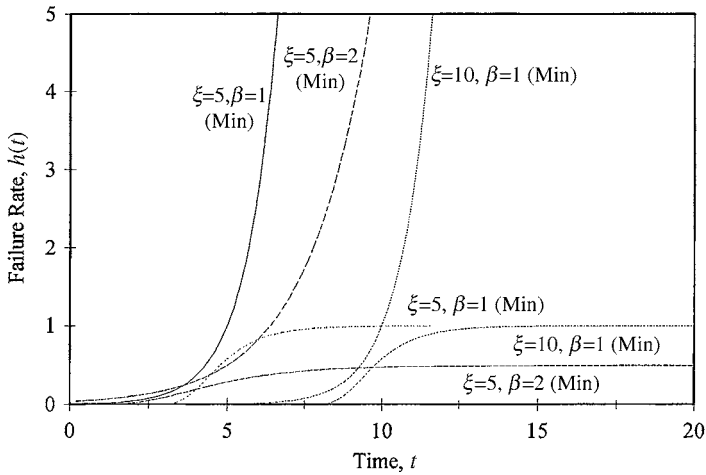
Figure 5.4 Failure rate for Gumbel failure density function.

hazard function can be interpreted as the time rate of change of the conditional failure probability for a system given that has survived up to time t. It is important to differentiate the meanings of the two quantities $f_t(t)\,dt$ and $h(t)\,dt$, with the former representing the probability that a component would experience failure during the time interval $(t, t + dt]$—it is unconditional—whereas the latter, $h(t)\,dt$, is the probability that a component would fail during the time interval $(t, t + \Delta t]$—conditional on the fact that the component has been in an operational state up to time instant t.
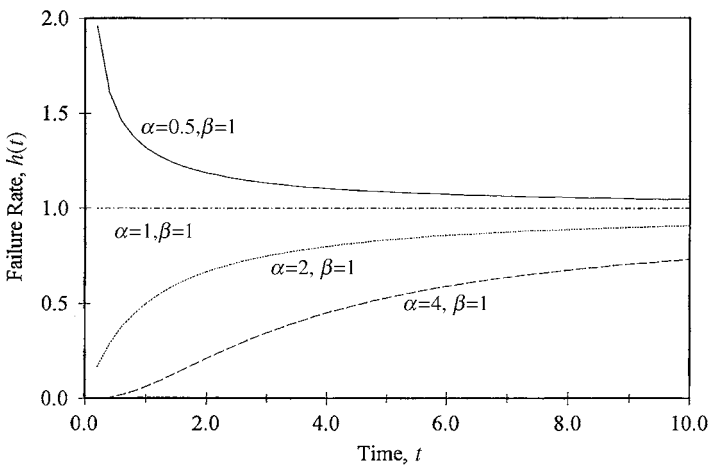
Figure 5.5 Failure rate for two-parameter gamma failure density function.
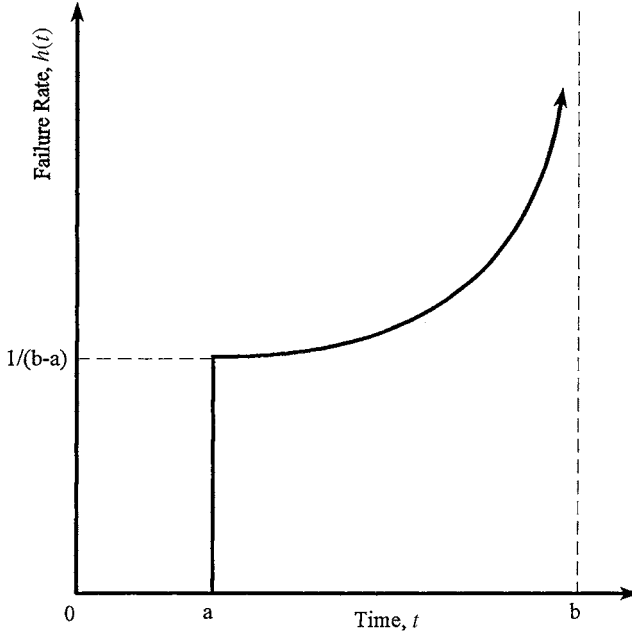
Figure 5.6 Failure rate for uniform failure density function.

5.2.3   Cumulative hazard function and average failure rate

Similar to the cumulative distribution function (CDF), the *cumulative hazard function* can be obtained from integrating the instantaneous hazard function $h(t)$ over time as

$$H(t) = \int_0^t h(t)\, dt \tag{5.5}$$

Referring to Eq. (5.3), the hazard function can be written as

$$h(t) = \frac{1}{p_s(t)} \frac{d\,[p_f(t)]}{dt} = -\frac{1}{p_s(t)} \frac{d\,[p_s(t)]}{dt} \tag{5.6}$$

Multiplying dt on both sides of Eq. (5.6) and integrating them over time yields

$$H(t) = \int_0^t h(t)\, dt = \int_0^t \frac{-d\,[p_s(t)]}{p_s(t)} = -\ln[p_s(t)]_0^t = -\ln[p_s(t)] \tag{5.7}$$

under the initial condition of $p_s(0) = 1$.

Unlike the CDF, interpretation of the cumulative hazard function is not simple and intuitive. However, Eq. (5.7) shows that the cumulative hazard function is equal to $\ln[1/p_s(t)]$. This identity relationship is especially useful in the statistical analysis of reliability data because the plot of the sample estimation

of $1/p_s(t)$ versus time on semi-log paper reveals the behavior of the cumulative hazard function. Then the slope of $\ln[1/p_s(t)]$ yields directly the hazard function $h(t)$. Numerical examples showing the analysis of reliability data can be found elsewhere (O'Connor, 1981, pp. 58–87; Tobias and Trindade, 1995, pp. 135–160).

Since the hazard function $h(t)$ varies over time, it is sometimes practical to use a single average value that is representative of the failure rate over a time interval of interest. The *averaged failure rate* (AFR) in the time interval $[t_1, t_2]$ can be defined as

$$\text{AFR}(t_1, t_2) = \frac{\int_{t_1}^{t_2} h(t)\, dt}{t_2 - t_1} = \frac{H(t_2) - H(t_1)}{t_2 - t_1} = \frac{\ln[p_s(t_1)] - \ln[p_s(t_2)]}{t_2 - t_1} \tag{5.8}$$

Therefore, the averaged failure rate of a component or system from the beginning over a time period $(0, t]$ can be computed as

$$\text{AFR}(0, t) = \frac{-\ln[p_s(t)]}{t} \tag{5.9}$$

The failure rate, in general, has the conventional unit of number of failures per unit time. For a component with a high reliability, the failure rate will be too small for the conventional unit to be appropriate. Therefore, the scale frequently used for the failure rate is the *percent per thousand hours* (*%K*) (Ramakumar, 1993; Tobias and Trindade, 1995). One percent per thousand hours means an expected rate of one failure for each 100 units operating 1000 hours. Another scale for even higher-reliability components is *parts per million per thousand hours* (*PPM/K*), which means the expected number of failures out of one million components operating for 1000 hours. The PPM/K is also called the *failures in time* (*FIT*). If the failure rate $h(t)$ has the scale of number of failures per hour, it is related to the %K and PPM/K as follows:

$$1\%\text{K} = 10^5 \times h(t) \qquad 1\,\text{PPM/K} = 1\,\text{FIT} = 10^8 \times h(t)$$

Example 5.1   Consider a pump unit that has an exponential failure density as

$$f_t(t) = \lambda e^{-\lambda t} \qquad \text{for } t \geq 0, \lambda > 0$$

in which λ is the number of failures per unit time. The reliability of the pump in time period $(0, t]$, according to Eq. (5.1), is

$$p_s(t) = \int_t^\infty \lambda e^{-\lambda t}\, dt = e^{-\lambda t}$$

as shown in Table 5.1. The failure rate for the pump, according to Eq. (5.3), is

$$h(t) = \frac{f_t(t)}{p_s(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

which is a constant. Since the instantaneous failure rate is a constant, the averaged failure rate for any time interval of interest also is a constant.

Example 5.2 Assume that the TTF has a normal distribution with the mean μ_t and standard deviation σ_t. Develop curves for the failure density function, reliability, and failure rate.

Solution For generality, it is easier to work on the standardized scale by which the random time to failure T is transformed according to $Z = (T - \mu_t)/\sigma_t$. In the standardized normal scale, the following table can be constructed easily:

(1) z	(2) $\phi(z)$	(3) $p_f(z) = \Phi(z)$	(4) $p_s(z)$	(5) $h(z)$	(6) $h(t) = h(z)/\sigma_t$
−3.5	0.0009	0.0002	0.9998	0.0009	0.0000018
−3.0	0.0044	0.0014	0.9986	0.0044	0.0000088
−2.5	0.0175	0.0062	0.9938	0.0176	0.0000352
−2.0	0.0540	0.0228	0.9772	0.0553	0.0001106
−1.5	0.1295	0.0668	0.9332	0.1388	0.0002776
−1.0	0.2420	0.1587	0.8413	0.2877	0.0005754
−0.5	0.3521	0.3085	0.6915	0.5092	0.0010184
0.0	0.3989	0.5000	0.5000	0.7978	0.0015956
0.5	0.3521	0.6915	0.3085	1.1413	0.0022826
1.0	0.2420	0.8413	0.1587	1.5249	0.0030498
1.5	0.1295	0.9332	0.0668	1.9386	0.0038772
2.0	0.0540	0.9772	0.0228	2.3684	0.0047368
2.5	0.0175	0.9938	0.0062	2.8226	0.0056452
3.0	0.0044	0.9986	0.0014	3.1429	0.0062858
3.5	0.0009	0.9998	0.0002	4.5000	0.0090000

NOTE: $\sigma_t = 500$ hours; $h(t)$ has a unit of failures/h, $t = \mu_t + \sigma_t z$.

Column (2) is simply the ordinate of the standard normal PDF computed by Eq. (2.59). Column (3) for the unreliability is the standard normal CDF, which can be obtained from Table 2.2 or computed by Eq. (2.63). Subtracting the unreliability in column (3) from one yields the reliability in column (4). Then failure rate $h(z)$ in column (5) is obtained by dividing column (2) by column (4) according to Eq. (5.3).

Note that the failure rate of the normal time to failure $h(t) = f_t(t)/p_s(t)$ is what the problem is after rather than $h(z)$. According to the transformation of variables, the following relationship holds:

$$f_t(t) = \phi(z)|dz/dt| = \phi(z)/\sigma_t$$

Since $p_s(t) = 1 - \Phi(z)$, the functional relationship between $h(t)$ and $h(z)$ can be derived as

$$h(t) = h(z)/\sigma_t$$

Column (5) of the table for $h(t)$ is obtained by assuming that $\sigma_t = 500$ hours. The relationships between the failure density function, reliability, and failure rate for the standardized and the original normal TTF are shown in Fig. 5.7. As can be seen, the failure rate for a normally distributed TTF increases monotonically as the system ages. Kapur and Lamberson (1977) showed that the failure-rate function associated with a normal TTF is a convex function of time. Owing to the monotonically increasing characteristics of the failure rate with time for a normally distributed TTF, it can be used to describe the system behavior during the wear-out period.
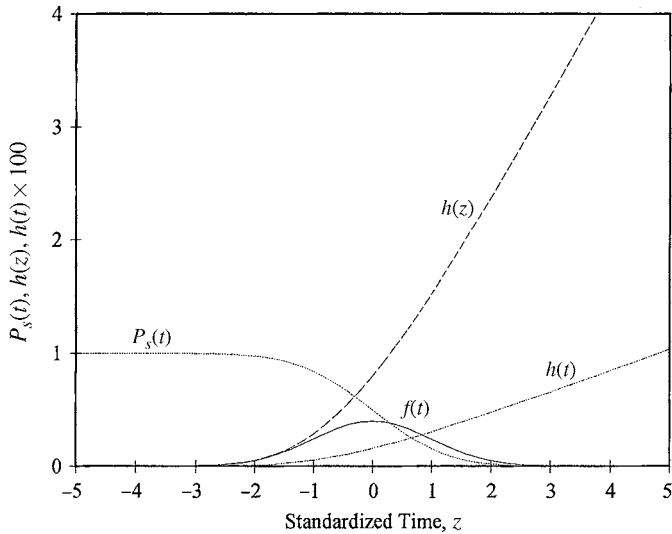
Figure 5.7 Reliability [$p_s(t)$], failure rate [$h(t)$], failure density function [$f_t(t)$] for a process/component the TTF of which follows a normal distribution as in Example 5.2.

5.2.5 Typical hazard functions

The failure rate for many systems or components has a bathtub shape, as shown in Fig. 5.8, in that three distinct life periods can be identified (Harr, 1987). They are the *early-life* (or *infant mortality*) *period*, *useful-life period*, and *wear-out-life period*. Kapur (1989b) differentiates three types of failure that result in the bathtub type of total failure rate, as indicated in Fig. 5.8. It is interesting to note that the failure rate in the early-life period is higher than during the useful-life period and has a decreasing trend with age. In this early-life period, quality failures and stress-related failures dominate, with little contribution from wear-out failures. During the useful-life period, all three types of failures contribute to the potential failure of the system or component, and the overall failure rate remains more or less constant over time. From Example 5.1, the exponential distribution could be used as the failure density function for the useful-life period. In the later part of life, the overall failure rate increases with age. In this life stage, wear-out failures and stress-related failures are the main contributors, and wear-out becomes an increasingly dominating factor for the failure of the system with age.

Quality failures, also called *break-in failures* (Wunderlich, 1993, 2004), are mainly related to the construction and production of the system, which could be caused by poor construction and manufacturing, poor quality control and workmanship, use of substandard materials and parts, improper installation, and human error. Failure rate of this type generally decreases with age. *Stress-related failures* generally are referred to as *chance failures*, which occur when loads on the system exceed its resistance, as described in Chap. 4. Possible causes of stress-related failures include insufficient safety factors, occurrence
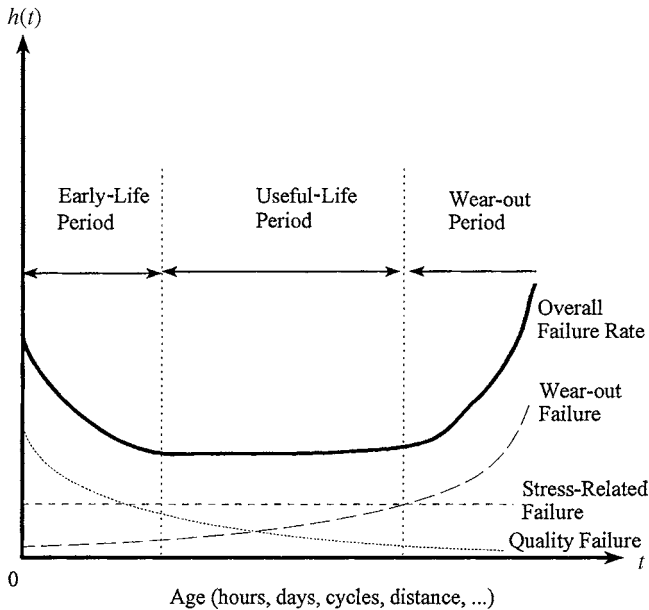
Figure 5.8 Bathtub failure rate with its three components.

of higher than expected loads or lower than expected random strength, misuse, abuse, and/or an act of God. *Wear-out failures* are caused primarily by aging; wear; deterioration and degradation in strength; fatigue, creep, and corrosion; or poor maintenance, repair, and replacement.

The failure of the 93-m-high Teton Dam in Idaho in 1976 was a typical example of break-in failure during the early-life period (Arthur, 1977; Jansen, 1988). The dam failed while the reservoir was being filled for the first time. Four hours after the first leakage was detected, the dam was fully breached. There are other examples of hydraulic structure failures during different stages of their service lives resulting from a variety of causes. For examples, in 1987 the foundation of a power plant on the Mississippi River failed after a 90-year service life (Barr and Heuer, 1989), and in the summer of 1993 an extraordinary sequence of storms caused the breach of levees in many parts along the Mississippi River. The failures and their impacts can be greatly reduced if proper maintenance and monitoring are actively implemented.

5.2.6 Relationships among failure density function, failure rate, and reliability

According to Eq. (5.3), given the failure density function $f_t(t)$ it is a straightforward task to derive the failure rate $h(t)$. Furthermore, based on Eq. (5.3), the reliability can be computed directly from the failure rate as

$$p_s(t) = \exp\left[-\int_0^t h(\tau)\,d\tau\right] \qquad (5.10)$$

Substituting Eq. (5.10) into Eq. (5.3), the failure density function $f_t(t)$ can be expressed in terms of the failure rate as

$$f_t(t) = h(t) \exp\left[-\int_0^t h(\tau)\,d\tau\right] \tag{5.11}$$

Example 5.3 (after Mays and Tung, 1992) Empirical equations have been developed for the break rates of water mains using data from a specific water distribution system. As an example, Walski and Pelliccia (1982) developed break-rate equations for the water distribution system in Binghamton, New York. These equations are

Pit cast iron: $N(t) = 0.02577e^{0.0207t}$

Sandspun cast iron: $N(t) = 0.0627e^{0.0137t}$

where $N(t)$ is the break rate (in number of breaks per mile per year), and t is the age of the pipe (in years). The break rates versus the ages of pipes for the preceding two types of cast iron pipes are shown in Fig. 5.9. Derive the expressions for the failure rate, reliability, and failure density function for a 5-mile water main of sandspun cast iron pipe.

Solution The break rate per year (i.e., failure rate or hazard function for the 5-mile water main) for sandspun cast iron pipe can be calculated as

$$h(t) = 5\text{ miles} \times N(t) = 0.3185e^{0.0137t}$$

The reliability of this 5-mile water main then can be computed using Eq. (5.10) as

$$p_s(t) = \exp\left(-\int_0^t 0.3185e^{0.0137\tau}\,d\tau\right) = \exp[23.25(1 - e^{0.0137t})]$$
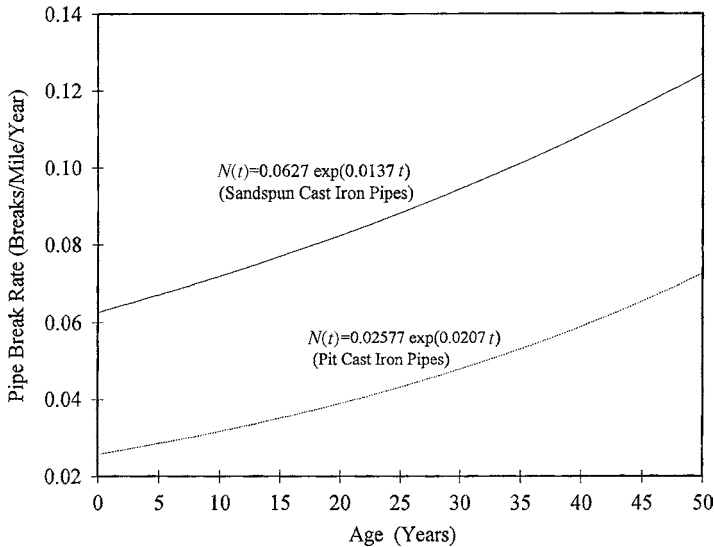
Figure 5.9 Break-rate curves for sandspun cast iron and pit cast iron pipes.

The failure density $f_t(t)$ can be calculated, using Eq. (5.11), as

$$f_t(t) = 0.3185e^{0.0137t} \times \exp[23.25(1 - e^{0.0137t})]$$

The curves for the failure rate, reliability, and failure density function of the 5-mile sandspun cast iron water main are shown in Fig. 5.10.

5.2.7 Effect of age on reliability

In general, the reliability of a system or a component is strongly dependent on its age. In other words, the probability that a system can be operational to perform its intended function satisfactorily is conditioned by its age. This *conditional reliability* can be expressed mathematically as

$$p_s(\xi \,|\, t) = \frac{P(\text{TTF} \geq t, \text{TTF} \geq t + \xi)}{P(\text{TTF} \geq t)} = \frac{P(\text{TTF} \geq t + \xi)}{P(\text{TTF} \geq t)} = \frac{p_s(t + \xi)}{p_s(t)} \quad (5.12)$$

in which t is the age of the system up to the point that the system has not failed, and $p_s(\xi \,|\, t)$ is the reliability over a new mission period ξ, having successfully operated over a period of $(0, t]$. In terms of failure rate, $p_s(\xi \,|\, t)$ can be written as

$$p_s(\xi \,|\, t) = \exp\left[-\int_t^{t+\xi} h(\tau)\, d\tau\right] \quad (5.13)$$
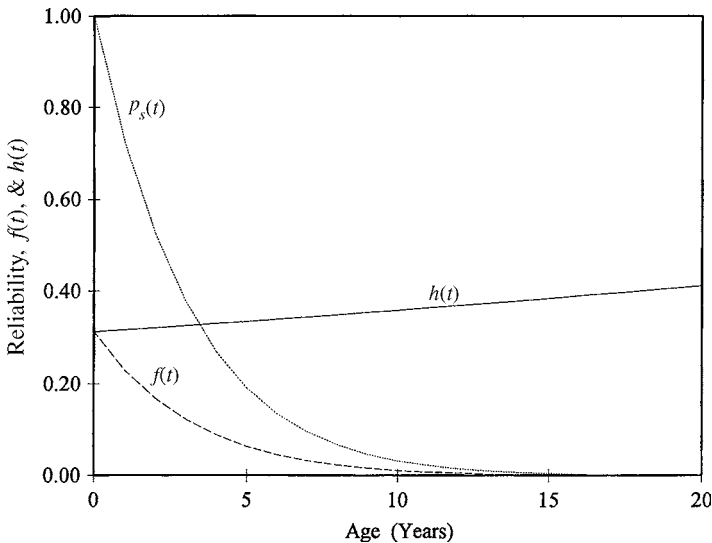
Figure 5.10 Curves for reliability $[p_s(t)]$, failure density $[f_t(t)]$, and failure rate $[h(t)]$ for the 5-mile sandspun cast iron pipe water main in Example 5.3.

Based on Eq. (5.2), the *conditional failure density function* can be obtained as

$$f(\xi \,|\, t) = \frac{-d\,[p_s(\xi \,|\, t)]}{d\xi} = -p_s(\xi \,|\, t) \times \frac{d}{d\xi}\left[-\int_t^{t+\xi} h(\tau)\,d\tau\right] = p_s(\xi \,|\, t)h(t+\xi)$$

(5.14)

For a process or component following the bathtub failure rate shown in Fig. 5.8 during the useful-life period, the failure rate is a constant, and the failure density function is an exponential distribution. Thus the failure rate $h(t) = \lambda$. The conditional reliability, according to Eq. (5.13), is

$$p_s(\xi \,|\, t) = e^{-\lambda t}$$

(5.15)

which shows that the conditional reliability depends only on the new mission period ξ regardless of the length of the previous operational period. Hence the time to failure of a system having an exponential failure density function is memoryless.

However, for nonconstant failure rates during the early-life and wear-out periods, the memoryless characteristics of the exponential failure density function no longer hold. Consider the Weibull failure density with $\alpha \neq 1$. Referring to Fig. 5.3, the condition $\alpha \neq 1$ precludes having a constant failure rate. According to Table 5.1, the conditional reliability for the Weibull failure density function is

$$p_s(\xi \,|\, t) = \frac{\exp\left[-\left(\dfrac{t+\xi-t_o}{\beta}\right)^{\alpha}\right]}{\exp\left[-\left(\dfrac{t-t_o}{\beta}\right)^{\alpha}\right]}$$

(5.16)

As can be seen, $p_s(\xi \,|\, t)$ will not be independent of the previous service period t when $\alpha \neq 1$. Consequently, to evaluate the reliability of a system for an additional service period in the future during the early-life and wear-out stages, it is necessary to know the length of the previous service period.

Example 5.4   Refer to Example 5.3. Derive the expression for the conditional reliability and conditional failure density of the 5-mile water main with sandspun cast iron pipe.

Solution   Based on the reliability function obtained in Example 5.3, the conditional reliability of the 5-mile sandspun cast iron pipe in the water distribution system can be derived, according to Eq. (5.12), as

$$p_s(\xi \,|\, t) = \frac{p_s(t+\xi)}{p_s(t)}$$

$$= \frac{\exp[23.25(1 - e^{0.0137(t+\xi)})]}{\exp[23.25(1 - e^{0.0137t})]}$$

$$= \exp[23.25e^{0.0137t}(1 - e^{0.0137\xi})]$$

The conditional failure density, according to Eq. (5.13), can be obtained as

$$f_t(\xi \,|\, t) = 0.3185 \times e^{0.0137(t+\xi)} \times \exp[23.25 e^{0.0137t}(1 - e^{0.0137\xi})]$$

Figure 5.11 shows the conditional reliability and conditional failure density of the pipe system for various service periods at different ages. Note that at age $t = 0$, the curve simply corresponds to the reliability function.

5.2.8   Mean time to failure

A commonly used reliability measure of system performance is the *mean time to failure* (MTTF), which is the expected TTF. The MTTF can be defined mathematically as

$$\text{MTTF} = E(\text{TTF}) = \int_0^\infty \tau \, f_t(\tau) \, d\tau \tag{5.17}$$

Referring to Eq. (2.30), the MTTF alternatively can be expressed in terms of reliability as

$$\text{MTTF} = \int_0^\infty p_s(t) \, dt \tag{5.18}$$

By Eq. (5.18), the MTTF geometrically is the area underneath the reliability function. The MTTF for some failure density functions are listed in the last column of Table 5.1. For illustration purposes, the MTTFs for some of the components in water distribution systems can be determined from mean time between failures (MTBF) and mean time to repair (MTTR) data listed in Tables 5.2 and 5.3.

Example 5.5   Refer to Example 5.3. Determine the expected elapsed time that a pipe break would occur in the 5-mile sandspun cast iron pipe water main.

Solution   The expected elapsed time over which a pipe break would occur can be computed, according to Eq. (5.17), as

$$\text{MTTF} = \int_0^\infty p_s(t) \, dt = \int_0^\infty \exp[23.25(1 - e^{0.0137t})] \, dt = 3.015 \, \text{years}$$

The main reason for using Eq. (5.18) is purely for computational considerations because the expression for $p_s(t)$ is much simpler than $f_t(t)$.

5.3   Repairable Systems

For repairable hydrosystems, such as pipe networks, pump stations, and storm runoff drainage structures, failed components within the system can be repaired or replaced so that the system can be put back into service. The time required to have the failed system repaired is uncertain, and consequently, the total time required to restore the system from its failure state to an operational state is a random variable.
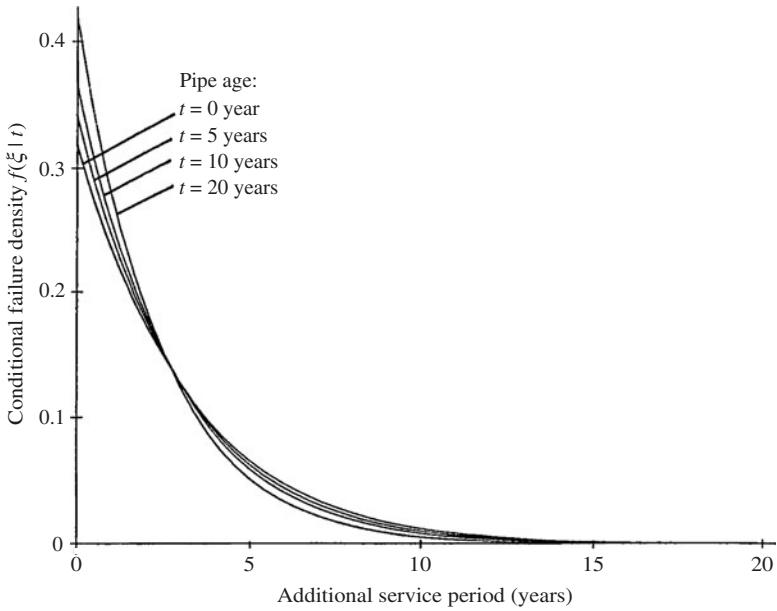
Figure 5.11 Conditional reliability and conditional failure density for the 5-mile water main made of sandspun cast iron pipe in Example 5.4.

TABLE 5.2   Reliability and Maintainability of Water Distribution Subsystems by Generic Group

Subsystem	MTBF* ($\times 10^6$ hours)	MTTR* (hours)
Pumps		
Centrifugal, open impeller	0.021660	7.825
Axial flow, propeller	0.074191	16.780
Power transmission		
Concentric reducer	0.122640	2.000
Parallel shaft	0.710910	32.000
Right angle shaft	0.019480	1.400
Vertical shaft	0.031470	2.023
Variable speed, hydraulic	0.349500	—
Variable speed, other	0.014200	2.500
Gear box	0.045780	3.530
Chain drive	0.017850	8.000
Belt drive	0.091210	1.800
Motors		
Multiphase	0.068000	6.853
Variable speed, ac	0.114820	8.000
Gas engine	0.023800	24.000
Valves		
Gate	0.008930	3.636
Ball	0.011460	—
Butterfly	0.032590	1.000
Plug	0.028520	—
Controls		
Electrical	0.100640	2.893
Mechanical	0.031230	8.000
Pressure (fluid)	0.035780	8.236
Pressure (air)	0.018690	3.556

*MTBF = mean time between failure; MTTR = mean time to repair; MTBF = MTTF + MTTR.
SOURCE: From Schultz and Parr (1981).

5.3.1   Repair density and repair probability

Like the time to failure, the random *time to repair* (TTR) has the *repair density function* $g_t(t)$ describing the random characteristics of the time required to repair a failed system when the failure occurs at time zero. The *repair probability* $G_t(t)$ is the probability that the failed system can be restored within a given time period $(0, t]$:

$$G_t(t) = P(\text{TTR} \leq t) = \int_0^t g_t(\tau)\, d\tau \tag{5.19}$$

The repair probability $G_t(t)$ is also called the *maintainability function* (Knezevic, 1993), which is one of the measures for maintainability (Kapur, 1988b). *Maintainability* is a design characteristic to achieve fast, easy maintenance at the lowest life-cycle cost. In addition to the maintainability function, other types of maintainability measures are derivable from the repair density function (Kraus, 1988; Knezevic, 1993), and they are the mean time to repair (described in Sec. 5.3.3), TTR_p, and the restoration success.

TABLE 5.3 Reliability and Maintainability of Water Distribution Subsystems by Size

Subsystem	MTBF*($\times 10^6$ hours)	MTTR* (hours)
Pumps (in gpm)		
1–10,000	0.039600	6.786
10,001–20,000	0.031100	7.800
20,001–100,000	0.081635	26.722
Over 100,000	0.008366	9.368
Power transmission (in horsepower)		
0–1	0.025370	1.815
2–5	0.011010	2.116
6–25	1.376400	25.000
26–100	0.058620	5.000
101–500	0.078380	2.600
Over 500	0.206450	32.000
Motors (in horsepower)		
0–1	0.206450	2.600
2–5	0.214700	—
6–25	0.565600	7.857
26–100	0.062100	4.967
101–500	0.046000	12.685
Over 500	0.064630	7.658
Valves (in inches)		
6–12	0.054590	—
13–24	0.010810	1.000
25–48	0.019070	42.000
Over 48	0.007500	2.667
Controls (in horsepower)		
0–1	2.009200	2.050
2–5	0.509500	—
6–25	4.684900	—
26–100	0.026109	2.377
101–500	0.099340	5.450
Over 500	0.037700	3.125

*MTBF = mean time between failure; MTTR = mean time to repair; MTBF = MTTF + MTTR.
  SOURCE : From Schultz and Parr (1981).

The TTR_p is the maintenance time by which $100p$ percent of the repair work is completed. The value of the TTR_p can be determined by solving

$$P(\text{TTR} \leq \text{TTR}_p) = \int_0^{\text{TTR}_p} g_t(\tau)\, d\tau = G_t(\text{TTR}_p) = p \qquad (5.20)$$

In other words, the TTR_p is the pth order quantile of the repair density function. In general, $p = 0.90$ is used commonly.

Note that the repair probability or maintainability function $G_t(t)$ represents the probability that the restoration can be completed before or at time t. Sometimes one may be interested in the probability that the system can be restored by time t_2, given that it has not been repaired at an earlier time t_1. This type of conditional repair probability, similar to the conditional reliability of Eq. (5.12),

is called the *restoration success* $RS(t_1, t_2)$, which is defined mathematically as

$$RS(t_1, t_2) = P\left[\text{TTR} \le t_2 \,|\, \text{TTR} > t_1\right] = \frac{G_t(t_2) - G_t(t_1)}{1 - G(t_1)} \qquad (5.21)$$

Kraus (1988) pointed out the difference in maintainability and maintenance; namely, maintainability is design-related, whereas maintenance is operation-related. Since the MTTF is a measure of maintainability, it includes those time elements that can be controlled by design. Elements involved in the evaluation of the time to repair are fault isolation, repair or replacement of a failed component, and verification time. Administrative times, such as mobilization time and time to reach and return from the maintenance site, are not included in the evaluation of the time to repair. The administrative times are considered under the context of *supportability* (see Sec. 5.3.4), which measures the ability of a system to be supported by the required resources for execution of the specified maintenance task (Knezevic, 1993).

5.3.2 Repair rate and its relationship with repair density and repair probability

The *repair rate* $r(t)$, similar to the failure rate, is the conditional probability that the system is repaired per unit time given that the system failed at time zero and is still not repaired at time t. The quantity $r(t)\,dt$ is the probability that the system is repaired during the time interval $(t, t + dt]$ given that the system fails at time t. Similar to Eq. (5.3), the relationship among repair density function, repair rate, and repair probability is

$$r(t) = \frac{g_t(t)}{1 - G_t(t)} \qquad (5.22)$$

Given a repair rate $r(t)$, the repair density function and the maintainability can be determined, respectively, as

$$g_t(t) = r(t) \times \exp\left[-\int_0^t r(\tau)\,d\tau\right] \qquad (5.23)$$

$$G_t(t) = 1 - \exp\left[-\int_0^t r(\tau)\,d\tau\right] \qquad (5.24)$$

5.3.3 Mean time to repair, mean time between failures, and mean time between repairs

The *mean time to repair* (MTTR) is the expected value of time to repair of a failed system, which can be calculated by

$$\text{MTTR} = \int_0^\infty \tau g_t(\tau)\,d\tau = \int_0^\infty [1 - G_t(\tau)]\,d\tau \qquad (5.25)$$

The MTTR measures the elapsed time required to perform the maintenance operation and is used to estimate the downtime of a system. The MTTR values for some components in a water distribution system are listed in the last columns of Tables 5.2 and 5.3. It is also a commonly used measure for the maintainability of a system.

The MTTF is a proper measure of the mean life span of a nonrepairable system. However, for a repairable system, the MTTF is no longer appropriate for representing the mean life span of the system. A more representative indicator for the fail-repair cycle is the *mean time between failures* (MTBF), which is the sum of MTTF and MTTR, that is,

$$\text{MTBF} = \text{MTTF} + \text{MTTR} \tag{5.26}$$

The *mean time between repairs* (MTBR) is the expected value of the time between two consecutive repairs, and it is equal to MTBF. The MTBF for some typical components in a water distribution system are listed in Tables 5.2 and 5.3.

Example 5.6   Consider a pump having a failure density function of

$$f_t(t) = 0.0008 \exp(-0.0008t) \qquad \text{for } t \geq 0$$

and a repair density function of

$$g_t(t) = 0.02 \exp(-0.02t) \qquad \text{for } t \geq 0$$

in which t is in hours. Determine the MTBF for the pump.

Solution   To compute the MTBF, the MTTF and MTTR of the pump should be calculated separately. Since the time to failure and time to repair are exponential random variables, the MTTF and MTTR, respectively, are

$$\text{MTTF} = 1/0.0008 = 1250 \, \text{hours}$$

$$\text{MTTR} = 1/0.02 = 50 \, \text{hours}$$

Therefore, MTBF = MTTF + MTTR = 1250 + 50 = 1300 hours.

5.3.4   Preventive maintenance

There are two basic categories of maintenance: corrective maintenance and preventive maintenance. *Corrective maintenance* is performed when the system experiences in-service failures. Corrective maintenance often involves the needed repair, adjustment, and replacement to restore the failed system back to its normal operating condition. Therefore, corrective maintenance can be regarded as repair, and its stochastic characteristics are describable by the repair function, MTTR, and other measures discussed previously in Secs. 5.3.1 through 5.3.3.

On the other hand, *preventive maintenance*, also called *scheduled maintenance*, is performed in a regular time interval involving periodic inspections,

even if the system is in working condition. In general, preventive maintenance involves not only repair but inspection and some replacements. Preventive maintenance is aimed at postponing failure and prolonging the life of the system to achieve a longer MTTF for the system. This section will focus on some basic features of preventive maintenance.

From the preceding discussions of what a preventive maintenance program wishes to achieve, it is obvious that preventive maintenance is only a waste of resources for a system having a decreasing or constant hazard function because such an activity cannot decrease the number of failures (see Example 5.7). If the maintenance is neither ideal nor perfect, it may even have an adverse impact on the functionality of the system. Therefore, preventive maintenance is a worthwhile consideration for a system having an increasing hazard function or an aged system (see Problems 5.18 and 5.19).

Ideal maintenance.   An *ideal maintenance* has two features: (1) zero time to complete, relatively speaking, as compared with the time interval between maintenance, and (2) system is restored to the "as new" condition. The second feature often implies a replacement.

Let t_M be the fixed time interval between the scheduled maintenance, and $p_{s,M}(t)$ is the *reliability function with preventive maintenance*. The reliability of the system at time t, after k preventive maintenances, with $kt_M < t \leq (k+1)t_M$, for $k = 0, 1, 2, \ldots$, is

$$p_{s,M}(t) = P \{\text{no failure in } (0, t_M], \text{no failure in } (t_M, 2\,t_M], \ldots,$$
$$\text{no failure in } ((k-1)t_M, kt_M], \text{no failure in } (kt_M, t]\}$$

$$= P \left\{ \bigcap_{i=1}^{k} \text{no failure in } ((i-1)t_M, it_M], \text{no failure in } (kt_M, t] \right\}$$

$$= [p_s(t_M)]^k \times p_s(t - kt_M) \tag{5.27}$$

where $p_{s,M}(t)$ is the unmaintained reliability function defined in Eq. (5.1a).

The *failure density function with maintenance* $f_M(t)$ can be obtained from Eq. (5.27), according to Eq. (5.2), as

$$f_M(t) = -\frac{d\,[p_{s,M}(t)]}{dt}$$

$$= -[p_s(t_M)]^k \times \frac{d\,[p_s(t - kt_M)]}{dt} = [p_s(t_M)]^k f_t(t - kt_M) \tag{5.28}$$

for $kt_M < t \leq (k+1)t_M$, with $k = 0, 1, 2, \ldots$. As can be seen from Eqs. (5.27) and (5.28), the reliability function and failure density function with maintenance in each time segment, defined by two consecutive preventive maintenances, are scaled down by a factor of $p_s(t_M)$ as compared with the proceeding segment.
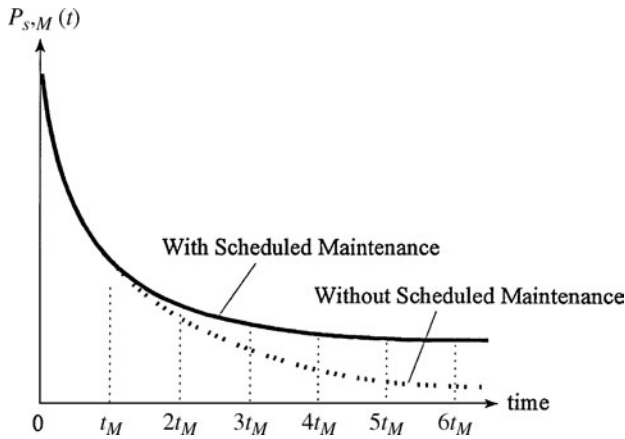
Figure 5.12 Reliability function with and without preventive maintenance.

The factor $p_s(t_M)$ is the fraction of the total components that will survive from one segment of the maintenance period to the next. Geometric illustrations of Eqs. (5.27) and (5.28) are shown in Figs. 5.12 and 5.13, respectively. The envelop curve in Fig. 5.12 (shown by a dashed line) exhibits an exponential decay with a factor of $p_s(t_M)$.

Similar to an unmaintained system, the *hazard function with maintenance* can be obtained, according Eq. (5.3), as

$$h_M(t) = \frac{f_M(t - kt_M)}{p_{s,M}(t - kt_M)} \qquad \text{for } kt_M < t \le (k+1)t_M, k = 0, 1, 2, \dots \qquad (5.29)$$
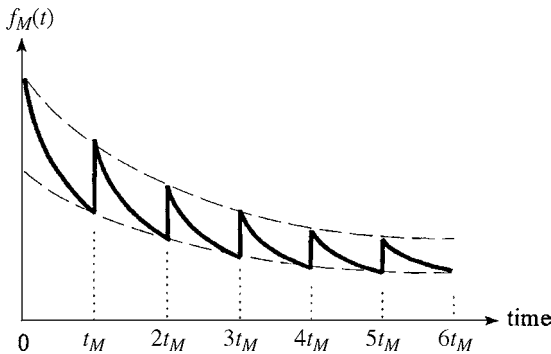
Figure 5.13 Failure density function with ideal preventive maintenance.

The *mean time-to-failure with maintenance* MTTF_M can be evaluated, according to Eq. (5.18), as

$$\mathrm{MTTF}_M = \int_0^\infty p_{s,M}(t)\,dt = \sum_{k=0}^\infty \int_{kt_M}^{(k+1)t_M} p_{s,M}(t)\,dt$$

$$= \sum_{k=0}^\infty [p_s(t_M)]^k \int_{kt_M}^{(k+1)t_M} p_s(t - kt_M)\,dt \qquad (5.30)$$

By letting $\tau = t - kt_M$, the preceding integration for computing the MTTF_M can be rewritten as

$$\mathrm{MTTF}_M = \left\{ \sum_{k=0}^\infty [p_s(t_M)]^k \right\} \left[\int_0^{t_M} p_s(\tau)\,d\tau \right] = \frac{\int_0^{t_M} p_s(\tau)\,d\tau}{1 - p_s(t_M)} \qquad (5.31)$$

using $1/(1-x) = 1 + x + x^2 + x^3 + x^4 + \dots$, for $0 < x < 1$.

A preventive maintenance program is worth considering if the reliability with maintenance is greater than the reliability without maintenance. That is,

$$\frac{p_{s,M}(t)}{p_s(t)} = \frac{[p_s(t_M)]^k p_s(t - kt_M)}{p_s(t)} > 1 \qquad \text{for } kt_M < t \le (k+1)t_M, k = 0, 1, 2, \dots$$

$$(5.32)$$

Letting $t = kt_M$ and assuming $p_s(0) = 1$, the preceding expression can be simplified as

$$\frac{p_{s,M}(kt_M)}{p_s(kt_M)} = \frac{[p_s(t_M)]^k}{p_s(kt_M)} > 1 \qquad \text{for } k = 0, 1, 2, \dots \qquad (5.33)$$

Similarly, the implementation of a preventive maintenance program is justifiable if $\mathrm{MTTF}_M > \mathrm{MTTF}$ or $h_M(t) > h(t)$ for all time t.

Example 5.7 Suppose that a system is implemented with preventive maintenance at a regular time interval of t_M. The failure density of the system is of an exponential type as

$$f_t(t) = \lambda e^{-\lambda t} \qquad \text{for } t \ge 0$$

Assuming that the maintenance is ideal, find the expression for the reliability function and the mean time to failure of the system.

Solution The reliability function of the system if no maintenance is in place is (from Table 5.1)

$$p_s(t) = e^{-\lambda t} \qquad \text{for } t \ge 0$$

The reliability of the system under a regular preventive maintenance of time interval t_M can be derived, according to Eq. (5.27), as

$$p_{s,M}(t) = (e^{-\lambda t_M})^k \times e^{-\lambda(t - kt_M)} \qquad \text{for } kt_M < t \le (k+1)t_M, k = 0, 1, 2, \dots$$

which can be reduced to

$$p_{s,M}(t) = e^{-\lambda t} \qquad \text{for } t \geq 0$$

The mean time to failure of the system with maintenance can be calculated, according to Eq. (5.31), as

$$\text{MMTF}_M = \frac{\int_0^{t_M} e^{-\lambda \tau} \, d\tau}{1 - e^{-\lambda t_M}} = \frac{\frac{1}{\lambda}(1 - e^{-\lambda t_M})}{1 - e^{-\lambda t_M}} = \frac{1}{\lambda}$$

As can be seen, with preventive maintenance in place, the reliability function and the mean time to failure of a system having an exponential failure density (constant failure rate) are identical to those without maintenance.

Example 5.8 Consider a system having a uniform failure density bounded in [0, 5 years]. Evaluate the reliability, hazard function, and MTTF for the system if a preventive maintenance program with a 1-year maintenance interval is implemented. Assume that the maintenance is ideal.

Solution The failure density function for the system is

$$f_t(t) = 1/5 \qquad \text{for } 0 \leq t \leq 5$$

From Table 5.1, the reliability function, hazard function, and MTTF of the system without maintenance, respectively, are

$$p_s(t) = (5 - t)/5 \qquad \text{for } 0 \leq t \leq 5$$

$$h(t) = 1/(t - 5) \qquad \text{for } 0 \leq t \leq 5$$

and
$$\text{MTTF} = 5/2 = 2.5 \text{ years}$$

With the maintenance interval $t_M = 1$ year, the reliability function, failure density, hazard function, and the MTTF can be derived, respectively, as

$$p_{s,M}(t) = (4/5)^k [(5 - t + k)/5] \qquad \text{for } k < t \leq k + 1, k = 0, 1, 2, \ldots$$

$$f_M(t) = (4/5)^k (1/5) \qquad \text{for } k < t \leq k + 1, k = 0, 1, 2, \ldots$$

$$h_M(t) = \frac{f_M(t)}{p_{s,M}(t)} = 1/(5 - t + k) \qquad \text{for } k < t \leq k + 1, k = 0, 1, 2, \ldots$$

and
$$\text{MMTF}_M = \frac{\int_0^1 \dfrac{5 - \tau}{5} \, d\tau}{1 - \dfrac{4}{5}} = \frac{9/10}{1/5} = 4.5 \text{ years}$$

Referring to Fig. 5.6, the hazard function for the system associated with a uniform failure density function is increasing with time. This example shows that the MTTF_M is larger than the MTTF, indicating that the scheduled maintenance is beneficial to the system under consideration. Furthermore, plots of the reliability, failure density, and hazard function for this example are shown in Fig. 5.14.

In the context of scheduled maintenance, the number of maintenance applications K_M before system failure occurs is a random variable of significant
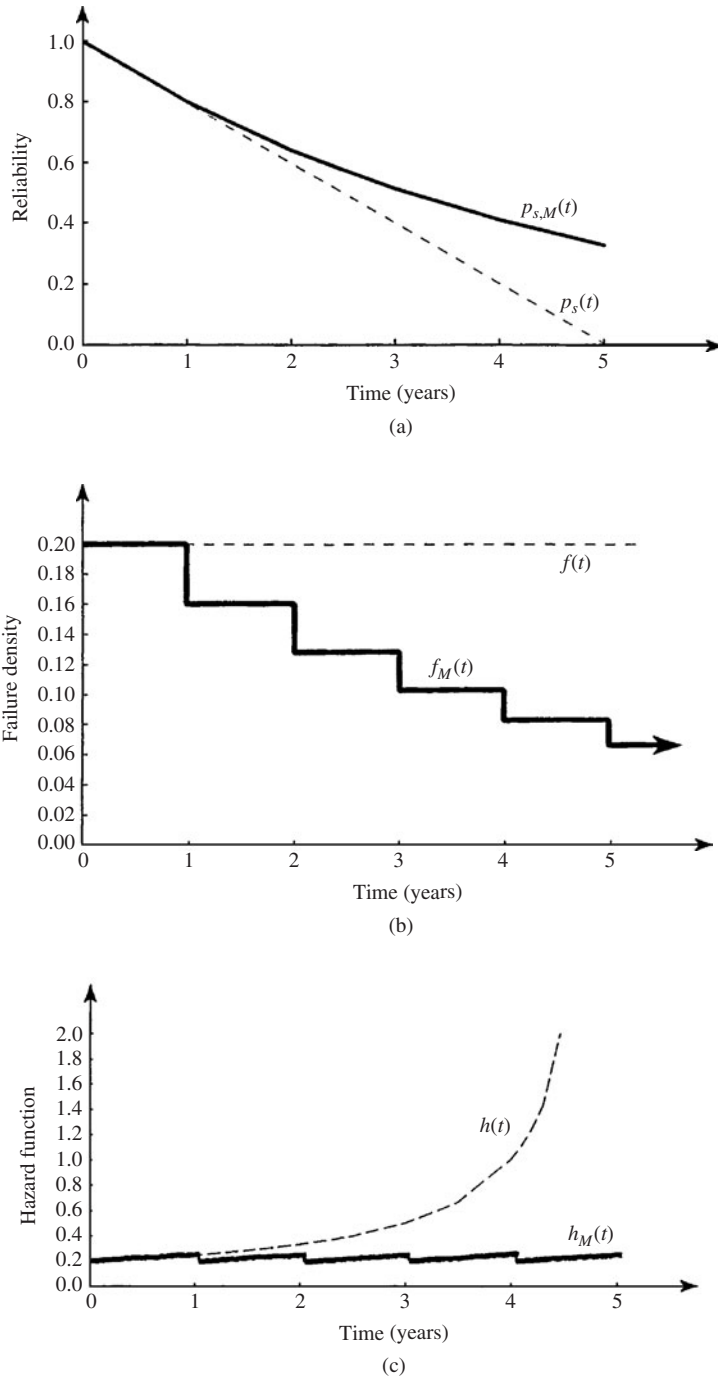
Figure 5.14 Reliability function (*a*) with [$p_{s,M}(t)$] and without [$p_s(t)$] preventive maintenance, failure density function (*b*) with [$f_M(t)$] and without [$f_t(t)$] preventive maintenance, and hazard function (*c*) with [$h_M(t)$] and without [$h(t)$] preventive maintenance for Example 5.8.

importance. The probability that the system will undergo exactly k preventive maintenance applications before failure is the probability that system failure occurs before $(k + 1)t_M$, which can be expressed as

$$q_k = [p_s(t_M)]^k [1 - p_s(t_M)] \qquad \text{for } k = 0, 1, 2, \ldots \tag{5.34}$$

which has the form of a *geometric distribution*. From Eq. (5.34), the *expected number of scheduled maintenance applications* before the occurrence of system failure is

$$E(K_M) = \sum_{k=0}^{\infty} k \times q_k = [1 - p_s(t_M)] \sum_{k=1}^{\infty} k \times [p_s(t_M)]^k = \frac{p_s(t_M)}{1 - p_s(t_M)} \tag{5.35}$$

and the variance of K_M is

$$\text{Var}(K_M) = \sum_{k=0}^{\infty} k^2 \times q_k - E^2(K_M) = [1 - p_s(t_M)] \sum_{k=1}^{\infty} k^2 [p_s(t_M)]^k - \left[\frac{p_s(t_M)}{1 - p_s(t_M)} \right]^2$$

$$= \frac{p_s(t_M)[1 + p_s(t_M)]}{[1 - p_s(t_M)]^2} - \left[\frac{p_s(t_M)}{1 - p_s(t_M)} \right]^2$$

$$= \frac{p_s(t_M)}{[1 - p_s(t_M)]^2} \tag{5.36}$$

The algebraic manipulations used in Eqs. (5.35) and (5.36) employ the following relationships under the condition $0 < x < 1$:

$$\sum_{i=1}^{\infty} i x^i = \frac{x}{(1-x)^2} \qquad \sum_{i=1}^{\infty} i^2 x^i = \frac{x(1+x)}{(1-x)^3} \tag{5.37}$$

Example 5.9 Referring to Example 5.7, having an exponential failure density function, derive the expressions for the expected value and variance of the number of scheduled maintenance applications before the system fails.

Solution According to Eq. (5.35), the expected number of scheduled maintenance applications before failure can be derived as

$$E(K_M) = \frac{p_s(t_M)}{1 - p_s(t_M)} = \frac{e^{-\lambda t_M}}{1 - e^{-\lambda t_M}} = \frac{1}{e^{\lambda t_M} - 1} = \frac{1}{e^{t_M/\text{MTTF}} - 1}$$

The variance of the number of scheduled maintenance applications before failure can be derived, according to Eq. (5.36), as

$$\text{Var}(K_M) = \frac{p_s(t_M)}{[1 - p_s(t_M)]^2} = \frac{e^{-\lambda t_M}}{(1 - e^{-\lambda t_M})^2} = \frac{e^{\lambda t_M}}{(e^{\lambda t_M} - 1)^2} = \frac{e^{t_M/\text{MTTF}}}{(e^{t_M/\text{MTTF}} - 1)^2}$$

Time variations of the expected value and standard deviation of the number of scheduled maintenance applications before failure for a system with an exponential failure density function are shown in Fig. 5.15. It is observed clearly that, as expected, when the time interval for scheduled maintenance t_M becomes longer relative to the MTTF,
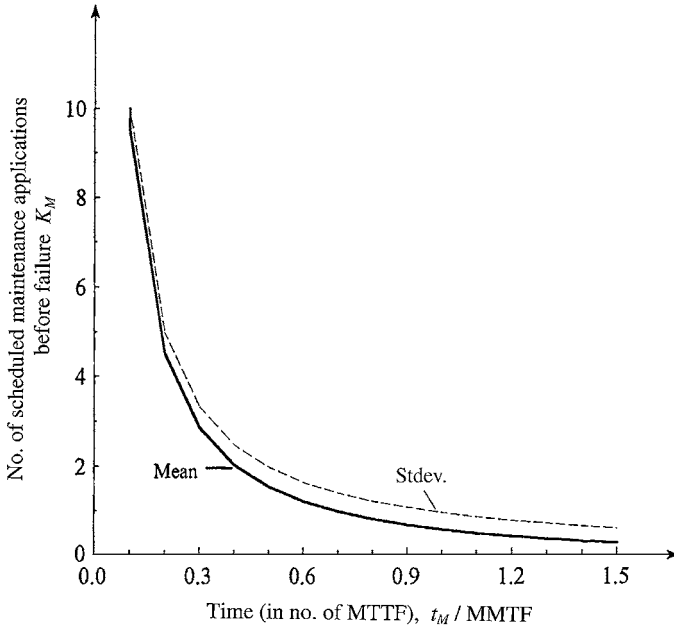
Figure 5.15 Time variations of the expected value and standard deviation of the number of scheduled maintenance applications before failure for a system with an exponential failure density function, as in Example 5.9.

the expected number of scheduled maintenance applications $E(K_M)$ and its associated standard deviation $\sigma(K_M)$ decrease. However, the coefficient of variation $\Omega(K_M)$ increases. Interestingly, when $t_M/\text{MTTF} = 1$, $E(K_M) = 1/(e-1) = 0.58$, indicating that failure could occur before maintenance if the scheduled maintenance time interval is set to be MTTF under the exponential failure density function.

Example 5.10 Referring to Example 5.8, with a uniform failure density function, compute the expected value and standard deviation of the number of scheduled maintenance applications before the system fails.

Solution According to Eq. (5.32), the expected number of scheduled maintenance applications before failure can be derived as

$$E(K_M) = \frac{p_s(t_M)}{1 - p_s(t_M)} = \frac{4/5}{1 - (4/5)} = \frac{4/5}{1/5} = 4$$

The variance of the number of scheduled maintenance applications before failure can be derived, according to Eq. (5.33), as

$$\text{Var}(K_M) = \frac{p_s(t_M)}{[1 - p_s(t_M)]^2} = \frac{4/5}{[1 - (4/5)]^2} = 20$$

The standard deviation of the number of scheduled maintenance applications before failure for the system is $\sqrt{20} = 4.47$ scheduled maintenance applications.

Imperfect maintenance. Owing to faulty maintenance as a result of human error, the system under repair could fail soon after the preventive maintenance application. If the probability of performing an imperfect maintenance is q, the reliability of the system is to be multiplied by $(1-q)$ each time maintenance is performed, that is,

$$p_{s,M}(t \mid q) = [(1-q)p_s(t_M)]^k \, p_s(t - kt_M) \qquad \text{for } kt_M < t \le (k+1)t_M, k = 0, 1, 2, \dots$$
(5.38)

An imperfect maintenance is justifiable only when, at $t = kt_M$,

$$\frac{p_{s,M}(kt_M \mid q)}{p_s(kt_M)} = \frac{[(1-q)p_s(t_M)]^k}{p_s(kt_M)} > 1 \qquad \text{for } k = 0, 1, 2, \dots$$
(5.39)

Example 5.11   Refer to Example 5.7, with an exponential failure density function. Show that implementing an imperfect maintenance is in fact damaging.

Solution   By Eq. (5.39), the ratio of reliability functions with and without imperfect maintenance for a system with an exponential failure density is

$$\frac{p_{s,M}(kt_M \mid q)}{p_s(kt_M)} = (1-q)^k \left(\frac{e^{-k\lambda t_M}}{e^{-k\lambda t_M}} \right) = (1-q)^k < 1 \qquad \text{for } k \ge 1$$

This indicates that performing an imperfect maintenance for a system with an exponential failure density function could reduce reliability.

5.3.5   Supportability

For a repairable component, supportability is an issue concerning the ability of the components, when they fail, to receive the required resources for carrying out the specified maintenance task. It is generally represented by the *time to support* (TTS), which may include administrative time, logistic time, and mobilization time. Similar to the TTF and TTR, TTS, in reality, also is randomly associated with a probability density function. The cumulative distribution function of the random TTS is called the *supportability function*, representing the probability that the resources will be available for conducting a repair task at a specified time. Also, other measures of supportability include the *mean time to support* (MTTS), TTS_p, and *support success*, similar to those defined for maintainability, with the repair density function replaced by the density function of the TTS.

5.4   Determinations of Availability and Unavailability

5.4.1   Terminology

A repairable system experiences a repetition of the repair-to-failure and failure-to-repair processes during its service life. Hence the probability that a system is in an operating condition at any given time t for a repairable system is different from that for a nonrepairable system. The term *availability $A(t)$* generally is

used for repairable systems to indicate the probability that the system is in an operating condition at any given time t. It also can be interpreted as the percentage of time that the system is in an operating condition within a specified time period. On the other hand, reliability $p_s(t)$ is appropriate for nonrepairable systems, indicating the probability that the system has been continuously in an operating state starting from time zero up to time t.

There are three types of availability (Kraus, 1988). *Inherent availability* is the probability of a system, when used under stated conditions and without consideration of any scheduled or preventive actions, in an ideal support environment, operating satisfactorily at a given time. It does not include ready time, preventive downtime, logistic time, and administrative time. *Achieved availability* considers preventive and corrective downtime and maintenance time. However, it does not include logistic time and administrative time. *Operational availability* considers the actual operating environment. In general, the inherent availability is higher than the achieved availability, followed by the operational availability (see Example 5.13). Of interest to design is the inherent availability; this is the type of availability discussed in this chapter.

In general, the availability and reliability of a system satisfy the following inequality relationship:

$$0 \le p_s(t) \le A(t) \le 1 \tag{5.40}$$

with the equality for $p_s(t)$ and $A(t)$ holding for nonrepairable systems. The reliability of a system decreases monotonically to zero as the system ages, whereas the availability of a repairable system decreases but converges to a positive probability (Fig. 5.16).
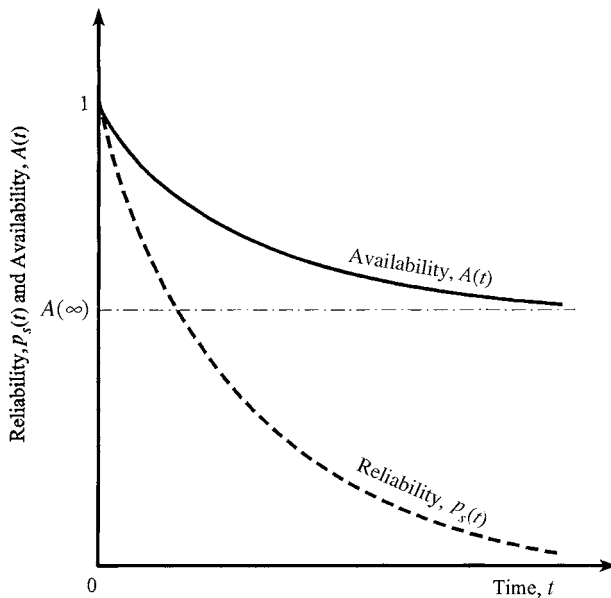
Figure 5.16 Comparison of reliability and availability.

The complement to the availability is the *unavailability* $U(t)$, which is the probability that a system is in a failed condition at time t, given that it was in an operating condition at time zero. In other words, unavailability is the percentage of time the system is not available for the intended service in time period $(0, t]$, given that it was operational at time zero. Availability, unavailability, and unreliability satisfy the following relationships:

$$A(t) + U(t) = 1 \tag{5.41}$$

$$0 \leq U(t) \leq p_f(t) \leq 1 \tag{5.42}$$

For a nonrepairable system, the unavailability is equal to the unreliability, that is, $U(t) = p_f(t)$.

Recall the failure rate in Sec. 5.2.2 as being the probability that a system experiences a failure per unit time at time t, given that the system was operational at time zero and has been in operation continuously up to time t. This notion is appropriate for nonrepairable systems. For a repairable system, the term *conditional failure intensity* $\mu(t)$ is used, which is defined as the probability that the system will fail per unit time at time t, given that the system was operational at time zero and also was in an operational state at time t. Therefore, the quantity $\mu(t) dt$ is the probability that the system fails during the time interval $(t, t + dt]$, given that the system was as good as new at time zero and was in an operating condition at time t. Both $\mu(t) dt$ and $h(t) dt$ are probabilities that the system fails during the time interval $(t, t + dt]$, being conditional on the fact that the system was operational at time zero. The difference is that the latter, $h(t) dt$, requires that the system has been in a continuously operating state from time zero to time t, whereas the former allows possible failures before time t, and the system is repaired to the operating state at time t. Hence $\mu(t) \neq h(t)$ for the general case, and they are equal for nonrepairable systems or when $h(t)$ is a constant (Henley and Kumamoto, 1981).

A related term is the *unconditional failure intensity* $w(t)$, which is defined as the probability that a system will fail per unit time at time t, given that the system is in an operating condition at time zero. Note that the unconditional failure intensity does not require that the system is operational at time t. For a nonrepairable system, the unconditional failure intensity is equal to the failure density $f_t(t)$. The *number of failures* experienced by the system within a specified time interval $[t_1, t_2]$ can be evaluated as

$$W(t_1, t_2) = \int_{t_1}^{t_2} w(\tau) \, d\tau \tag{5.43}$$

Hence, for a nonrepairable system, $W(0, t)$ is equal to the unreliability, which approaches unity as t increases. However, for repairable systems, $W(0, t)$ would diverge to infinite as t gets larger (Fig. 5.17).

On the repair aspect of the system, there are elements similar to those of the failure aspect. The *conditional repair intensity* $\rho(t)$ is defined as the probability that a system is repaired per unit time at time t, given that the system was in
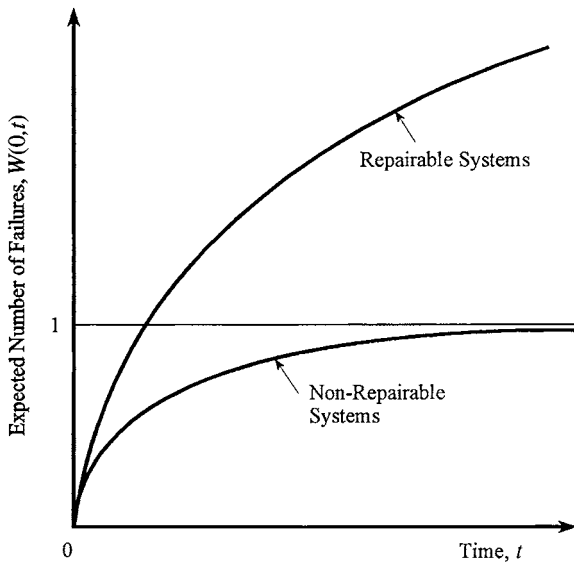
Figure 5.17 Expected number of failures for repairable and nonrepairable systems.

an operational state initially at time zero but in a failed condition at time t. The *unconditional repair intensity* $\gamma(t)$ is the probability that a failed system will be repaired per unit time at time t, given that it was initially in an operating condition at time zero. The *number of repairs* over a specified time period (t_1, t_2), analogous to Eq. (5.43), can be expressed as

$$\Gamma(t_1, t_2) = \int_{t_1}^{t_2} \gamma(\tau) \, d\tau \tag{5.44}$$

in which $\Gamma(0, t)$ is the expected number of repairs for a repairable system within the time interval $[t_1, t_2]$. A repairable system has $\Gamma(0, t)$ approaching infinity as t increases, whereas it is equal to zero for a nonrepairable system. It will be shown in the next subsection that the difference between $W(0, t)$ and $\Gamma(0, t)$ is the unavailability $U(t)$.

5.4.2 Determinations of availability and unavailability

Determination of the availability or unavailability of a system requires a full accounting of the failure and repair processes. The basic elements that describe such processes are the failure density function $f_t(t)$ and the repair density function $g_t(t)$. In this section computation of the availability of a single component or system is described under the condition of an ideal supportability. That is, the availability, strictly speaking, is the inherent availability. Discussions of the subject matter for a complex system are given in Chap. 7.

Consider a specified time interval $(0, t]$, and assume that the system is initially in an operating condition at time zero. Therefore, at any given time instance t, the system is in an operating state if the number of failures and repairs $w(0, t)$ are equal, whereas the system is in a failed state if the number of failures exceeds the number of repairs by one. Let $N_F(t)$ and $N_R(t)$ be the random variables representing the numbers of failures and repairs in time interval $(0, t]$, respectively. The state of the system at time instance t, failed or operating, can be indicated by a new variable $I(t)$ defined as

$$I(t) = N_F(t) - N_R(t) \tag{5.45}$$

Note that $I(t)$ also is a random variable. As described earlier, the indicator variable $I(t)$ is binary by nature, that is,

$$I(t) = \begin{cases} 1 & \text{if system is in a failed state} \\ 0 & \text{otherwise} \end{cases}$$

Recall that the unavailability is the probability that a system is in the failed state, given that the system was initially operational at time zero. Hence the unavailability of a system is the probability that the indicator variable $I(t)$ takes the value of 1, which is equal to the expected value of $I(t)$. Accordingly,

$$U(t) = E[I(t)] = E[N_F(t)] - E[N_R(t)] = W(0, t) - \Gamma(0, t) \tag{5.46}$$

indicating that the unavailability is equal to the expected number of failures $W(0, t)$ minus the expected number of repairs $\Gamma(0, t)$ in time interval $(0, t]$. The values of $W(0, t)$ and $\Gamma(0, t)$ can be computed by Eqs. (5.43) and (5.44), respectively.

To compute $W(0, t)$ and $\Gamma(0, t)$, knowledge of the unconditional failure intensity $w(t)$ and the unconditional repair intensity $\gamma(t)$ is required. The unconditional failure intensity can be derived by the total probability theorem as

$$w(t) = f_t(t) + \int_0^t \gamma(\tau) f_t(t - \tau) \, d\tau \tag{5.47}$$

in which, on the right-hand side, the first term, $f_t(t)$, is for the case that the probability of failure is at time t, given that the system has survived up to time t; the second term accounts for the case that the system is repaired at time $\tau < t$ and later fails at time t. This is shown in Fig. 5.18.

For the unconditional repair intensity $\gamma(t)$ one would need only to consider one possible case, as shown in Fig. 5.19. That is, the system is in an operating state at time t given that the system is operational initially and is in a failed state at time $\tau < t$. The probability that this condition occurs is

$$\gamma(t) = \int_0^t w(\tau) g(t - \tau) \, d\tau \tag{5.48}$$
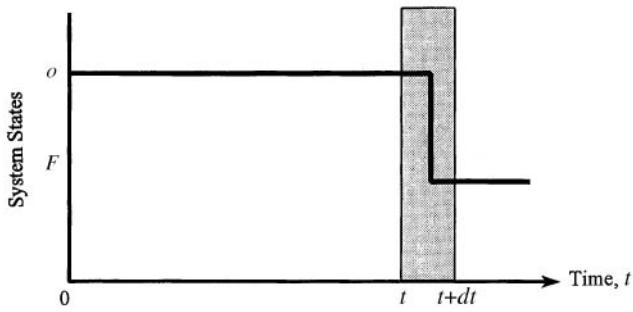
(a)
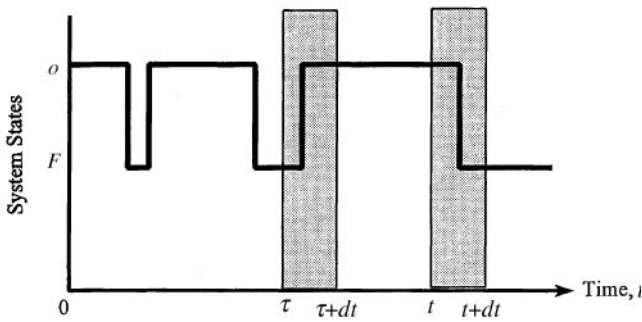
(b)

Figure 5.18 Two different cases for a system to be in a failed state during $(t, dt]$: (a) the system has been operational up to time t and failed during $(t, t + dt)$, given that it was good at $t = 0$ and no repair has been done during $(0, t)$; (b) the system has been operational up to time t and failed during $(t, t + dt)$, given that it was good at $t = 0$ and was repaired at $t = \tau$. (*After Henley and Kumamoto, 1981.*)
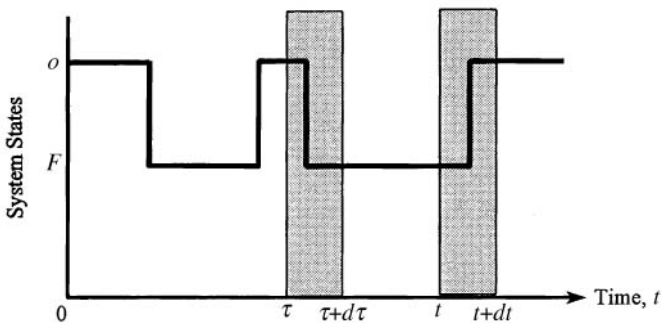
Figure 5.19 The case for a system to be repaired during $(t, dt]$. (*After Henley and Kumamoto, 1981.*)

Note that given the failure density $f_t(t)$ and the repair density $g_t(t)$, the unconditional failure intensity $w(t)$ and the unconditional repair intensity $\gamma(t)$ are related to one another in an implicit fashion, as shown in Eqs. (5.47) and (5.48). Hence the calculations of $w(t)$ and $\gamma(t)$ are solved by iterative numerical integration. Analytically, the Laplace transform technique can be applied to derive $w(t)$ and $\gamma(t)$ owing to the convolution nature of the two integrals.

Based on the unavailability and unconditional failure intensity, the conditional failure intensity $\mu(t)$ can be computed as

$$\mu(t) = \frac{w(t)}{1 - U(t)} = \frac{w(t)}{A(t)} \tag{5.49}$$

which is analogous to Eq. (5.3). For the repair process, the conditional repair intensity $\rho(t)$, unconditional repair intensity $\gamma(t)$, and unavailability are related as

$$\rho(t) = \frac{\gamma(t)}{U(t)} \tag{5.50}$$

The general relationships among the various parameters in the failure and repair processes are summarized in Table 5.4.

TABLE 5.4 Relationship among Parameters in Time-to-Failure Analysis

Repairable systems	Nonrepairable systems
General relations	
$A(t) + U(t) = 1$	$A(t) + U(t) = 1$
$A(t) > p_s(t)$	$A(t) = p_s(t)$
$U(t) < p_f(t)$	$U(t) = p_f(t)$
$w(t) = f_t(t) + \int f_t(t - \tau)\gamma(\tau)\,d\tau$	$w(t) = f_t(t)$
$\gamma(t) = \int g_t(t - \tau)w(\tau)\,d\tau$	$\gamma(t) = 0$
$W(t_1, t_2) = \int w(\tau)\,d\tau$	$W(t_1, t_2) = p_s(t_2) - p_s(t_1)$
$\Gamma(t_1, t_2) = \int \gamma(\tau)\,d\tau$	$\Gamma(t_1, t_2) = 0$
$U(t) = W(0, t) - \Gamma(0, t)$	$U(t) = p_f(t)$
$\mu(t) = w(t)/A(t)$	$h(t) = f_t(t)/p_s(t)$
$\rho(t) = \gamma(t)/U(t)$	$r(t) = 0$
Stationary values	
$\text{MTBF} = \text{MTBR} = \text{MTTF} + \text{MTTR}$	$\text{MTBF} = \text{MTBR} = \infty$
$0 < A(\infty), U(\infty) < 1$	$A(\infty) = 0, U(\infty) = 1$
$0 < w(\infty), \gamma(\infty) < \infty$	$w(\infty) = 0, \gamma(\infty) = 0$
$w(\infty) = \gamma(\infty)$	$w(\infty) = \gamma(\infty) = 0$
$W(0, \infty) = \Gamma(0, \infty) = \infty$	$W(0, \infty) = 1, \Gamma(0, \infty) = 0$
Remarks	
$w(t) \neq \mu(t), \gamma(t) \neq \rho(t)$	$w(t) \neq \mu(t), \gamma(t) = \rho(t) = 0$
$\mu(t) \neq h(t), \rho(t) \neq r(t)$	$\mu(t) = h(t), \rho(t) = r(t) = 0$
$w(t) \neq f_t(t), \gamma(t) \neq g_t(t)$	$w(t) = f_t(t), \gamma(t) = g_t(t) = 0$

SOURCE: After Henley and Kumamoto (1981).

Example 5.12 For a given failure density function $f_t(t)$ and repair density function $g_t(t)$, solve for the unconditional failure intensity $w(t)$ and the unconditional repair intensity $\gamma(t)$ by the Laplace transform technique.

Solution Note that the integrations in Eqs. (5.47) and (5.48) are in fact convolutions of two functions. According to the properties of the Laplace transform described in Appendix 5A, the Laplace transforms of Eqs. (5.47) and (5.48) result in the following two equations, respectively:

$$\mathscr{L}[w(t)] = \mathscr{L}[f_t(t)] + \mathscr{L}[\gamma(t)] \times \mathscr{L}[f_t(t)] \qquad (5.51a)$$

$$\mathscr{L}[\gamma(t)] = \mathscr{L}[w(t)] \times \mathscr{L}[g_t(t)] \qquad (5.51b)$$

in which $\mathscr{L}(\cdot)$ is the Laplace transform operator. Solving Eqs. (5.51a) and (5.51b) simultaneously, one has

$$\mathscr{L}[w(t)] = \frac{\mathscr{L}[f_t(t)]}{1 - \mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]} \qquad (5.52a)$$

$$\mathscr{L}[\gamma(t)] = \frac{\mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]}{1 - \mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]} \qquad (5.52b)$$

To derive $w(t)$ and $\gamma(t)$, the inverse transform can be applied to Eqs. (5.52a) and (5.52b), and the results are

$$w(t) = \mathscr{L}^{-1}\left\{ \frac{\mathscr{L}[f_t(t)]}{1 - \mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]} \right\} \qquad (5.53a)$$

and

$$\gamma(t) = \mathscr{L}^{-1}\left\{ \frac{\mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]}{1 - \mathscr{L}[f_t(t)] \times \mathscr{L}[g_t(t)]} \right\} \qquad (5.53b)$$

Example 5.13 (Constant failure rate and repair rate) Consider that the failure density function $f_t(t)$ and the repair density function $g_t(t)$ are both exponential distributions given as

$$f_t(t) = \lambda e^{-\lambda t} \qquad \text{for } \lambda \geq 0, t \geq 0$$

$$g_t(t) = \eta e^{-\eta t} \qquad \text{for } \eta \geq 0, t \geq 0$$

Derive the expressions for their availability and unavailability.

Solution The Laplace transform of the exponential failure density $f_t(t)$ is

$$\mathscr{L}[f_t(t)] = \int_0^\infty e^{-st}\, f_t(t)\, dt = \lambda \int_0^\infty e^{-(s+\lambda)t}\, dt = \frac{\lambda}{\lambda + s}$$

Similarly, $\mathscr{L}[g_t(t)] = \eta/(\eta + s)$. Substituting $\mathscr{L}[f_t(t)]$ and $\mathscr{L}[g_t(t)]$ into Eqs. (5.52a) and (5.52b), one has

$$\mathscr{L}[w(t)] = \frac{\lambda \eta}{\lambda + \eta}\left(\frac{1}{s}\right) + \frac{\lambda^2}{\lambda + \eta}\left(\frac{1}{s + \lambda + \eta}\right)$$

$$\mathscr{L}[\gamma(t)] = \frac{\lambda \eta}{\lambda + \eta}\left(\frac{1}{s}\right) \frac{\lambda \eta}{\lambda + \eta}\left(\frac{1}{s + \lambda + \eta}\right)$$

Taking the inverse transform for the preceding two equations, the results are

$$w(t) = \frac{\lambda\eta}{\lambda+\eta}\mathscr{L}^{-1}\left(\frac{1}{s}\right) + \frac{\lambda^2}{\lambda+\eta}\mathscr{L}^{-1}\left(\frac{1}{s+\lambda+\eta}\right)$$

$$\gamma(t) = \frac{\lambda\eta}{\lambda+\eta}\mathscr{L}^{-1}\left(\frac{1}{s}\right) - \frac{\lambda\eta}{\lambda+\eta}\mathscr{L}^{-1}\left(\frac{1}{s+\lambda+\eta}\right)$$

which can be finalized, after some algebraic manipulations, as

$$w(t) = \frac{\lambda\eta}{\lambda+\eta} + \frac{\lambda^2}{\lambda+\eta}e^{-(\lambda+\eta)t} \tag{5.54}$$

$$\gamma(t) = \frac{\lambda\eta}{\lambda+\eta} - \frac{\lambda\eta}{\lambda+\eta}e^{-(\lambda+\eta)t} \tag{5.55}$$

According to Eq. (5.43), the expected number of failures within time period $(0, t]$ can be calculated as

$$W(0, t) = \int_0^t w(\tau)\,d\tau = \left(\frac{\lambda\eta}{\lambda+\eta}\right)t + \frac{\lambda^2}{(\lambda+\eta)^2}(1 - e^{-(\lambda+\eta)t}) \tag{5.56}$$

Similarly, the expected number of repairs in time period $(0, t)$ is

$$\Gamma(0, t) = \int_0^t \gamma(\tau)\,d\tau = \left(\frac{\lambda\eta}{\lambda+\eta}\right)t - \frac{\lambda\eta}{(\lambda+\eta)^2}(1 - e^{-(\lambda+\eta)t}) \tag{5.57}$$

Once $W(0, t)$ and $\Gamma(0, t)$ are computed, the unavailability $U(t)$ can be determined, according to Eq. (5.46), as

$$U(t) = W(0, t) - \Gamma(0, t) = \frac{\lambda}{\lambda+\eta}(1 - e^{-(\lambda+\eta)t}) \tag{5.58}$$

The availability $A(t)$ then is

$$A(t) = 1 - U(t) = \frac{\eta}{\lambda+\eta} + \frac{\lambda}{\lambda+\eta}e^{-(\lambda+\eta)t} \tag{5.59}$$

As the time approaches infinity ($t \to \infty$), the system reaches its stationary condition. Then the stationary availability $A(\infty)$ and unavailability $U(\infty)$, are

$$A(\infty) = \frac{\eta}{\lambda+\eta} = \frac{1/\lambda}{1/\lambda + 1/\eta} = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} \tag{5.60}$$

$$U(\infty) = \frac{\lambda}{\lambda+\eta} = \frac{1/\eta}{1/\lambda + 1/\eta} = \frac{\text{MTTR}}{\text{MTTF} + \text{MTTR}} \tag{5.61}$$

Other properties for a system with constant failure and repair rates are summarized in Table 5.5. Results obtained in this example also can be derived based on the Markov analysis (Henley and Kumamoto, 1981; Ang and Tang, 1984).

Strictly speaking, the preceding expressions for the availability are the inherent availability under the condition of an ideal supportability with which the mean time to support (MTTS) is zero. In the case that the failed system requires some time to respond and prepare before the repair task is undertaken, the actual availability is

$$A(\infty) = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR} + \text{MTTS}} \tag{5.62}$$

which, as compared with Eq. (5.60), is less than the inherent availability.

TABLE 5.5 Summary of the Constant Rate Model

Repairable systems	Nonrepairable systems
Failure process	
$h(t) = \lambda$	$h(t) = \lambda$
$p_s(t) = e^{-\lambda t}$	$p_s(t) = e^{-\lambda t}$
$p_f(t) = 1 - e^{-\lambda t}$	$p_f(t) = 1 - e^{-\lambda t}$
$f_t(t) = \lambda e^{-\lambda t}$	$f_t(t) = \lambda e^{-\lambda t}$
MTTF $= 1/\lambda$	MTTF $= 1/\lambda$
Repair process	
$r(t) = \eta$	$r(t) = 0$
$G_t(t) = 1 - e^{-\eta t}$	$G_t(t) = 0$
$g_t(t) = \eta e^{-\eta t}$	$g_t(t) = 0$
MTTR $= 1/\eta$	MTTR $= \infty$
Dynamic behavior of whole process	
$U(t) = \lambda/(\lambda + \eta)(1 - e^{-(\lambda+\eta)t})$	$U(t) = 1 - e^{-\lambda t} = p_f(t)$
$A(t) = \eta/(\lambda + \eta) + \lambda/(\lambda + \eta)(1 - e^{-(\lambda+\eta)t})$	$A(t) = e^{-\lambda t} = p_s(t)$
$\omega(t) = \lambda\eta/(\lambda + \eta) + \lambda^2/(\lambda + \eta)(1 - e^{-(\lambda+\eta)t})$	$w(t) = f_t(t) = \lambda e^{-\lambda t}$
$\gamma(t) = \lambda\eta/(\lambda + \eta)(1 - e^{-(\lambda+\eta)t})$	$\gamma(t) = 0$
$W(0, t) = \lambda\eta t/(\lambda + \eta) + \lambda^2/(\lambda + \eta)(1 - e^{-(\lambda+\eta)t})$	$W(0, t) = p_f(t)$
$\Gamma(0, t) = \lambda\eta t/(\lambda + \eta) - \lambda\eta/(\lambda + \eta)^2(1 - e^{-(\lambda+\eta)t})$	$\Gamma(0, t) = 0$
Stationary values of whole process	
$U(\infty) = \lambda/(\lambda + \eta) = $ MTTR/(MTTF + MTTR)	$U(\infty) = 1$
$A(\infty) = \eta/(\lambda + \eta) = $ MTTF/(MTTF + MTTR)	$A(\infty) = 0$
$w(\infty) = \lambda\eta/(\lambda + \eta) = 1/($MTTF + MTTR$)$	$w(\infty) = 0$
$\gamma(\infty) = \lambda\eta/(\lambda + \eta) = w(\infty)$	$\gamma(\infty) = 0$

SOURCE: After Henley and Kumamoto (1981).

Example 5.14 Referring to Example 5.12, with exponential failure and repair density functions, determine the availability and unavailability of the pump.

Solution Since the failure density and repair density functions are both exponential, the unavailability $U(t)$ of the pump, according to Eq. (5.58), is

$$U(t) = \frac{\lambda}{\lambda + \eta}(1 - e^{-(\lambda+\eta)t}) = \frac{0.0008}{0.0008 + 0.02}(1 - e^{-0.0208t})$$

$$= 0.03846(1 - e^{-0.0208t})$$

The availability $A(t)$ then is

$$A(t) = 1 - U(t) = 0.9615 + 0.03846e^{-0.0208t}$$

The stationary availability and unavailability are

$$A(\infty) = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} = \frac{1/\lambda}{1/\lambda + 1/\eta} = \frac{1250}{1250 + 50} = 0.96154$$

$$U(\infty) = 1 - A(\infty) = 1 - 0.96154 = 0.03846$$

TABLE 5.6   Operation Properties of the Laplace Transform on a Function

Property	Function	Variable	Laplace transform
Standard	$f_x(x)$	X	$\mathscr{L}_x(s)$
Scaling	$f_x(ax)$	X	$a^{-1}\mathscr{L}_x(s/a)$
Linear	$af_x(x)$	X	$a\mathscr{L}_x(s)$
Translation-1	$e^{ax}f_x(x)$	X	$\mathscr{L}_x(s+a)$
Translation-2	$f_x(x-a)$	X	$e^{as}\mathscr{L}_x(s), x > a$

Appendix 5A: Laplace Transform*

The *Laplace transforms* of a function $f_x(x)$ are defined, respectively, as

$$\mathscr{L}_x(s) = \int_{-\infty}^{\infty} e^{sx} f_x(x)\,dx \tag{5A.1}$$

In a case where $f_x(x)$ is the PDF of a random variable, the Laplace transform defined in Eqs. (5A.1) can be stated as

$$\mathscr{L}_x(s) = E[e^{sX}] \qquad \text{for } x \geq 0 \tag{5A.2}$$

Useful operational properties of the Laplace transform on a PDF are given in Table 5.6. The transformed function given by Eq. (5A.1) of a PDF is called the *moment-generating function* (MGF) and is shown in Table 5.7 for some commonly used probability distribution functions. Some useful operational rules relevant to the Laplace transform are given in Table 5.8.

---

*Extracted from Tung and Yen (2005).

TABLE 5.7   Laplace Transform (Moment-Generating Functions) of Some Commonly Used Distribution Functions

Distribution	PDF	Laplace transform
Uniform	Eq. (2.100)	$\dfrac{e^{bs} - e^{as}}{(b-a)s}$
Normal	Eq. (2.58)	$\exp(\mu s - 0.5s^2\sigma^2)$
Gamma	Eq. (2.72)	$\left[\dfrac{1/\beta}{(1/\beta)-s}\right]^{\alpha}$
Exponential	Eq. (2.79)	$\dfrac{1/\beta}{(1/\beta)-s}$
Extreme value I (max)	Eq. (2.85)	$e^{\xi s}\Gamma(1-\beta s)$
Chi-square	Eq. (2.102)	$(1-2s)^{-K/2}$

TABLE 5.8 Operational Rules for the Laplace Transform

$W = cX$	$\mathscr{L}_w(s) = \mathscr{L}_x(cs)$, $c = a$ constant.
$W = c + X$	$\mathscr{L}_w(s) = e^{cs}\mathscr{L}_x(s)$, $c = a$ constant.
$W = \Sigma_k X_k$	$\mathscr{L}_w(s) = \Pi_k \mathscr{L}_k(s)$, when all X_k are independent.
$W = \Sigma_k c_k X_k$	$\mathscr{L}_w(s) = \Pi_k \mathscr{L}_k(c_k s)$, when all X_k are independent.

Problems

5.1 Consider the following hazard function:

$$h(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[\left(\frac{t}{\alpha}\right)^{\beta}\right]$$

Derive the expressions for the corresponding failure density function and reliability function.

5.2 Refer to Problem 5.1, and let $\tau = t/\alpha$. Obtain the expression for $h(\tau)$, and plot $h(\tau)$ versus τ for $\beta = 1.0, 0.5, 0.2$.

5.3 Fifty pumps of identical models are tested. The following table contains data on the pump failures from the test. Propose a scheme to determine the parameter values in the failure-rate function given in Problem 5.1.

No. of failures	Time to failure (h)	No. of failures	Time to failure (h)
1	26	16	3400
2	65	18	4000
3	300	20	4400
4	700	22	4500
5	900	24	4600
6	1100	26	4800
7	1200	28	5000
8	1500	32	5500
9	1700	36	6000
10	1900	40	6500
12	2400	44	7000
14	2800	50	7400

5.4 Consider the following hazard function:

$$h(t) = \alpha t + \frac{\gamma}{1 + \beta t} \qquad t \geq 0, \ \alpha, \beta, \gamma \geq 0$$

Derive the expressions for the corresponding failure density function and reliability function.

5.5 The failure-rate function given in Problem 5.4 is a very versatile function, capable of modeling increasing, decreasing, or a bathtub failure-rate behavior of a system. For example, it corresponds to the Rayleigh failure density function for $\gamma = 0$ and to the exponential failure density for $\alpha = \beta = 0$. Furthermore, when $\alpha = 0$, $h(t)$ is a decreasing function; when $\alpha \geq \beta\gamma$, $h(t)$ is an increasing failure rate; for $0 < \alpha < \beta\gamma$, $h(t)$ is a bathtub failure rate. Plot the failure-rate function for (a) $\alpha = 0, \gamma = 2, \beta = 3$; (b) $\alpha = 1, \gamma = 2, \beta = 3$; and (c) $\alpha = 6, \gamma = 2, \beta = 3$.

5.6   Given the following failure density function for a piece of equipment:

$$f_t(t) = \frac{0.7}{850} \left(\frac{t}{850} \right)^{-0.3} \exp\left[-\left(\frac{t}{850} \right)^{0.7} \right]$$

derive the corresponding reliability function $p_s(t)$ and hazard function $h(t)$. Also, construct plots for $f_t(t)$, $p_s(t)$, and $h(t)$.

5.7   Repeat Example 5.3 to derive the expressions for the failure rate, reliability, and failure density for a 10-mile water main of sandspun cast iron pipe. Compare the results with those from Example 5.3.

5.8   Refer to Example 5.3. Assume that the pipe main made of sandspun cast iron is 5 years old. Let x be the length of the pipe system. Derive the expressions for the failure rate, failure density, and reliability for a system with the total pipe length. Construct curves for the three quantities as a function of the pipe length.

5.9   Determine the mean length to failure based on the results from Problem 5.8.

5.10 Repeat Example 5.3 for pit cast iron pipe.

5.11 Repeat Problems 5.8 and 5.9 for pit cast iron pipe.

5.12 Consider the sandspun cast iron pipe in Example 5.3. The pipe break rate is a function of the age and length of the pipe, which can be generalized as

$$N(x, t) = 0.0627 x e^{0.0137t}$$

Derive the expressions for the failure density, reliability, and failure rate as a function of the age and pipe length. Furthermore, construct figures for the three functions for different values of x and t.

5.13 Goodrich et al. (1989) presented the following break-rate equation for a cast iron pipe at St. Louis, MO, based on the 1985 pipe break data:

$$h(d) = 0.819 e^{0.1363d}$$

in which $h(d)$ is the break rate (in breaks per mile per year), and d is pipe diameter (in inches). Assume that a brand new pipe system is to be installed that follows the given break-rate function. Derive expressions for the failure density function, reliability, and failure rate in terms of the size, age, and length of the pipe for St. Louis.

5.14 Walski and Pelliccia (1982) also developed a regression equation for the average time required to repair pipe breaks:

$$t_r = 6.5 d^{0.285}$$

where t_r is the time to repair in hours per break, and d is pipe diameter in inches. Assume that the preceding regression equation has a 25 percent error associated with it and that the time to repair is lognormally distributed. Derive the expressions for the repair density function, repair probability, and repair rate.

5.15 Refer to Example 5.3 and Problem 5.14. Compute the MTBF for a 5-year-old, 12-inch, 5-mile-long sandspun cast iron pipe.

5.16 Given the following unreliability and repair probabilities:

$$p_f(t) = 1 - \frac{8}{7}e^{-t} + \frac{1}{7}e^{-8t}$$

$$G(t) = 1 - e^{-6t}$$

derive the failure density, failure rate, repair density, and repair rate.

5.17 Assume that a system has an ideal preventive maintenance with a regular inspection interval of t_M. The failure density function follows a Weibull distribution (with $t_o = 0$), as defined in Table 5.1. Derive the expressions, under the maintenance condition, for the reliability function, failure density, hazard function, mean time to failure, and mean and variance for the number of scheduled maintenances before failure.

5.18 Based on the reliability function derived in Problem 5.17 for the system with preventive maintenance, compare it with the reliability function without maintenance, and derive the condition under which the ideal maintenance is beneficial. Furthermore, derive the condition under which faulty maintenance is desirable.

5.19 Find the condition under which preventive maintenance would be beneficial for a system having the following failure density function (after Rao, 1992):

$$f_t(t) = \alpha^2 t e^{-\alpha t} \qquad t \geq 0$$

if (a) the maintenance is ideal and (b) the maintenance is imperfect.

5.20 A turbine is known to sustain two types of failures: bearing failure and blade failure. The bearing failure times follow an exponential distribution, with a failure rate of 0.0005 per hour, and the blade failure times follow the following Weibull distribution:

$$f_t(t) = \frac{1}{100}\left(\frac{t}{200}\right)\exp\left[-\left(\frac{t}{200}\right)^2\right] \qquad t \geq 0$$

   (a) Find the reliability of the unmaintained turbine system after 10,000 hours of operation.

   (b) If the reliability of the turbine is to be increased by 20 percent at the end of the 10,000-hour operating period by replacing the turbine blades at times $t_M, 2t_M, 3t_M, 4t_M, \ldots$, find the value of t_M.

5.21 Suppose that there is a new water main, 2-miles long, conveying raw water from the source to the treatment plant. Let the break rate of the pipe be defined by the function in Problem 5.12.

   (a) Derive the unmaintained reliability function and the MTTF.

   (b) Derive the reliability function and the MTTF under the condition that the pipe has a scheduled maintenance of 6 months.

   (c) Determine the inspection frequency such that the reliability of the pipe is 25 percent higher than that of the unmaintained reliability at the end of the fifth year.

5.22 Based on the results from Problem 5.16, derive the expressions for the unconditional failure intensity $w(t)$ and unconditional repair intensity $r(t)$.

5.23 Based on the results from Problem 5.22, derive the availability, unavailability, and average availability over period $(0, T]$.

5.24 Given the following failure density and repair density functions:

$$f_t(t) = \frac{1}{2}(e^{-t} + 3e^{-3t})$$

$$g_t(t) = 1.5e^{-1.5t}$$

derive the expressions for the availability, unavailability, and average availability over period $(0, T]$.

5.25 Show that the average availability for a system with constant failure rate λ and repair rate η with $A(0) = 1$ is

$$\overline{A}(0, T) = \frac{\eta}{\lambda + \eta} + \frac{\lambda}{(\lambda + \eta)^2 T} - \frac{\lambda}{(\lambda + \eta)^2 T} e^{-(\lambda + \eta)T}$$

References

Ang, A. H. S., and Tang, W. H. (1984). *Probability Concepts in Engineering Planning and Design,* Vol. II: *Decision, Risk, and Reliability*, John Wiley and Sons, New York.

Arthur, H. G. (1977). Teton Dam failure, *The Evaluation of Dam Safety*, ASCE, New York, pp. 61–68.

Barr, D. W., and Heuer, K. L. (1989). Quick response on the Mississippi, *Civil Engineering*, ASCE. 59(9):50–52.

Goodrich, J., Mays, L. W., Su, Yu-Chun, and Woodburn, J. (1989). Data base management systems, in *Reliability Analysis of Water Distribution Systems*, ed. by L. W. Mays, ASCE, New York, pp. 123–149.

Harr, M. E. (1987). *Reliability-Based Design in Civil Engineering*, McGraw-Hill, New York.

Henley, E. J., and Kumamoto, H. (1981). *Reliability Engineering and Risk Assessment*. Prentice-Hall, Englewood Cliffs, NJ.

Jansen, R. B. (1988). Dam safety in America, *Hydro Review*, 17(3):10–20.

Kapur, K. C. (1988a). Technique of estimating reliability at design stage, in *Handbook of Reliability Engineering and Management*, ed. by W. G. Ireson and C. F. Crombs, Jr., McGraw-Hill, New York.

Kapur, K. C. (1988b). Mathematical and statistical methods and models in reliability and life studies, in *Handbook of Reliability Engineering and Management*, ed. by W. G. Ireson and C. F. Crombs, Jr., McGraw-Hill, New York.

Kapur, K. C., and Lamberson, L. R. (1977). *Reliability in Engineering Design*, John Wiley and Sons, New York.

Kraus, J. W. (1988). Maintainability and reliability, in *Handbook of Reliability Engineering and Management*, ed. by W. G. Ireson and C. F. Crombs, Jr., McGraw-Hill, New York.

Knezevic, J. (1993). *Reliability, Maintainability, and Supportability*, McGraw-Hill, New York.

Mays, L. W. and Tung, Y. K. (1992). *Hydrosystems Engineering and Management*. McGraw-Hill, New York.

O'Connor, P. D. T. (1981). *Practical Reliability Engineering*, John Wiley and Sons, New York.

Pieruschka, E. (1963). *Principles of Reliability*, Prentice-Hall, Englewood Cliffs, NJ.

Ramakumar, R. (1993). *Engineering Reliability: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ.

Rao, S. S. (1992). *Reliability-Based Designs*, McGraw-Hill, New York.

Shultz, D. W., and Parr, V. B. (1981). Evaluation and documentation of mechanical reliability of conventional wastewater treatment plant components, *Report*, U.S. Environmental Protection Agency, Cincinnati, OH.

Tobias, P. A., and Trindade, D. C. (1995). *Applied Reliability*, 2nd ed., Van Nostrand Reinhold, New York.

Walski, T. M., and Pelliccia, A. (1982). Economic analysis of water main breaks, *Journal of the American Water Works Association*, 79(3):140–147.

Wunderlich, W. O. (1993). Probabilistic methods for maintenance of hydraulic structures, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. B. C. Yen and Y. K. Tung, ASCE, New York, pp. 191–206.

Wunderlich, W. O. (2004). Hydraulic Structures: Probabilistic Approaches to Maintenance, ASCE, Restor, VA.

This page intentionally left blank

6

Monte Carlo Simulation

6.1 Introduction

As uncertainty and reliability related issues are becoming more critical in engineering design and analysis, proper assessment of the probabilistic behavior of an engineering system is essential. The true distribution for the system response subject to parameter uncertainty should be derived, if possible. However, owing to the complexity of physical systems and mathematical functions, derivation of the exact solution for the probabilistic characteristics of the system response is difficult, if not impossible. In such cases, Monte Carlo simulation is a viable tool to provide numerical estimations of the stochastic features of the system response.

Simulation is a process of replicating the real world based on a set of assumptions and conceived models of reality (Ang and Tang, 1984, pp. 274–332). Since the purpose of a simulation model is to duplicate reality, it is an effective tool for evaluating the effects of different designs on a system's performance. *Monte Carlo simulation* is a numerical procedure to reproduce random variables that preserve the specified distributional properties. In Monte Carlo simulation, the system response of interest is repeatedly measured under various system parameter sets generated from known or assumed probabilistic laws. It offers a practical approach to uncertainty analysis because the random behavior of the system response can be duplicated probabilistically.

Two major concerns in practical applications of Monte Carlo simulation in uncertainty and reliability analyses are (1) the requirement of a large number of computations for generating random variates and (2) the presence of correlation among stochastic basic parameters. However, as computing power increases, the concern with the computation cost diminishes, and Monte Carlo simulations are becoming more practical and viable for uncertainty analyses. In fact, Beck (1985) notes that "when the computing power is available, there can, in general, be no strong argument against the use of Monte Carlo simulation."

As noted previously, the accuracy of the model output statistics and probability distribution (e.g., probability that a specified safety level will be exceeded) obtained from Monte Carlo simulation is a function of the number of simulations performed. For models or problems with a large number of uncertain basic variables and for which low probabilities (<0.1) are of interest, tens of thousands of simulations may be required. Rules for determining the number of simulations required for convergence are not available, and thus replication of the Monte Carlo simulation runs for a given number of simulations is the only way to check convergence (Melching, 1995). Cheng et al. (1982) considered the convergence characteristics of Monte Carlo simulation for a simple case of $Z = X_3 X_4 - (X_1 + X_2)$, where the distributions and statistics of the variables are listed in Table 6.1. They found that failure probabilities (i.e., probability of $Z < 0$) down to 0.0025 could be estimated reliably with 32,000 simulations, failure probabilities down to 0.015 could be estimated reliably with 8000 simulations, and failure probabilities down to 0.2 could be estimated reliably with 1000 simulations.

Problems involving more complex system functions Z and more basic variables may require more simulations to obtain similar accuracy. For example, Melching (1992) found that 1000 simulations were adequate to estimate the mean, standard deviation, and quantiles above 0.2 for an application of the HEC-1 (U.S. Army Corps of Engineers, 1990) and RORB (Laurenson and Mein, 1985) rainfall-runoff models and that 10,000 simulations were needed to accurately estimate quantiles between 0.01 and 0.2. Brown and Barnwell (1987) reported that for the QUAL2E multiple-constituent (dissolved oxygen, nitrogen cycle, algae, etc.) steady-state surface water-quality model, 2000 simulations were required to obtain accurate estimates of the output standard deviation. With the computational speed of today's computers, making even 10,000 runs is not prohibitive for simpler models. However, increased computational speed has made possible the use of computational fluid dynamics codes in three dimensions for hydrosystems design work. When such codes are applied, the variance-reduction techniques described in Sec. 6.7 may be preferred to Monte Carlo simulation.

This chapter focuses on the basic principles and applications of Monte Carlo simulations in the reliability analysis of hydrosystems engineering problems. Section 6.2 describes some basic concepts of generating random numbers, followed by discussions on the classifications of algorithms for a generation of random variates in Sec. 6.3. Algorithms for generating univariate random numbers

TABLE 6.1   Basic Variable Statistics and Distributions for Evaluation of Monte Carlo Simulation of Convergence

Variable	Mean value	Coefficient of variation	Distribution function
X_1	0.5	0.2	Uniform
X_2	1.5	0.4	Uniform
X_3	1.0	0.005	Lognormal
X_4	1.5	0.1	Lognormal

SOURCE: After Cheng et al. (1982).

are described in Sec. 6.4 for several commonly used distribution functions. In Sec. 6.5, attention is given to algorithms that generate multivariate random numbers. As reliability assessment involves mathematical integration, Sec. 6.6 describes several Monte Carlo simulation techniques for reliability evaluation. Given that Monte Carlo simulations, in essence, are sampling techniques, they provide only estimations, which inevitably are subject to certain degrees of errors. To improve the accuracy of the Monte Carlo estimation while reducing excessive computational time, several variance-reduction techniques are discussed in Sec. 6.7. Finally, resampling techniques are described in Sec. 6.8, which allow for assessment of the uncertainty of the quantity of interest based on the available random data without having to make assumptions about the underlying probabilistic structures.

6.2 Generation of Random Numbers

The most commonly used techniques to generate a sequence of pseudorandom numbers are those that apply some form of recursive computation. In principle, such recursive formulas are based on calculating the residuals modulo of some integers of a linear transformation. The process of producing a random number sequence is completely deterministic. However, the generated sequence would appear to be uniformly distributed and independent.

Congruential methods for generating n random numbers are based on the fundamental congruence relationship, which can be expressed as (Lehmer, 1951)

$$X_{i+1} = \{aX_i + c\}(\text{mod } m) \qquad i = 1, 2, \ldots, n \tag{6.1}$$

in which a is the multiplier, c is the increment, and m is an integer-valued modulus. The modulo notation (mod m) in Eq. (6.1) represents that

$$X_{i+1} = aX_i + c - mI_i \tag{6.2}$$

with $I_i = [(aX_i + c)/m]$ denoting the largest positive integer value in $(aX_i + c)/m$. In other words, X_{i+1} determined by Eq. (6.1) is the residual resulting from $(aX_i + c)/m$. Therefore, the values of the number sequence generated by Eq. (6.1) would satisfy $X_i < m$, for all $i = 1, 2, \ldots, n$. Random number generators that produce a number sequence according to Eq. (6.1) are called *mixed congruential generators*.

Applying Eq. (6.1) to generate a random number sequence requires the specification of a, c, and m, along with X_0, called the *seed*. Once the sequence of random number Xs are generated, the random number from the unit interval $u_i \in [0, 1]$ can be obtained as

$$U_i = \frac{X_i}{m} \qquad i = 1, 2, \ldots, n \tag{6.3}$$

It should be pointed out that the process of generating uniform random numbers is the building block in Monte Carlo simulation.

Owing to the deterministic nature of the number generation, it is clear that the number sequence produced by Eq. (6.1) is periodic, which will repeat itself

in, at most, m steps. This implies that the sequence would contain, at most, m distinct numbers and will have a maximum period of length $m - 1$ beyond which the sequence will get into a loop. For example, consider $X_{i+1} = 2X_i + 3(\mathrm{mod}\ m = 5)$, with $X_0 = 3$; the number sequence generated would be $4, 1, 0, 3, 4, 1, 0, \ldots$.

From the practical application viewpoint, it is desirable that the generated number sequence have a very long periodicity to ensure that sufficiently large amounts of distinct numbers are produced before the cycle occurs. Therefore, one would choose the value of the modulus m to be as large as possible. However, the length of the periodicity in a sequence also depends on the values of multiplier a and increment c. Knuth (1981) derived three conditions under which a sequence from Eq. (6.1) has a *full period m*. Based on the three conditions of Knuth (1981), Rubinstein (1981) showed that for a computer with a binary digit system, using $m = 2^\beta$, with β being the word length of the computer, along with an odd number for parameter c and $a = 2^r + 1$, $r \geq 2$ would produce a full period sequence. The literature (Hull and Dobell, 1964; MacLaren and Marsagalia, 1965; Olmstead, 1946) indicates that good statistical results can be achieved by using $m = 2^{35}$, $a = 2^7 + 1$, and $c = 1$. Table 6.2 lists suggested values for the parameters in Eq. (6.1) for different computers.

TABLE 6.2   Suggested Values for Parameters in Congruential Methods

Constants for portable random number generators							
Overflow at	m	a	c	Overflow at	m	a	c
2^{20}	6075	106	1283	2^{28}	117128	1277	24749
2^{21}	7875	211	1663		312500	741	66037
2^{22}	7875	421	1663		121500	2041	25673
2^{23}	11979	430	2531	2^{29}	120050	2311	25367
	6655	936	1399		214326	1807	45289
	6075	1366	1283		244944	1597	51749
2^{24}	53125	171	11213		233280	1861	49297
	11979	859	2531		175000	2661	36979
	29282	419	6173		121500	4081	25673
	14406	967	3041		145800	3661	30809
2^{25}	134456	141	28411	2^{30}	139968	3877	29573
	31104	625	6571		214326	3613	45289
	14000	1741	2957		714025	1366	150889
	12960	1741	2731	2^{31}	134456	8121	28411
	21870	1291	4621		243000	4561	51349
	139968	205	29573		259200	7141	54773
2^{26}	81000	421	17117	2^{32}	233280	9301	49297
	29282	1255	6173		714025	4096	150889
	134456	281	27411	2^{33}	1771875	2416	374441
2^{27}	86463	1093	18257	2^{34}	510300	17221	107839
	259200	421	54773		312500	36261	66037
	116640	1021	24631	2^{35}	217728	84589	45989
	121500	1021	25673				

SOURCE: After Press et al. (1989).

A second commonly used generator is called the *multiplicative generator*:

$$X_{i+1} = \{aX_i\}(\bmod\ m) \qquad i = 1, 2, \ldots, n \tag{6.4}$$

which is a special case of the mixed generator with $c = 0$. Knuth (1981) showed that a maximal period can be achieved for the multiplicative generator in a binary computer system when $m = 2^\beta$ and $a = 8r \pm 3$, with r being any positive integer.

Another type of generator is called the *additive congruential generator* having the recursive relationship as

$$X_{i+1} = \{X_i + X_{i-t}\}(\bmod\ m) \qquad t = 1, 2, \ldots, i - 1 \tag{6.5}$$

As can be seen, the random numbers generated by the additive congruential generator depend on more than one of its preceding values. When $t = 1$, Eq. (6.5) would generate a sequence of Fibonacci numbers, which are not satisfactorily random. However, the statistical properties improve as t gets larger.

In summary, to ensure that a sequence of random numbers generated by the congruential methods would have satisfactory statistical properties, Knuth (1981) recommended the following principles to choose the parameters a, c, m, and X_0.

1. The seed X_0 can be chosen arbitrarily. If different random number sequences are to be generated, a practical way is to set X_0 equal to the date and time when the sequence is to be generated.

2. The modulus m must be large. It may be set conveniently to the word length of the computer because this would enhance computational efficiency. The computation of $\{aX + c\}(\bmod\ m)$ must be done exactly without round-off errors.

3. If modulus m is a power of 2 (for binary computers), select the multiplier a so that $a(\bmod\ 8) = 5$. If m is a power of 10 (for decimal computers), pick a such that $a(\bmod\ 200) = 21$. Selection of the multiplier a in this fashion, along with the choice of increment c described below, would ensure that the random number generator will produce all m distinct possible values in the sequence before repeating itself.

4. The multiplier a should be larger than $\sqrt{m}$, preferably larger than $m/100$, but smaller than $m - \sqrt{m}$. The best policy is to take some haphazard constant to be the multiplier satisfying both conditions 3 and 4.

5. The increment parameter c should be an odd number when the modulus m is a power of 2 and c should not be a multiple of 5 when m is a power of 10.

6.3 Classifications of Random Variates Generation Algorithms

6.3.1 CDF-inverse method

Let a random variable X have the cumulative distribution function (CDF) $F_x(x)$. From Sec. 2.3.1, $F_x(x)$ is a nondecreasing function with respect to the value of x, and $0 \le F_x(x) \le 1$. Therefore, $F_x^{-1}(u)$ may be defined for any value of u between 0 and 1 as $F_x^{-1}(u)$ is the smallest x satisfying $F_x(x) \ge u$.

For the majority of continuous probability distributions applied in hydrosystems engineering and analysis, $F_x(x)$ is a strictly increasing function of x. Hence a unique relationship exists between $F_x(x)$ and u; that is, $u = F_x(x)$, as shown in Fig. 6.1. Furthermore, it can be shown that if U is a standard uniform random variable defined over the unit interval [0, 1], denoted by $U \sim U(0, 1)$, the following relationship holds:

$$X = F_x^{-1}(U) \tag{6.6}$$

Note that X is a random variable because it is a function of the random variable U. From Eq. (6.6), the one-to-one correspondence between X and U, through the CDF, enables the generation of random numbers $X \sim F_x(x)$ from the standard uniform random numbers. The algorithm using the CDF-inverse method for generating continuous random numbers from a CDF $F_x(x)$ can be stated as follows:

1. Generate n uniform random numbers $u_1, u_2, \ldots, u_n$ from $U(0, 1)$.

2. Solve for $x_i = F_x^{-1}(u_i)$, for $i = 1, 2, \ldots, n$.
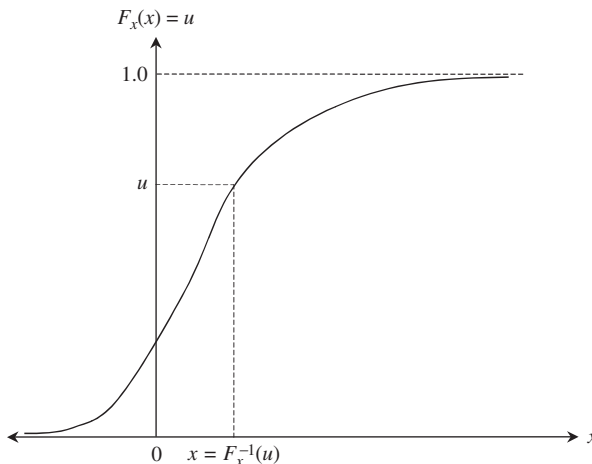
Figure 6.1 Schematic diagram of the inverse-CDF method for generating random variates.

Example 6.1 Consider that the Manning roughness coefficient X of a cast iron pipe is uncertain, having a uniform distribution $f_x(x) = 1/(b - a)$, $a \leq x \leq b$. Develop an algorithm using the CDF-inverse method to generate a random Manning roughness coefficient.

Solution Using the CDF-inverse method, the expression of the CDF of the random variable is first sought. The CDF for this example can be derived as

$$F_x(x) = \int_a^x \frac{1}{b-a} \, dx = \frac{x-a}{b-a} \qquad \text{for } a \leq x \leq b$$

From this expression for the CDF, the random variate of Manning's roughness coefficient x is obtained, in terms of $F_x(x)$, as

$$x = a + (b - a)F_x(x)$$

A simple algorithm for generating uniform random variates from U(a, b) is

1. Generate n standard uniform random variates $u_1, u_2, \ldots, u_n$ from U(0, 1).
2. Calculate the corresponding uniform random variates $x_i = a + (b - a)u_i$, $i = 1, 2, \ldots, n$.

In the case that the random variables under consideration are discrete, the value of x_j corresponding to the generated standard uniform random variate u must satisfy

$$F_x(x_{j-1}) = \sum_{i=1}^{j-1} f_x(x_i) < u \leq F_x(x_j) = \sum_{i=1}^{j} f_x(x_i) \qquad (6.7)$$

The CDF-inverse algorithm for generating discrete random variates can be implemented as follows:

1. Generate the uniform random number u from U(0, 1).
2. Initialize $i = 0$ and set $p = 0$.
3. Let $i = i + 1$, and compute $p = p + f_x(x_i)$.
4. If $p < u$, go to step 3; otherwise, stop, and x_i is the random variate sought.

Example 6.2 Suppose that the number of snow storms X at a location in January has a discrete uniform distribution

$$f_x(x) = 1/5 \qquad \text{for } x = 0, 1, 2, 3, 4$$

Develop an algorithm to generate a sequence of random number of snow storms.

Solution The CDF for the number of snow storms can be written as

$$F_x(x) = (x + 1)/5 \qquad \text{for } x = 0, 1, 2, 3, 4$$

The algorithm for this example can be outlined as follows.

1. Generate the uniform random number u from U(0, 1).
2. Initialize $x = 0$ and $p_1 = 0$, and compute $F_x(0)$.

3. Test if $p_1 < u \le F_x(x)$. If yes, x is the solution; otherwise, go to step 4.

4. Let $p_1 = F_x(x)$, $x = x + 1$, and compute $F_x(x + 1)$. Go to step 3.

To apply the CDF-inverse method for generating random numbers efficiently, an explicit expression between X and U is essential so that X can be obtained analytically from the generated U. The distributions the inverse forms of which are analytically expressible include exponential, uniform, Weibull, and Gumbel. Table 6.3 lists some distributions that are used in hydrosystems the CDF inverses of which are analytically expressible.

When the analytical forms of the CDF inverse are not available, applying the CDF-inverse method would require solving

$$u = \int_{-\infty}^{x} f_x(t)\, dt \tag{6.8}$$

for x from the known u. For many commonly used distributions such as normal, lognormal, and gamma, solving Eq. (6.8) is inefficient and difficult. More efficient algorithms have been developed to generate random variates from those distributions; some of these are described in Sec. 6.4.

6.3.2 Acceptance-rejection methods

Consider a problem for which random variates are to be generated from a specified probability density function (PDF) $f_x(x)$. The basic idea of the

TABLE 6.3 List of Distributions the Cumulative Distribution Function (CDF) Inverses of which Are Analytically Expressible

Distribution	$F_x(x) =$	$x = F_x^{-1}(u)$
Exponential	$1 - \exp(-\beta x), x > 0$	$-\beta \ln(1 - F)$
Uniform	$(x - a)/(b - a)$	$a + (b - a)F$
Gumbel	$\exp\{-\exp[-(x - \xi)/\beta)]\}$	$\xi - \beta \ln[-\ln(F)]$
Weibull	$1 - \exp\{-[(x - \xi)/\beta]^{\alpha}\}$	$\xi + \beta[-\ln(1 - F)]^{1/\alpha}$
Pareto	$1 - x^{-\alpha}$	$(1 - F)^{-(1/\alpha)}$
Wakeby	Not explicitly defined	$\xi + (\alpha/\beta)[1 - (1 - F)^{\beta}] - (\gamma/\delta)[1 - (1 - F)^{-\delta}]$
Kappa	$\{1 - h[1 - \alpha(x - \xi)/\beta]^{1/\alpha}\}^{1/h}$	$\xi + (\beta/\alpha)\{1 - [(1 - F^h)/h]^{\alpha}\}$
Burr	$1 - (1 + x^{\alpha})^{-\beta}$	$[(1 - F)^{-1/\beta} - 1]^{1/\alpha}$
Cauchy	$0.5 + \tan^{-1}(x)/\pi$	$\tan[\pi(F - 0.5)]$
Rayleigh	$1 - \exp[-(x - \xi)^2/2\beta^2]$	$\xi + \{-2\beta^2 \ln(1 - F)\}^{1/2}$
Generalized lambda	Not explicitly defined	$\xi + \alpha F^{\beta} - \gamma(1 - F)^{\delta}$
Generalized extreme value	$\exp[-\exp(-y)]$ where $y = -\alpha^{-1}\ln\{1 - \alpha(x - \xi)/\beta\}, \alpha \ne 0$ $= (x - \xi)/\beta, \alpha = 0$	$\xi + \beta\{1 - [-\ln(F)]^{\alpha}\}/\alpha, \alpha \ne 0$ $\xi - \beta \ln[-\ln(F)], \alpha = 0$
Generalized logistic	$1/[1 + \exp(-y)]$ where $y = -\alpha^{-1}\ln\{1 - \alpha(x - \xi)/\beta\}, \alpha \ne 0$ $= (x - \xi)/\beta, \alpha = 0$	$\xi + \beta\{1 - [(1 - F)/F]^{\alpha}\}/\alpha, \alpha \ne 0$ $\xi - \beta \ln[(1 - F)/F], \alpha = 0$
Generalized Pareto	$1 - \exp(-y)$ where $y = -\alpha^{-1}\ln\{1 - \alpha(x - \xi)/\beta\}, \kappa \ne 0$ $= (x - \xi)/\beta, \kappa = 0$	$\xi + \beta[1 - (1 - F)^{\alpha}]/\alpha, \alpha \ne 0$ $\xi - \beta \ln(1 - F), \alpha = 0$

acceptance-rejection (AR) method is to replace the original $f_x(x)$ by an appropriate PDF $h_x(x)$ from which random variates can be produced easily and efficiently. The generated random variate from $h_x(x)$, then, is subject to testing before it is accepted as one from the original $f_x(x)$. This approach for generating random numbers has become widely used.

In AR methods, the PDF $f_x(x)$ from which a random variate x to be generated is represented, in terms of $h_x(x)$, by

$$f_x(x) = \varepsilon h_x(x) g(x) \tag{6.9}$$

in which $\varepsilon \geq 1$ and $0 < g(x) \leq 1$. Figure 6.2 illustrates the AR method in that the constant $\varepsilon \geq 1$ is chosen such that $\psi(x) = \varepsilon h_x(x)$ over the sample space of the random variable X. The problem then is to find a function $\psi(x) = \varepsilon h_x(x)$ such that $\psi(x) \geq f_x(x)$ and a function $h_x(x) = \psi(x)/\varepsilon$, from which random variates are generated. The constant ε that satisfies $\psi(x) \geq f_x(x)$ can be obtained from

$$\varepsilon = \max_x \left[\frac{f_x(x)}{h_x(x)} \right] \tag{6.10}$$

The algorithm of a generic AR method is the following:

1. Generate a uniform random number u from U(0, 1).
2. Generate a random variate y from $h_x(x)$.
3. If $u \leq g(y) = f_x(y)/\varepsilon h_x(y)$, accept y as the random variate from $f_x(x)$. Otherwise, reject both u and y, and go to step 1.

The efficiency of an AR method is determined by $P\{U \leq g(Y)\}$, which represents the probability that each individually generated Y from $h_x(x)$ will be accepted by the test. The higher the probability, the faster the task of generating a random number can be accomplished. It can be shown that $P\{U \leq g(Y)\} = 1/\varepsilon$
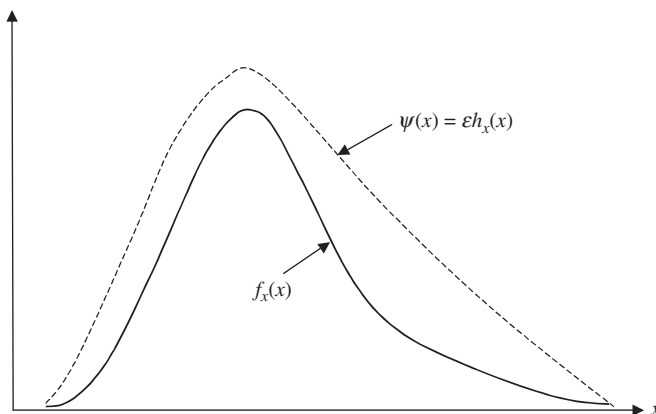
Figure 6.2 Illustration of the von Neumann acceptance-rejection (AR) procedure.

(see Problem 6.4). Intuitively, the maximum achievable efficiency for an AR method occurs when $\psi(x) = f_x(x)$. In this case, $\varepsilon = 1, g(x) = 1$, and the corresponding probability of acceptance $P\{U \leq g(Y)\} = 1$. Therefore, consideration must be given to two aspects when selecting $h_x(x)$ for AR methods: (1) the efficiency and exactness of generating a random number from $h_x(x)$ and (2) the closeness of $h_x(x)$ in imitating $f_x(x)$.

Example 6.3 Consider that Manning's roughness coefficient X of a cast iron pipe is uncertain with a density function $f_x(x), a \leq x \leq b$. Develop an AR algorithm using $\psi(x) = c$ and $h_x(x) = 1/(b - a)$, for $a \leq x \leq b$.

Solution Since $\psi(x) = c$ and $h_x(x) = 1/(b - a)$, the efficiency constant ε and $g(x)$ are

$$\varepsilon = \frac{\psi(x)}{h_x(x)} = c(b - a) \qquad g(x) = \frac{f_x(x)}{\psi(x)} = \frac{f_x(x)}{c} \qquad a \leq x \leq b$$

The AR algorithm for this example, then, can be outlined as the following:

1. Generate u_1 from U(0, 1).
2. Generate u_2 from U(0, 1) from which $y = a + (b - a)u_2$.
3. Determine if

$$u_1 \leq g(y) = \frac{f_x[a + (b - a)u_2]}{c}$$

   holds. If yes, accept y; otherwise, reject (u_1, y), and return to step 1.

In fact, this is the von Neumann (1951) algorithm for the AR method.

AR methods are important tools for random number generation because they can be very fast in comparison with the CDF-inverse method for distribution models the analytical forms of CDF inverse of which are not available. This approach has been applied to some distributions, such as gamma, resulting in extremely simple and efficient algorithms (Dagpunar, 1988).

6.3.3 Variable transformation method

The *variable transformation method* generates a random variate of interest based on its known statistical relationship with other random variables the variates of which can be produced easily. For example, one is interested in generating chi-square random variates with n degrees of freedom. The CDF-inverse method is not appropriate in this case because the chi-square CDF is not analytically expressible. However, knowing the fact that the sum of n squared independent standard normal random variables gives a chi-square random variable with n degrees of freedom (see Sec. 2.6.6), one could generate chi-square random variates from first producing n standard normal random variates, then squaring them, and finally adding them together. Therefore, the variable transformation method is sometimes effective for generating random variates from a complicated distribution based on variates produced from simple distributions. In fact, many algorithms described in the next section are based on the idea of variable transformation.

6.4 Generation of Univariate Random Numbers for Some Distributions

This section briefly outlines efficient algorithms for generating random variates for some probability distributions commonly used in hydrosystems engineering and analysis.

6.4.1 Normal distribution

A normal random variable with a mean μ_x and standard deviation σ_x, denoted as $X \sim \mathrm{N}(\mu_x, \sigma_x)$, has a PDF given in Eq. (2.58). The relationship between X and the standardized normal variable Z is

$$X = \mu_x + \sigma_x Z \tag{6.11}$$

in which Z is the standard normal random variable having a mean 0 and unit standard deviation, denoted as $Z \sim \mathrm{N}(0, 1)$. Based on Eq. (6.11), normal random variates with a specified mean and standard deviation can be generated from standard normal variates. Herein, three simple algorithms for generating standard normal variates are described.

Box-Muller algorithm. The algorithm (Box and Muller, 1958) produces a pair of independent $\mathrm{N}(0, 1)$ variates as

$$
\begin{aligned}
z_1 &= \sqrt{-2\ln(u_1)}\cos(2\pi u_2) \\
z_2 &= \sqrt{-2\ln(u_2)}\sin(2\pi u_2)
\end{aligned}
\tag{6.12}
$$

in which u_1 and u_2 are independent uniform variates from $\mathrm{U}(0, 1)$. The algorithm involves the following steps:

1. Generate two independent uniform random variates u_1 and u_2 from $\mathrm{U}(0, 1)$.
2. Compute z_1 and z_2 simultaneously using u_1 and u_2 according to Eq. (6.12).

Marsagalia-Bray algorithm. Marsagalia and Bray (1964) proposed an alternative algorithm that avoids using trigonometric evaluations. In their algorithm, two independent uniform random variates u_1 and u_2 are produced to evaluate the following three expressions:

$$
\begin{aligned}
V_1 &= 2U_1 - 1 \\
V_2 &= 2U_2 - 1 \\
R &= V_1^2 + V_2^2
\end{aligned}
\tag{6.13}
$$

If $R > 1$, the pair (u_1, u_2) is rejected from further consideration, and a new pair (u_1, u_2) is generated. For the accepted pair, the corresponding standard

normal variates are computed by

$$Z_1 = V_1 \sqrt{\frac{-2\ln(R)}{R}} \qquad Z_2 = V_2 \sqrt{\frac{-2\ln(R)}{R}} \tag{6.14}$$

The Marsagalia-Bray algorithm involves the following steps:

1. Generate two independent uniform random variates u_1 and u_2 from U(0, 1).
2. Compute V_1, V_2, and R according to Eq. (6.13).
3. Check if $R \leq 1$. If it is true, compute the two corresponding N(0, 1) variates using Eq. (6.14). Otherwise, reject (u_1, u_2) and return to step 1.

Algorithm based on the central limit theorem. This algorithm is based on the central limit theorem, which states that the sum of independent random variables approaches a normal distribution as the number of random variables increases. Specifically, consider the sum of J independent standard uniform random variates from U(0, 1). The following relationships are true:

$$E\left(\sum_{j=1}^{J} U_j\right) = \frac{J}{2} \tag{6.15}$$

$$\text{Var}\left(\sum_{j=1}^{J} U_j\right) = \frac{J}{12} \tag{6.16}$$

By the central limit theorem, this sum of J independent U's would approach a normal distribution with the mean and variance given in Eqs. (6.15) and (6.16), respectively. Constrained by the unit variance of the standard normal variates, Eq. (6.16) yields $J = 12$. Then a standard normal variate is generated by

$$Z = \left(\sum_{j=1}^{12} U_j\right) - 6 \tag{6.17}$$

The central limit theorem–based algorithm can be implemented as

1. Generate 12 uniform random variates from U(0, 1).
2. Compute the corresponding standard normal variate by Eq. (6.17).

There are many other efficient algorithms developed for generating normal random variates using the variable transformation method and AR method. For these algorithms readers are referred to Rubinstein (1981).

6.4.2 Lognormal distribution

Consider a random variable X having a lognormal distribution with a mean μ_x and standard deviation σ_x, that is, $X \sim \mathrm{LN}(\mu_x, \sigma_x)$. For a lognormal random variable X, its logarithmic transform $Y = \ln(X)$ leads to a normal distribution for Y. The PDF of X is given in Eq. (2.65). In the log-transformed space, the mean and standard deviation of $\ln(X)$ can be computed, in terms of μ_x and σ_x, by Eqs. (2.67a) and (2.67b). Since $Y = \ln(X)$ is normally distributed, the generation of lognormal random variates from $X \sim \mathrm{LN}(\mu_x, \sigma_x)$ can be obtained by the following steps:

1. Calculate the mean $\mu_{\ln x}$ and standard deviation $\sigma_{\ln x}$ of log-transformed variable $\ln(X)$ by Eqs. (2.67a) and (2.67b), respectively.
2. Generate the standard normal variate z from $\mathrm{N}(0, 1)$.
3. Compute $y = \mu_{\ln x} + \sigma_{\ln x} z$.
4. Compute the lognormal random variate $x = e^{y}$.

6.4.3 Exponential distribution

The exponential distribution is used frequently in reliability computation in the framework of time-to-failure analysis. It is often used to describe the stochastic behavior of time to failure and time-to-repair of a system or component. A random variable X having an exponential distribution with parameter β, denoted by $X \sim \mathrm{EXP}(\beta)$, is described by Eq. (2.79). By the CDF-inverse method,

$$u = F_x(x) = 1 - e^{-x/\beta} \tag{6.18}$$

so that

$$X = -\beta \ln(1 - U) \tag{6.19}$$

Since $1 - U$ is distributed in the same way as U, Eq. (6.19) is reduced to

$$X = -\beta \ln(U) \tag{6.20}$$

Equation (6.20) is also valid for random variables with the standard exponential distribution, that is, $V \sim \exp(\beta = 1)$. The algorithm for generating exponential variates is

1. Generate uniform random variate u from $\mathrm{U}(0, 1)$.
2. Compute the standard exponential random variate $v = -\ln(u)$.
3. Calculate $x = v\beta$.

6.4.4 Gamma distribution

The gamma distribution is used frequently in the statistical analysis of hydrologic data. For example, Pearson type III and log-Pearson type III distributions used in the flood frequency analysis are members of the gamma distribution family. It is a very versatile distribution the PDF of which can take many forms (see Fig. 2.20). The PDF of a two-parameter gamma random variable, denoted by $X \sim \text{GAM}(\alpha, \beta)$, is given by Eq. (2.72). The standard gamma PDF involving one-parameter α can be derived using variable transformation by letting $Y = X/\beta$. The PDF of the standard gamma random variable Y, denoted by $Y \sim \text{GAM}(\alpha)$, is shown in Eq. (2.78). The standard gamma distribution is used in all algorithms to generate gamma random variate Ys from which random variates from a two-parameter gamma distribution are obtained from $X = \beta Y$.

The simplest case in generating gamma random variates is when the shape parameter α is a positive integer (Erlang distribution). In such a case, the random variable $Y \sim \text{GAM}(\alpha)$ is a sum of α independent and identical standard exponential random variables with parameter $\beta = 1$. The random variates from $Y \sim \text{GAM}(\alpha)$, then, can be obtained as

$$Y = \sum_{i=1}^{\alpha} -\ln(U_i) \tag{6.21}$$

To avoid large numbers of logarithmic evaluations (when α is large), Eq. (6.21) alternatively can be expressed as

$$Y = -\ln\left(\prod_{i=1}^{\alpha} U_i\right) \tag{6.22}$$

Although simplicity is the idea, this algorithm for generating gamma random variates has three disadvantages: (1) It is only applicable to integer-valued shape parameter α, (2) the algorithm becomes extremely slow when α is large, and (3) for a large α, numerical underflow on a computer could occur.

Several algorithms have been developed for generating standard gamma random variates for a real-valued α. The algorithms can be classified into those which are applicable for the full range ($\alpha \geq 0$), $0 \leq \alpha \leq 1$, and $\alpha \geq 1$. Dagpunar (1988) showed that through a numerical experiment, algorithms developed for a full range of α are not efficient in comparison with those especially tailored for subregions. The two efficient AR-based algorithms are presented in Dagpunar (1988).

6.4.5 Poisson distribution

The Poisson random variable is discrete, having a PMF $f_x(x_i) = P(X = x_i)$ given in Eq. (2.53). Dagpunar (1988) presented a simple algorithm and used the CDF-inverse method based on Eq. (6.7). When generating Poisson random variates, care should be taken so that e^{-v} is not smaller than the machine's smallest positive real value. This could occur especially when the Poisson

parameter v is large. An algorithm for generating Poisson random variates is as follows:

1. Generate $u \sim U(0, 1)$ and initialize $x = 0$ and $y = e^{-v}$

2. If $y < u$, go to step 3. Otherwise, x is the Poisson random variate sought.

3. Let $u = u - y$, $x = x + 1$, and update $y = vy/x$. Then go to step 2.

This algorithm is efficient when $v < 20$. For a large v, the Poisson distribution can be approximated by a normal distribution with a mean $v - 0.5$ and standard deviation of $\sqrt{v}$. Then a Poisson random variate is set to the round-off normal random variate from $N(v - 0.5, \sqrt{v})$.

Other algorithms have been developed for generating Poisson random variates. Rubinstein (1981) used the fact that the interarrival time between events for a Poisson process has an exponential distribution with parameter $1/v$. Atkinson (1979) applied the AR method using a logistic distribution as the enveloping PDF.

6.4.6 Other univariate distributions and computer programs

The algorithms described in the preceding subsections are for some probability distributions commonly used in hydrosystems engineering and analysis. One might encounter other types of probability distributions in an analysis that are not described herein. There are several books that have been written for generating univariate random numbers (Rubinstein, 1981; Dagpunar, 1988; Gould and Tobochnik, 1988; Law and Kelton, 1991). To facilitate the implementation of Monte Carlo simulation, computer subroutines in different languages are available (Press et al., 1989, 1992, 2002; IMSL, 1980). In addition, many other spreadsheet-based computer software, such as Microsoft Excel, @Risk, and Crystal Ball, contain statistical functions allowing the generation of random variates of various distributions.

6.5 Generation of Vectors of Multivariate Random Variables

In preceding sections, discussions focused on generating univariate random variates. It is not uncommon for hydrosystems engineering problems to involve multiple random variables that are correlated and statistically dependent. For example, many data show that the peak discharge and volume of a runoff hydrograph are positively correlated. To simulate systems involving correlated random variables, generated random variates must preserve the probabilistic characteristics of the variables and the correlation structure among them. Although multivariate random number generation is an extension of the univariate case, mathematical difficulty and complexity associated with multivariate problems increase rapidly as the dimension of the problem gets larger.

Compared with generating univariate random variates, multivariate random variate generation is much more restricted to fewer joint distributions, such as multivariate normal, multivariate lognormal, and multivariate gamma (Ronning, 1977; Johnson, 1987; Parrish, 1990). Nevertheless, the algorithms for generating univariate random variates serve as the foundation for many multivariate Monte Carlo algorithms.

6.5.1 CDF-inverse method

This method is an extension of the univariate case described in Sec. 6.3.1. Consider a vector of K random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$ having a joint PDF of

$$f_x(\boldsymbol{x}) = f_{1,2,\ldots,K}(x_1, x_2, \ldots, x_K) \tag{6.23}$$

This joint PDF can be decomposed to

$$f_x(\boldsymbol{x}) = f_1(x_1) \times f_2(x_2|x_1) \times \cdots \times f_K(x_K|x_1, x_2, \ldots, x_{K-1}) \tag{6.24}$$

in which $f_1(x_1)$ and $f_k(x_k|x_1, x_2, \ldots, x_{k-1})$ are, respectively, the marginal PDF and the conditional PDF of random variables X_1 and X_k. In the case when all K random variables are statistically independent, Eq. (6.23) is simplified to

$$f_x(\boldsymbol{x}) = \prod_{k=1}^{K} f_k(x_k) \tag{6.25}$$

One observes that from Eq. (6.25) the joint PDF of several independent random variables is simply the product of the marginal PDF of the individual random variable. Therefore, generation of a vector of independent random variables can be accomplished by treating each individual random variable separately, as in the case of the univariate problem. However, treatment of random variables cannot be made separately in the case when they are correlated. Under such circumstances, as can be seen from Eq. (6.24), the joint PDF is the product of conditional distributions. Referring to Eq. (6.24), generation of K random variates following the prescribed joint PDF can proceed as follows:

1. Generate random variates for X_1 from its marginal PDF $f_1(x_1)$.
2. Given $X_1 = x_1$ obtained from step 1, generate X_2 from the conditional PDF $f_2(x_2|x_1)$.
3. With $X_1 = x_1$ and $X_2 = x_2$ obtained from steps 1 and 2, produce X_3 based on $f_3(x_3|x_1, x_2)$.
4. Repeat the procedure until all K random variables are generated.

To generate multivariate random variates by the CDF-inverse method, it is required that the analytical relationship between the value of the variate

and the conditional distribution function is available. Following Eq. (6.24), the product relationship also holds in terms of CDFs as

$$F_x(\boldsymbol{x}) = F_1(x_1) \times F_2(x_2|x_1) \times \cdots \times F_K(x_K|x_1, x_2, \ldots, x_{K-1}) \qquad (6.26)$$

in which $F_1(x_1)$ and $F_k(x_k|x_1, x_2, \ldots, x_{k-1})$ are the marginal CDF and conditional CDF of random variables X_1 and X_k, respectively. Based on Eq. (6.26), the algorithm using the CDF-inverse method to generate n sets of K multivariate random variates from a specified joint distribution is described below (Rosenblatt, 1952):

1. Generate K standard uniform random variates $u_1, u_2, \ldots, u_K$ from U(0, 1).
2. Compute

$$
\begin{aligned}
x_1 &= F_1^{-1}(u_1) \\
x_2 &= F_2^{-1}(u_2\,|x_1) \\
&\;\vdots \\
x_K &= F_K^{-1}(u_k\,|x_1, x_2, \ldots, x_{K-1})
\end{aligned}
\qquad (6.27)
$$

3. Repeat steps 1 and 2 for n sets of random vectors.

There are $K!$ ways to implement this algorithm in which different orders of random variates $X_k, k = 1, 2, \ldots, K$, are taken to form the random vector $\boldsymbol{X}$. In general, the order adopted could affect the efficiency of the algorithm.

Example 6.4 This example is extracted from Nguyen and Chowdhury (1985). Consider a box cut of an open strip coal mine, as shown in Fig. 6.3. The overburden has a phreatic aquifer overlying the coal seam. In the next bench of operation, excavation is to be made 50 m ($d = 50$ m) behind the box-cut high wall. It is suggested that for
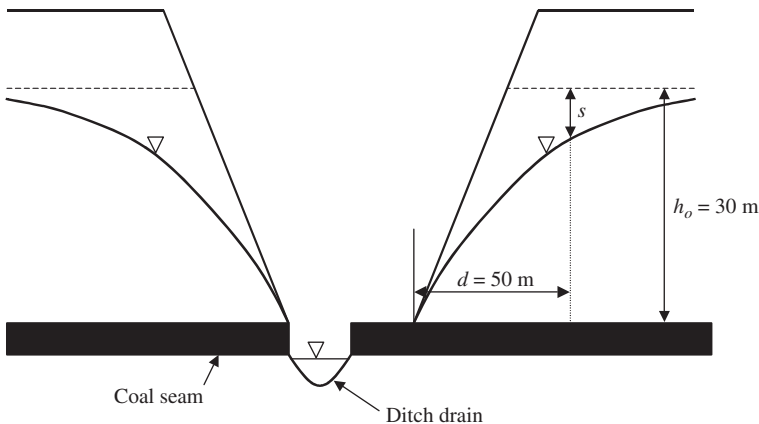
Figure 6.3 Box cut of an open strip coal mine resulting in water drawdown. (*After Nguyen and Chowdhury, 1985.*)

safety reasons of preventing slope instability the excavation should start at the time when the drawdown in the overburden $d = 50$ m away from the excavation point has reached at least 50 percent of the total aquifer depth (h_o).

Nguyen and Raudkivi (1983) gave the transient drawdown equation for this problem as

$$\frac{s}{h_o} = 1 - \text{erf}\left(\frac{d}{\sqrt{2K_h h_o t/S}}\right) \tag{6.28}$$

in which s is the drawdown (in meters) at a distance d (in meters) from the toe of the embankment, h_o is the original thickness of the water bearing aquifer, t is the drawdown recess time (in days), K_h is the aquifer permeability, S is the aquifer storage coefficient; and $\text{erf}(x)$ is the error function, referring to Eq. (2.69), as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-v^2} \, dv$$

with v being a dummy variable of integration.

From a field investigation through a pump test, data indicate that the aquifer permeability has approximately a normal distribution with a mean of 0.1 m/day and coefficient of variation of 10 percent. The storage coefficient of the aquifer has a mean of 0.05 with a standard deviation of 0.005. Further, the correlation coefficient between the permeability and storage coefficient is about 0.5.

Since the aquifer properties are random variables, the time required for the drawdown to reach the safe level for excavation also is a random variable. Apply the CDF-inverse method (using $n = 400$ repetitions) to estimate the statistical properties of the time of recess, including its mean, standard deviation, and skewness coefficient.

Solution The required drawdown recess time for a safe excavation can be obtained by solving Eq. (6.28), with $s/h_o = 0.5$ and $\text{erf}^{-1}(0.5) = 0.477$ (Abramowitz and Stegun, 1972; or by Eq. (2.69), as

$$t = \left(\frac{d}{2 \times 0.477}\right)^2 \frac{S}{K_h h_o} \tag{6.29}$$

The problem is a bivariate normal distribution (see Sec. 2.7.2) with two correlated random variables. The permeability K_h and storage coefficient S, referring to Eq. (2.108), have the joint PDF

$$f_{K_h,S}(k_h, S) = \frac{1}{2\pi \sigma_{kh}\sigma_s \sqrt{1 - \rho_{kh,s}^2}} e^{-Q}$$

with $\quad Q = \frac{1}{2(1 - \rho_{kh,s}^2)} \left[\frac{(k_h - \mu_{kh})^2}{\sigma_{kh}^2} - 2\rho_{kh,s}\frac{(k_h - \mu_{kh})(s - \mu_s)}{\sigma_{kh}\sigma_s} + \frac{(s - \mu_s)^2}{\sigma_s^2}\right)$

where $\rho_{kh,s}$ is the correlation coefficient between K_h and S, which is 0.5; σ_{kh} is the standard deviation of permeability, $0.1 \times 0.1 = 0.01$ m/day; σ_s is the standard deviation of the storage coefficient, 0.005; μ_{kh} is the mean of permeability, 0.1 m/day; and μ_s is the mean storage coefficient, 0.05.

To generate bivariate random variates according to Eq. (6.27), the marginal PDF of permeability K_h and the conditional PDF of storage coefficient S, or vice versa, are

required. They can be derived, respectively, according to Eq. (2.109), as

$$f_{K_h}(k_h) = \frac{1}{\sqrt{2\pi}\,\sigma_{kh}} \exp\left[-\frac{1}{2}\left\{\frac{k_h - \mu_{kh}}{\sigma_{kh}}\right\}^2\right] \tag{6.30}$$

$$f_{s|k_h}(s|k_h) = \frac{1}{\sqrt{2\pi}\,\sigma_s\sqrt{1 - \rho_{kh,s}^2}} \exp\left\{-\frac{1}{2}\left[\frac{(s - \mu_s) - \rho_{kh,s}(\sigma_s/\sigma_{kh})(k_h - \mu_{kh})}{\sigma_s\sqrt{1 - \rho_{kh,s}^2}}\right]^2\right\} \tag{6.31}$$

From the conditional PDF given earlier, the conditional expectation and conditional standard deviation of storage coefficient S, given a specified value of permeability $K_h = k_h$, can be derived, respectively, according to Eqs. (2.110) and (2.111), as

$$\mu_{S|kh} = E(S|K_h = k_h) = \mu_s + \rho_{kh,s}\frac{\sigma_s}{\sigma_{kh}}(k_h - \mu_{kh}) \tag{6.32}$$

$$\sigma_{s|kh} = \sigma_s\sqrt{1 - \rho_{kh,s}^2} \tag{6.33}$$

Therefore, the algorithm for generating bivariate normal random variates to estimate the statistical properties of the drawdown recess time can be outlined as follows:

1. Generate a pair of independent standard normal variates z_1' and z_2'.
2. Compute the corresponding value of permeability $k_h = \mu_{kh} + \sigma_{kh}z_1'$.
3. Based on the value of permeability obtained in step 2, compute the conditional mean and conditional standard deviation of the storage coefficient according to Eqs. (6.32) and (6.33), respectively. Then calculate the corresponding storage coefficient as $s = \mu_{s|kh} + \sigma_{s|kh}z_2'$.
4. Use $K_h = k_h$ and $S = s$ generated in steps 3 and 4 in Eq. (6.29) to compute the corresponding drawdown recess time t.
5. Repeat steps 1 through 4 $n = 400$ times to obtain 400 realizations of drawdown recess times $\{t_1, t_2, \ldots, t_{400}\}$.
6. Compute the sample mean, standard deviation, and skewness coefficient of the drawdown recess time according to the last column of Table 2.1.

The histogram of the drawdown recess time resulting from 400 simulations is shown in Fig. 6.4. The statistical properties of the drawdown recess time are estimated as

$$\text{Mean } \mu_t = 45.73 \text{ days}$$

$$\text{Standard deviation } \sigma_t = 4.72 \text{ days}$$

$$\text{Skewness coefficient } \gamma_t = 0.487$$

6.5.2 Generating multivariate normal random variates

A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$ has a multivariate normal distribution with a mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{C}_x$, denoted as $\boldsymbol{X} \sim \text{N}(\boldsymbol{\mu}_x, \boldsymbol{C}_x)$. The joint PDF of K normal random variables is given in Eq. (2.112). To generate high-dimensional multivariate normal random variates with specified $\boldsymbol{\mu}_x$
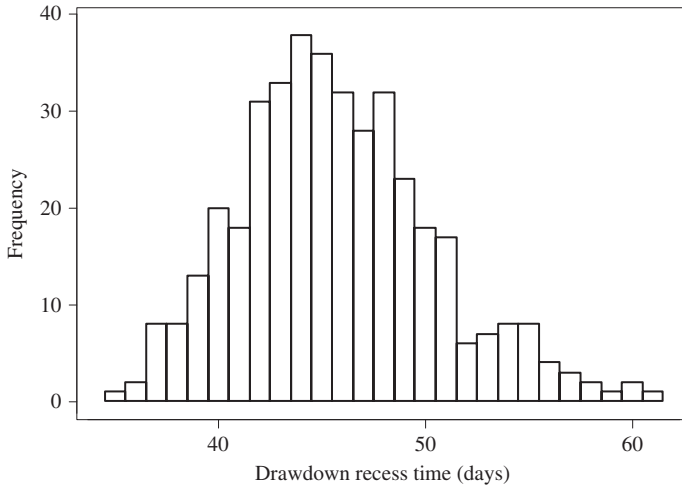
Figure 6.4 Histogram of simulated drawdown recess time for Example 6.4.

and $\boldsymbol{C}_x$, the CDF-inverse algorithm described in Sec. 6.5.1 might not be efficient. In this section, two alternative algorithms for generating multivariate normal random variates are described. Both algorithms are based on orthogonal transformation using the covariance matrix $\boldsymbol{C}_x$ or correlation matrix $\boldsymbol{R}_x$ described in Sec. 2.7.1. The result of the transformation is a vector of independent normal variables, which can be generated easily by the algorithms described in Sec. 6.4.1.

Square-root method. The *square-root algorithm* decomposes the covariance matrix $\boldsymbol{C}_x$ or correlation matrix $\boldsymbol{R}_x$ into

$$\boldsymbol{R}_x = \boldsymbol{L}\boldsymbol{L}^t \qquad \boldsymbol{C}_x = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^t$$

as shown in Appendix 4B, in which $\boldsymbol{L}$ and $\tilde{\boldsymbol{L}}$ are $K \times K$ lower triangular matrices associated with the correlation and covariance matrices, respectively. According to Eq. (4B.12), $\tilde{\boldsymbol{L}} = \boldsymbol{D}_x^{1/2}\boldsymbol{L}$, with $\boldsymbol{D}_x$ being the $K \times K$ diagonal matrix of variances of the K involved random variables.

In addition to being symmetric, if $\boldsymbol{R}_x$ or $\boldsymbol{C}_x$ is a positive-definite matrix, the Cholesky decomposition (see Appendix 4B) is an efficient method for finding the unique lower triangular matrices $\boldsymbol{L}$ or $\tilde{\boldsymbol{L}}$ (Young and Gregory, 1973; Golub and Van Loan, 1989). Using the matrix $\boldsymbol{L}$ or $\tilde{\boldsymbol{L}}$, the vector of multivariate normal random variables can be expressed as

$$\boldsymbol{X} = \mu_x + \tilde{\boldsymbol{L}}\boldsymbol{Z}' = \mu_x + \boldsymbol{D}_x^{1/2}\boldsymbol{L}\boldsymbol{Z}' \qquad (6.34)$$

in which $\boldsymbol{Z}'$ is an $K \times 1$ column vector of independent standard normal variables. It was shown easily in Appendix 4B that the expectation vector and the covariance matrix of the right-hand side in Eq. (6.34), $E(\mu_x + \tilde{\boldsymbol{L}}\boldsymbol{Z}')$, are equal

to $\boldsymbol{\mu}_x$ and $\boldsymbol{C}_x$, respectively. Based on Eq. (6.34), the square-root algorithm for generating multivariate normal random variates can be outlined as follows:

1. Compute the lower triangular matrix associated with the correlation or covariance matrix by the Cholesky decomposition method.
2. Generate K independent standard normal random variates $\boldsymbol{z}' = (z_1', z_2', \ldots, z_K')^t$ from N(0, 1).
3. Compute the corresponding normal random variates by Eq. (6.34).
4. Repeat steps 1 through 3 to generate the desired number of sets of normal random vectors.

Example 6.5 Refer to Example 6.4. Apply the square-root algorithm to estimate the statistical properties of the drawdown recess time, including its mean, standard deviation, and skewness coefficient. Compare the results with Example 6.4.

Solution By the square-root algorithm, the covariance matrix of permeability K_h and storage coefficient S,

$$\boldsymbol{C}(K_h, S) = \begin{bmatrix} 0.01^2 & 0.5(0.01)(0.005) \\ 0.5(0.01)(0.005) & 0.005^2 \end{bmatrix} = \begin{bmatrix} 0.0001 & 0.000025 \\ 0.000025 & 0.000025 \end{bmatrix}$$

is decomposed into the multiplication of the two lower triangular matrices, by the Cholesky decomposition, as

$$\tilde{\boldsymbol{L}} = \begin{bmatrix} 0.01 & 0 \\ 0.0025 & 0.00443 \end{bmatrix}$$

The Monte Carlo simulation can be carried out by the following steps:

1. Generate a pair of standard normal variates z_1' and z_2'.
2. Compute the permeability K_h and storage coefficient S simultaneously as

$$\begin{bmatrix} k_h \\ s \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 0.01 & 0 \\ 0.0025 & 0.00433 \end{bmatrix} \begin{bmatrix} z_1' \\ z_2' \end{bmatrix}$$

3. Use (k_h, s) generated from step 2 in Eq. (6.29) to compute the corresponding drawdown recess time t.
4. Repeat steps 1 through 3 $n = 400$ times to obtain 400 realizations of drawdown recess times $\{t_1, t_2, \ldots, t_{400}\}$.
5. Compute the mean, standard deviation, and skewness coefficient of the drawdown recess time.

The results from carrying out the numerical simulation are

$$\text{Mean } \mu_t = 45.94 \text{ days}$$

$$\text{Standard deviation } \sigma_t = 4.69 \text{ days}$$

$$\text{Skewness coefficient } \gamma_t = 0.301$$

The histogram of 400 simulated drawdown recess times is shown in Fig. 6.5. The mean and standard deviation are very close to those obtained in Example 6.4, whereas the
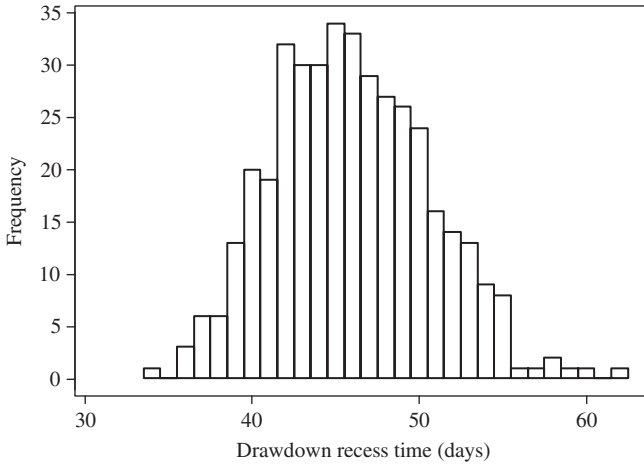
Figure 6.5 Histogram of simulated drawdown recess time for Example 6.5.

skewness coefficient is 62 percent of that found in Example 6.4. This indicates that 400 simulations are sufficient to estimate the mean and standard deviation accurately, but more simulations are needed to estimate the skewness coefficient accurately.

Spectral decomposition method. The basic idea of spectral decomposition is described in Appendix 4B. The method finds the eigenvalues and eigenvectors of the correlation or covariance matrix of the multivariate normal random variables. Through the spectral decomposition, the original vector of multivariate normal random variables X, then, is related to a vector of independent standard normal random variables $Z' \sim N(\mathbf{0}, \mathbf{I})$ as

$$X = \mu_x + D_x^{1/2} V \Lambda^{1/2} Z' = \mu_x + \tilde{V} \tilde{\Lambda}^{1/2} Z' \qquad (6.36)$$

in which $\tilde{V}$ and $\tilde{\Lambda}$ are the eigenvector and diagonal eigenvalue matrices of C_x, respectively, whereas V and Λ are the eigenvector and diagonal eigenvalue matrices of R_x, respectively. Equation (6.36) clearly reveals the necessary computations for generating multivariate normal random vectors. The spectral decomposition algorithm for generating multivariate normal random variates involves the following steps:

1. Obtain the eigenvector matrix and diagonal eigenvalue matrix of the correlation matrix R_x or covariance matrix C_x.

2. Generate K independent standard normal random variates $z' = (z'_1, z'_2, \ldots, z'_K)^t$.

3. Compute the correlated normal random variates X by Eq. (6.36).

Many efficient algorithms have been developed to determine the eigenvalues and eigenvectors of a symmetric matrix. For the details of such techniques, readers are referred to Golub and Van Loan (1989) and Press et al. (1992).

6.5.3 Generating multivariate random variates with known marginal pdfs and correlations

In many practical hydrosystems engineering problems, random variables often are statistically and physically dependent. Furthermore, distribution types for the random variables involved can be a mixture of different distributions, of which the corresponding joint PDF or CDF is difficult to establish. As a practical alternative, to replicate such systems properly, the Monte Carlo simulation should be able to preserve the correlation relationships among the stochastic variables and their marginal distributions.

In a multivariate setting, the joint PDF represents the complete information describing the probabilistic structures of the random variables involved. When the joint PDF or CDF is known, the marginal distribution and conditional distributions can be derived, from which the generation of multivariate random variates can be made straightforwardly in the framework of Rosenblatt (1952). However, in most practical engineering problems involving multivariate random variables, the derivation of the joint CDF generally is difficult, and the availability of such information is rare. The level of difficulty, in both theory and practice, increases with the number of random variables and perhaps even more so by the type of corresponding distributions. Therefore, more often than not, one has to be content with preserving incomplete information represented by the marginal distribution of each individual random variable and the correlation structure. In doing so, the difficulty of requiring a complete joint PDF in the multivariate Monte Carlo simulation is circumvented.

To generate correlated random variables with a mixture of marginal distributions, a methodology adopting a bivariate distribution model was first suggested by Li and Hammond (1975). The practicality of the approach was advanced by Der Kiureghian and Liu (1985), who, based on the Nataf bivariate distribution model (Nataf, 1962), developed a set of semiempirical formulas so that the necessary calculations to preserve the original correlation structure in the normal transformed space are reduced (see Table 4.5). Chang et al. (1994) used this set of formulas, which transforms the correlation coefficient of a pair of nonnormal random variables to its equivalent correlation coefficient in the bivariate standard normal space, for multivariate simulation. Other practical alternatives, such as the polynomial normal transformation (Vale and Maurelli, 1983; Chen and Tung, 2003), can serve the same purpose. Through a proper normal transformation, the multivariate Monte Carlo simulation can be performed in a correlated standard normal space in which efficient algorithms, such as those described in Sec. 6.5.2, can be applied.

The Monte Carlo simulation that preserves marginal PDFs and correlation structure of the involved random variables consists of following two basic steps:

Step 1. Transformation to a standard normal space. Through proper normal transformation, the operational domain is transformed to a standard normal space in which the transformed random variables are treated as if they were multivariate standard normal with the correlation matrix $\boldsymbol{R}_z$. As a result, multivariate normal random variates can be generated by the techniques described in Sec. 6.5.2.

Step 2. Inverse transformation. Once the standardized multivariate normal random variates are generated, then one can do the inverse transformation

$$X_k = F_k^{-1}[\Phi(Z_k)] \qquad \text{for } k = 1, 2, \ldots, K \qquad (6.37)$$

to compute the values of multivariate random variates in the original space.

6.5.4 Generating multivariate random variates subject to linear constraints

Procedures described in Sec. 6.5.2 are for generating multivariate normal (Gaussian) random variables without imposing constraints or restriction on the values of variates. The procedures under this category are also called *unconditional* (or *nonconditional*) *simulation* (Borgman and Faucette, 1993; Chilès and Delfiner, 1999). In hydrosystems modeling, random variables often exist for which, in addition to their statistical correlation, they are physically related in certain functional forms. In particular, this section describes the procedures for generating multivariate Gaussian random variates that must satisfy prescribed linear relationships. An example is the use of unit hydrograph model for estimating design runoff based on a design rainfall excess hyetograph. The *unit hydrograph* is applied as follows:

$$\boldsymbol{P}\boldsymbol{u} = \boldsymbol{q} \qquad (6.38)$$

where $\boldsymbol{P}$ is an $n \times J$ Toeplitz matrix defining the design effective rainfall hyetograph, $\boldsymbol{u}$ is a $J \times 1$ column vector of unit hydrograph ordinates, and $\boldsymbol{q}$ is the $n \times 1$ column vector of direct runoff hydrograph ordinates. In the process of deriving a unit hydrograph for a watershed, there exist various uncertainties rendering $\boldsymbol{u}$ uncertain. Hence the design runoff hydrograph $\boldsymbol{q}$ obtained from Eq. (6.38) is subject to uncertainty. Therefore, to generate a plausible direct runoff hydrograph for a design rainfall excess hyetograph, one could generate unit hydrographs that must consider the following physical constraint:

$$\sum_{j=1}^{J} U_j = c \qquad (6.39)$$

in which c is a constant to ensure that the volume of unit the hydrograph is one unit of effective rainfall.

The linearly constrained Monte Carlo simulation can be conducted by using the acceptance-rejection method first proposed by von Neumann (1951). The AR method generally requires a large number of simulations to satisfy

the constraint and, therefore, is not computationally efficient. Borgman and Faucettee (1993) developed a practical method to convert a Gaussian linearly constrained simulation into a Gaussian conditional simulation that can be implemented straightforwardly. The following discussions will concentrate on the method of Borgman and Faucette (1993).

Conditional simulation (CS) was developed in the field of geostatistics for modeling spatial uncertainty to generate a plausible random field that honors the actual observational values at the sample points (Chilès and Delfiner, 1999). In other words, conditional simulation yields special subsets of realizations from an unconditional simulation in that the generated random variates match with the observations at the sample points. For the multivariate normal case, the Gaussian conditional simulation is to simulate a normal random vector X_2 conditional on the normal random vector $X_1 = x_{1*}$. To implement the conditional simulation, define a new random vector X encompassing of X_1 and X_2 as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}(\mu_x, C_x) = \mathrm{N}\left(\begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix}, \begin{bmatrix} C_{x,11} & C_{x,12} \\ C_{x,21} & C_{x,22} \end{bmatrix} \right) \tag{6.40}$$

in which $\mu_x = (\mu_{x_1}, \mu_{x_2})^t$, and C_x is the covariance matrix of X, which is broken down into $C_{x,ij}$ representing the covariance matrix between random vectors X_i and X_j for $i, j = 1, 2$.

Based on the random vector $x = (x_1, x_2)^t$ generated from the unconditional simulation, the values of random variates for x_{2*}, conditioned on $X_1 = x_{1*}$, can be obtained, analogous to Eq. (2.110), by

$$x_{2*} = x_2 + C_{x,12}^t C_{x,11}^{-1}(x_{1*} - x_1) \tag{6.41}$$

Consider a problem involving K correlated random variables the values $X = x$ of which are subject to the following linear constraints:

$$A_{m \times K} x_{K \times 1} = b_{m \times 1} \tag{6.42}$$

in which A is an $m \times K$ matrix of constants, and b is an $m \times 1$ column vector of constants. To generate K multivariate normal random variates satisfying Eq. (6.42), one can define a new $(m + K)$-element random vector Y, analogous to Eq. (6.40), as

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} AX \\ X \end{bmatrix} = TX \sim \mathrm{N}\left(\begin{bmatrix} \mu_{y_1} \\ \mu_{y_2} \end{bmatrix}, \begin{bmatrix} C_{y,11} & C_{y,12} \\ C_{y,21} & C_{y,22} \end{bmatrix} \right) \tag{6.43}$$

where T is an $(m + K) \times K$ matrix. The mean and covariance matrix of random vector Y can be obtained, respectively, as

$$\mu_y = \begin{bmatrix} \mu_{y1} \\ \mu_{y2} \end{bmatrix} = \begin{bmatrix} A\mu_x \\ \mu_x \end{bmatrix} \qquad C_y = TC_xT^t \tag{6.44}$$

To generate multivariate normal random vector X subject to linear constraints (Eq. 6.42) is equivalent to a conditional simulation in that random vector y_{2*} is

generated conditioned on $\boldsymbol{y}_1 = \boldsymbol{b}$. Hence, using the spectral decomposition described in Sec. 6.5.2.2, random vector $\boldsymbol{X}$ subject to linear constraints Eq. (6.42) can be obtained in the following two steps:

1. Calculate $(m + K)$-dimensional multivariate normal random vector $\boldsymbol{y}$ by unconditional simulation as

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \tilde{\boldsymbol{V}}_y \tilde{\Lambda}_y^{0.5} \boldsymbol{Z}' + \boldsymbol{\mu}_y \tag{6.45}$$

where $\boldsymbol{y}_1$ is an $m \times 1$ column vector, $\boldsymbol{y}_2$ is a $K \times 1$ column vector; $\tilde{\boldsymbol{V}}_y$ is an $(m + K) \times (m + K)$ eigenvector matrix of $\boldsymbol{C}_y$, and $\tilde{\Lambda}_y$ is a diagonal matrix of eigenvalues of $\boldsymbol{C}_y$, and $\boldsymbol{Z}'$ is an $(m + K)$ column vector of independent standard normal variates.

2. Calculate the linearly constrained K-dimensional vector of random variates $\boldsymbol{x}$, according to Eq. (6.41), as

$$\boldsymbol{x} = \boldsymbol{y}_{2*} = \boldsymbol{y}_2 + \boldsymbol{C}_{y,12}^t \boldsymbol{C}_{y,11}^{-1} (\boldsymbol{b} - \boldsymbol{y}_1) \tag{6.46}$$

This constrained multivariate normal simulation has been applied, by considering the uncertainties in the unit hydrograph and geomorphologic instantaneous unit hydrograph, to reliability analysis of hydrosystems engineering infrastructures (Zhao et al., 1997a, 1997b; Wang and Tung, 2005).

6.6   Monte Carlo Integration

In reliability analysis, computations of system and/or component reliability and other related quantities, such as mean time to failure, essentially involve integration operations. A simple example is the time-to-failure analysis in which the reliability of a system within a time interval $(0, t)$ is obtained from

$$p_s = \int_t^\infty f_t(t) \, dt \tag{6.47}$$

where $f_t(t)$ is the failure density function. A more complex example of the reliability computation is by load-resistance interference in that the reliability is

$$p_s = P[R(\boldsymbol{X}_R) \geq L(\boldsymbol{X}_L)] = P[W(\boldsymbol{X}_R, \boldsymbol{X}_L) \geq 0] = P[W(\boldsymbol{X}) \geq 0]$$
$$= \int_{W(\boldsymbol{x}) \geq 0} f_x(\boldsymbol{x}) \, d\boldsymbol{x} \tag{6.48}$$

where $R(\boldsymbol{X}_R)$ and $L(\boldsymbol{X}_L)$ are, respectively, resistance and load functions, which are dependent on some basic stochastic variables $\boldsymbol{X}_R = (X_1, X_2, \ldots, X_m)$ and $\boldsymbol{X}_L = (X_{m+1}, X_{m+2}, \ldots, X_K)$, and $W(\boldsymbol{X})$ is the performance function. As can be seen, computation of reliability by Eq. (6.48) involves K-dimensional integrations.

For cases of integration in one or two dimensions, such as Eq. (6.47), where the integrands are well behaved (e.g., no discontinuity), conventional numerical integration methods, such as the trapezoidal approximation or Simpson's rule (see Appendix 4A), are efficient and accurate. For example, using Simpson's rule, the error in a one-dimensional integration is $O(n^{-4})$, with n being the number of discretizations, and the error in a two-dimensional integration is $O(n^{-2})$. Gould and Tobochnik (1988) show that, in general, if the error for the one-dimensional integration is $O(n^{-a})$, the error with a K-dimensional integration would be $O(n^{-a/K})$. As can be seen, the accuracy of conventional numerical integration schemes decreases rapidly as the dimension of integration increases. For multiple integrals, such as Eq. (6.48), the Monte Carlo method becomes a more suitable numerical technique for integration.

To illustrate the basic idea of the Monte Carlo integration, consider a simple one-dimensional integration

$$G = \int_a^b g(x)\, dx \tag{6.49}$$

which represents the area under the function $g(x)$, as shown in Fig. 6.6. Two simple Monte Carlo integration techniques are presented here.
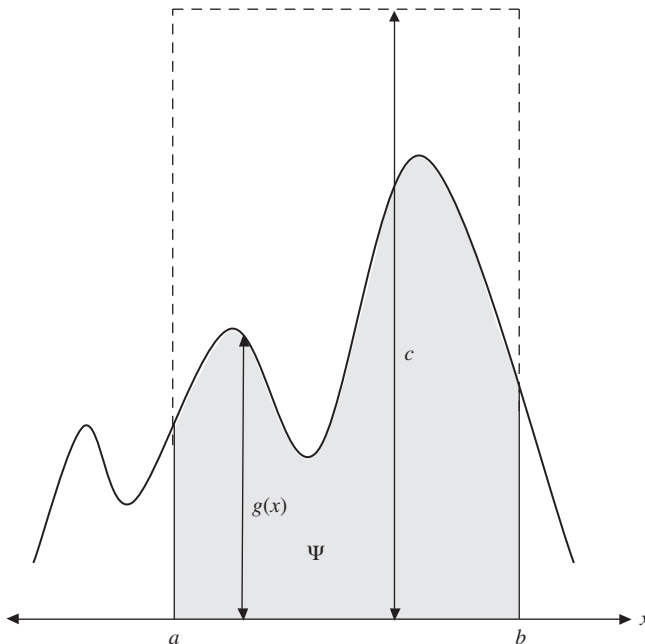
Figure 6.6 Schematic diagram of the hit-and-miss Monte Carlo integration in a one-dimensional integration.

6.6.1   The hit-and-miss method

Referring to Fig. 6.6, a rectangular region $\Re = \{(x, y)|a \leq x \leq b, 0 \leq y \leq c\}$ is superimposed to enclose the area $\Psi = \{(x, y)|a \leq x \leq b, 0 \leq y = g(x) \leq c\}$ represented by Eq. (6.49). By the *hit-and-miss method*, the rectangular region $\Re$ containing the area under $g(x)$, that is, Ψ, is hung on the wall, and one is to throw n darts on it. Assume that the darts fly in a random fashion and that all n darts hit within the rectangular region. The area under $g(x)$, then, can be estimated as the proportion of n darts hitting the target multiplied by the known area of rectangular region $\Re$, that is,

$$\hat{G} = A\left(\frac{n_h}{n}\right) \tag{6.50}$$

where $\hat{G}$ is the estimate of the true area G under $g(x)$, $A = c(b-a)$ is the area of the rectangular region, and n_h is the number of darts hitting the target out of a total of n trials.

The hit-and-miss method can be implemented numerically on a computer. The two coordinates (X_i, Y_i) on the rectangular region $\Re$, which represents the location where the ith dart lands, are treated as two independent random variables that can be generated from two uniform distributions. That is, X_i is generated from $\mathrm{U}(a, b)$ and Y_i from $\mathrm{U}(0, c)$. When $Y_i \leq g(X_i)$, the dart hits its target; otherwise, the dart misses the target. A simple hit-and-miss algorithm is given as follows:

1. Generate $2n$ uniform random variates from $\mathrm{U}(0, 1)$. Form them arbitrarily into n pairs, that is, $(u_1, u'_1), (u_2, u'_2), \ldots, (u_n, u'_n)$.
2. Compute $x_i = a + (b-a)u_i$ and $g(x_i)$, for $i = 1, 2, \ldots, n$.
3. Count the number of cases n_h that $g(x_i) \geq cu'_i$.
4. Estimate the integral G by Eq. (6.50).

Note that $\hat{G}$ is an estimator of the integral G; it is therefore also a random variable. It can be shown that $\hat{G}$ is unbiased, namely,

$$E(\hat{G}) = A \times E\left(\frac{n_h}{n}\right) = Ap = A\left(\frac{G}{A}\right) = G \tag{6.51}$$

where n_h/n, the proportion of n darts hitting the target, is an unbiased estimator of the true probability of hits, and p simply is the ratio of the area under $g(x)$ to the area of the rectangular region. Furthermore, the *standard error* associated with the estimator $\hat{G}$ is

$$\sigma_{\hat{G}} = \sqrt{\frac{G(A-G)}{n}} \tag{6.52}$$

As can be seen, the precision associated with $\hat{G}$, represented by its inverse of standard deviation, using the hit-and-miss Monte Carlo integration method increases with $n^{1/2}$.

A practical question is how many trials have to be carried out so that the estimated $\hat{G}$ satisfies a specified accuracy requirement. In other words, one would like to determine a minimum number of trials n such that the following relationship holds:

$$P(|\hat{G} - G| \leq \varepsilon) \geq \alpha \tag{6.53}$$

in which ε is the specified maximum error between G and $\hat{G}$, and α is the minimum probability that $\hat{G}$ would be within ε around the exact solution. Applying the *Chebyshev inequality*, the minimum number of trials required to achieve Eq. (6.53) can be determined as (Rubinstein, 1981)

$$n \geq \frac{(1-p)p[c(b-a)]^2}{(1-\alpha)\varepsilon^2} = \frac{(1-p)pA^2}{(1-\alpha)\varepsilon^2} \tag{6.54}$$

Note that the required number of trials n increases as the specified error level ε decreases and as the confidence level α increases. In addition, for the specified ε and α, Eq. (6.54) indicates that the required number of trials n can be reduced by letting p approach 1. This implies that selecting an enclosed region $\Re$ as close to Ψ as possible would reduce the required number of trials. However, consideration must be given to the ease of generating random variates for U' in the algorithm.

When the number of trials n is sufficiently large, the random variable T,

$$T = \frac{\hat{G} - G}{s_{\hat{G}}} \tag{6.55}$$

approximately, has the standard normal distribution, that is, $T \sim N(0, 1)$, with s_G being the sample estimator of σ_G, that is,

$$s_{\hat{G}} = \sqrt{\frac{\hat{G}(A - \hat{G})}{n}} \tag{6.56}$$

Hence the $(1 - 2\alpha)$-percent ($\alpha < 0.5$) confidence interval for G then can be obtained as

$$\hat{G} \pm s_{\hat{G}} z_\alpha \tag{6.57}$$

with $z_\alpha = \Phi^{-1}(1 - \alpha)$.

Example 6.6 Suppose that the time to failure of a pump in a water distribution system follows an exponential distribution with the parameter $\beta = 0.0008/h$ (i.e., 7 failures per year). The PDF of the time to failure of the pump can be expressed as

$$f_t(t) = 0.0008e^{-0.0008t} \qquad \text{for } t \geq 0$$

Determine the failure probability of the pump within its first 200 hours of operation by the hit-and-miss algorithm with $n = 2000$. Also compute the standard deviation associated with the estimated failure probability and derive the 95 percent confidence interval containing the exact failure probability.

Solution The probability that the pump would fail within 200 hours can be computed as

$$p_f = \int_0^{200} f_t(t)\,dt = \int_0^{200} 0.0008 e^{-0.0008\,t}\,dt$$

which is the area under the PDF between 0 and 200 hours (Fig. 6.7). Using the hit-and-miss Monte Carlo method, a rectangular area with a height of 0.0010 over the interval [0, 200] is imposed to contain the area representing the pump failure probability.

The area of the rectangle can be easily determined as $A = 0.001(200) = 0.2$. The hit-and-miss algorithm then can be outlined in the following steps:

1. Initialize $i = 0$ and $n_h = 0$.
2. Let $i = i + 1$, and generate a pair of standard uniform random variates (u_i, u_i') from U(0, 1).
3. Let $t_i = 200 u_i$, and compute $f_t(t_i) = 0.0008 e^{-0.0008\,t_i}$, $y = 0.001 u_i'$.
4. If $f_t(t_i) \geq y$, $n_h = n_h + 1$. If $i = 2000$, go to step 5; otherwise, go to step 1.
5. Estimate the pump failure probability as $\hat{p}_f = A(n_h/n) = 0.2(n_h/n)$.

Using the preceding algorithm, 2000 simulations were made, and the estimated pump failure probability is $\hat{p}_f = 0.2(n_h/n) = 0.2(1500/2000) = 0.15$. Comparing with the exact failure probability $p_f = 1 - \exp(-0.16) = 0.147856$, the estimated
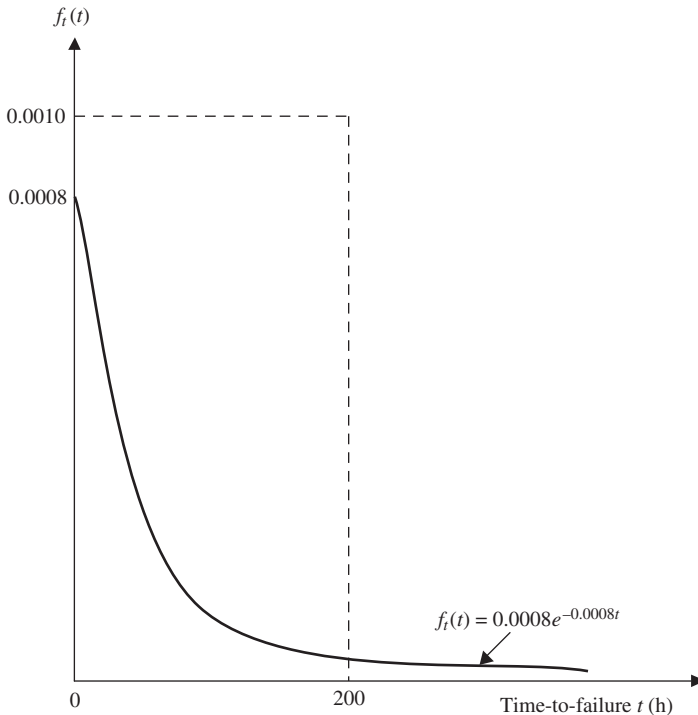
Figure 6.7 The hit-and-miss Monte Carlo integration for Example 6.6.

failure probability by the hit-and-miss method, with $n = 2000$ and the rectangular area chosen, has a 1.45 percent error relative to the exact solution.

The associated standard error can be computed according to Eq. (6.56) as

$$s_{\hat{p}_f} = \sqrt{\frac{\hat{p}_f (A - \hat{p}_f)}{n}} = \sqrt{\frac{0.15(0.2 - 0.15)}{2000}} = 0.00194$$

Assuming normality for the estimated pump failure probability, the 95 percent confidence interval containing the exact failure probability p_f is

$$\hat{p}_f \pm z_{0.975} s_{\hat{p}_f} = (0.1462, 0.1538)$$

where $z_{0.975} = 1.96$.

6.6.2 The sample-mean method

The *sample-mean Monte Carlo integration* is based on the idea that the computation of the integral by Eq. (6.49) alternatively can be carried out by

$$G = \int_a^b \left[\frac{g(x)}{f_x(x)} \right] f_x(x) \, dx \qquad \text{for } a \leq x \leq b \tag{6.58}$$

in which $f_x(x) \geq 0$ is a PDF defined over $a \leq x \leq b$. The transformed integral given by Eq. (6.49) is equivalent to the computation of expectation of $g(X)/f_x(X)$, namely,

$$G = E \left[\frac{g(X)}{f_x(X)} \right] \tag{6.59}$$

with X being a random variable having a PDF $f_x(x)$ defined over $a \leq x \leq b$. The estimation of $E[g(X)/f_x(X)]$ by the sample-mean Monte Carlo integration method is

$$\hat{G} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(x_i)}{f_x(x_i)} \tag{6.60}$$

in which x_i is the random variate generated according to $f_x(x)$, and n is the number of random variates produced. The sample estimator given by Eq. (6.60) has a variance

$$\text{Var}(\hat{G}) = \int_a^b \left[\frac{g(x)}{f_x(x)} \right]^2 f_x(x) \, dx - G^2 \tag{6.61}$$

The sample-mean Monte Carlo integration algorithm can be implemented as follows:

1. Select $f_x(x)$ defined over the region of the integral from which n random variates are generated.

2. Compute $g(x_i)/f_x(x_i)$, for $i = 1, 2, \ldots, n$.

3. Calculate the sample average based on Eq. (6.60) as the estimate for G.

For simplicity, consider that $X \sim U(a, b)$ has a PDF

$$f_x(x) = \frac{1}{b - a} \qquad \text{for } a \leq x \leq b$$

The unbiased estimator of G is the sample mean

$$\hat{G} = \frac{b - a}{n} \sum_{i=1}^{n} g(x_i) \tag{6.62}$$

and the associated with a standard error is

$$\sigma_{\hat{G}} = \sqrt{\frac{b - a}{n} \sum_{i=1}^{n} g^2(x_i) - G^2} \tag{6.63}$$

Example 6.7 Repeat Example 6.6 using the sample-mean Monte Carlo integration algorithm.

Solution Using the sample-mean Monte Carlo integration method, select a uniform distribution over the interval [0, 200]. The required height for the rectangle is 0.005, which satisfies the condition that the area of the rectangle is unity (Fig. 6.8).
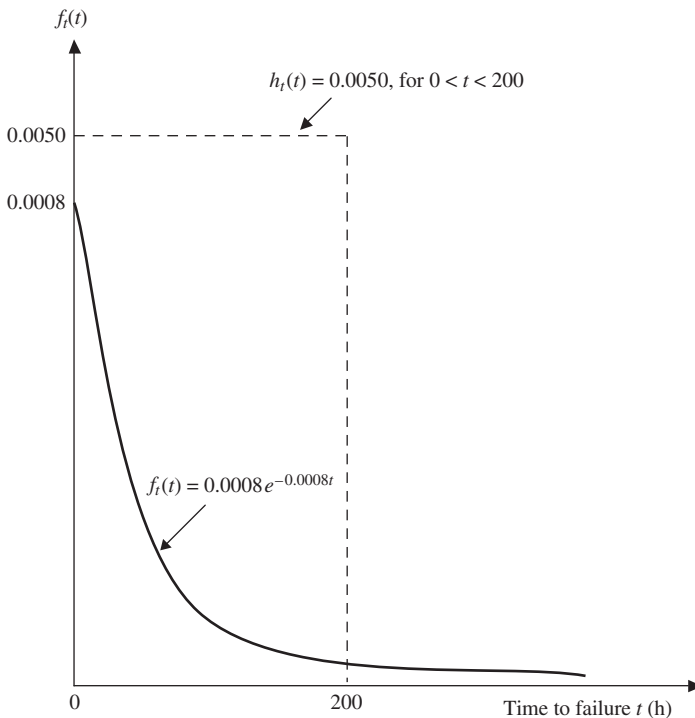
Figure 6.8 The sample-mean Monte Carlo integration for Example 6.7.

The sample-mean algorithm, then, can be outlined as the following:

1. Generate n standard uniform random variates u_i from U(0, 1).
2. Let $t_i = 200u_i$, which is a uniform random variate from U(0, 200), and compute $f_t(t_i)$.
3. Estimate the pump failure probability as

$$\hat{p}_f = \frac{1}{n} \sum_{i=1}^{n} \frac{f_t(t_i)}{h_t(t_i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_t(t_i)}{1/200} = \frac{200}{n} \sum_{i=1}^{n} f_t(t_i)$$

4. To assess the error associated with the estimated pump failure probability by the preceding equation, compute the following quantity:

$$\left\langle \hat{p}_f^2 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{f_t(t_i)}{h_t(t_i)} \right]^2 = \frac{40,000}{n} \sum_{i=1}^{n} [f_t(t_i)]^2$$

where $\langle \cdot \rangle$ is the operator for the mean of the quantity inside.

Using this algorithm for 2000 simulations, the estimated pump failure probability is $\hat{p}_f = 0.14797$. Comparing with the exact failure probability, $p_f = 0.147856$, the estimated failure probability by the sample-mean method, with $n = 2000$ and the simple uniform distribution chosen, has an error of 0.0771 percent relative to the exact solution.

The associated standard error can be computed according to Eq. (6.63) as

$$s_{\hat{p}_f} = \sqrt{\left\langle \hat{p}_f^2 \right\rangle - (\hat{p}_f)^2} = 0.00015$$

Assuming normality for the estimated pump failure probability, the 95 percent confidence interval containing the exact failure probability p_f is

$$\hat{p}_f + 1.96\, s_{\hat{p}_f} = (0.14767, 0.14826)$$

Comparing the solutions with those of Example 6.6, it is observed that for the same number of samples n, the sample-mean algorithm yields a significantly more accurate estimation than the hit-and-miss algorithm. Furthermore, the precision, represented by the standard error, associated with the estimated failure probability by the sample-mean method, is smaller than that of the hit-and-miss algorithm. Consequently, the confidence interval with the same level of significance will be tighter.

6.6.3 Directional Monte Carlo simulation algorithm

Consider the reliability computation involving a multidimensional integral as Eq. (6.48). Without losing generality, the following discussions assume that the stochastic variables in the original X-space have been transformed to the independent standard normal Z'-space (see Sec. 2.7.2). Consequently, the original performance function $W(X)$ can be expressed as $W(Z')$. In terms of Z', Eq. (6.48) can be written as

$$p_s = \int_{W(z') \geq 0} \phi(z')\, dz' \tag{6.64}$$

in which $\phi(\mathbf{z}')$ is the K-dimensional joint PDF of independent standard normal random variables $\mathbf{Z}'$.

Analytical solutions to Eq. (6.64) exist for only a few special cases. For most problems, Eq. (6.64) is solved by approximation methods, such as the first- and second-order reliability methods described in Secs. 4.4 through 4.6. Note that the advanced first-order second-moment methods and the second-order reliability methods require identification of the design point (or points) on the failure surface defined by $W(\mathbf{z}') = 0$. For problems involving multiple design points or a single design point with several points having almost the same distance, they can be cast into the system reliability framework described in Chap. 5. Nevertheless, the process of identifying designing points involves non-linear optimization, by which the search for all design points is a difficult task for problems having a complex failure surface defined by several performance functions (Fig. 6.9).

By the simple Monte Carlo simulation, n sets of random vector $\mathbf{z}'$ are produced to compute the corresponding values of the performance function $W(\mathbf{z}')$. The unbiased estimator of the reliability of a system is the ratio between the number of outcomes in the safe region [with $W(\mathbf{z}') \geq 0$] and the total number of random sets generated n (Fig. 6.10). The Monte Carlo simulation applying simple random sampling, in general, is not efficient, especially when the failure probability is very small. *Directional simulation* is a simple procedure based on the idea of conditional probability to improve the efficiency of the Monte Carlo
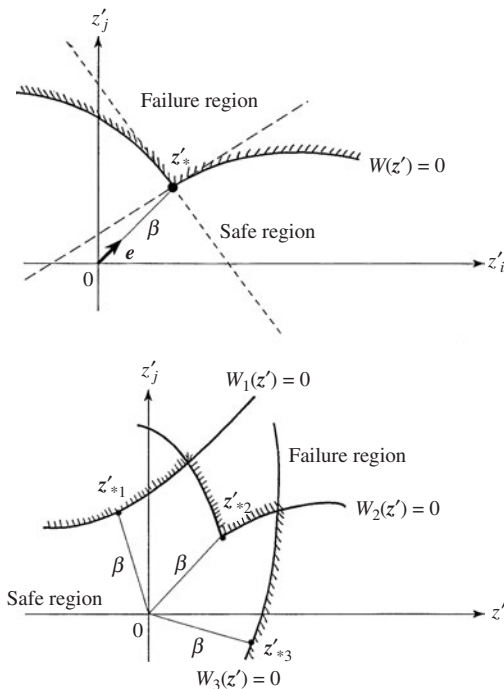
Figure 6.9 Failure surface with multiple design points.
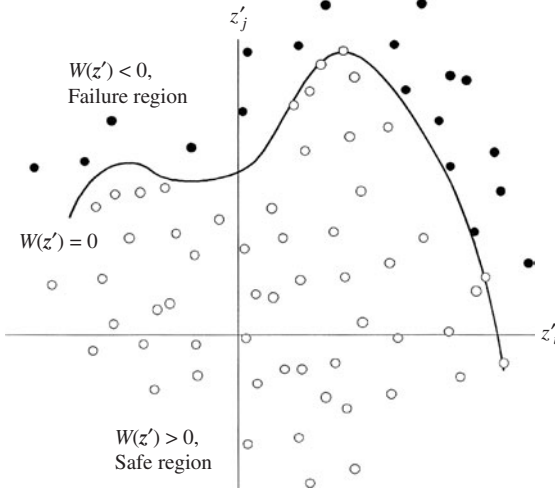
Figure 6.10 Determination of the failure probability by the Monte Carlo simulation with simple random sampling.

solution in Eq. (6.64). The procedure can be applied jointly with the variance reduction techniques described in Sec. 6.7 to further improve the computational efficiency and numerical accuracy of the Monte Carlo simulation of reliability problems.

In the K-dimensional $\boldsymbol{Z}'$-space, any Gaussian random vector $\boldsymbol{Z}'$ can be expressed as

$$\boldsymbol{Z}' = \boldsymbol{R}\boldsymbol{E} \tag{6.65}$$

where $\boldsymbol{R} \geq 0$ is chi-square random variable with K degrees of freedom, and $\boldsymbol{E} = (E_1, E_2, \ldots, E_K)$, an independent random unit vector of length one, that is, $|\boldsymbol{E}| = 1$. The random unit vector $\boldsymbol{E}$ is uniformly distributed on the K-dimensional unit hypersphere R_K. Along a specific direction $\boldsymbol{E} = \boldsymbol{e}$, the conditional reliability $p_s | \boldsymbol{e}$ is

$$p_s | \boldsymbol{e} = P\left[W(\boldsymbol{z}'_e) \geq 0\right] = P\left[W(R\boldsymbol{e}) \geq 0\right] = P\left[R \leq r_e\right] = F_{\chi_K^2}\left(r_e^2\right) \tag{6.66}$$

in which $\boldsymbol{z}'_e = R\boldsymbol{e}$ is a vector having a random length R along the direction defined by the vector $\boldsymbol{e}$, r_e is the distance from the origin to the failure surface along the vector $\boldsymbol{e}$ satisfying $W(r_e\boldsymbol{e}) = 0$, and $F_{\chi_K^2}(\cdot)$ is the χ^2 CDF with K degrees of freedom. The geometric definitions of the terms in Eq. (6.66) are shown in Fig. 6.11. Note that the distance r_e has to be found by a suitable method. For a complicated performance function, numerical root-finding techniques have to be used. If the safe region is nonclosed, the root is $+\infty$ for some $\boldsymbol{e}$. As can be seen, if the failure surface is a hypersphere, the reliability can be found by a single trial in the directional simulation.
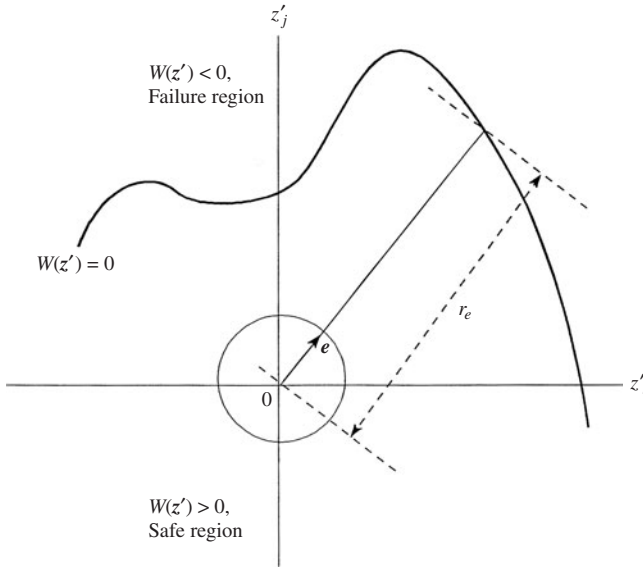
Figure 6.11 Schematic diagram of directional simulation.

From $p_s|e$, the reliability can be obtained using the total probability theorem (Sec. 2.2.4) as

$$p_s = \int_{e \in \boldsymbol{R}_k} (p_s|\boldsymbol{e}) f_e(\boldsymbol{e}) \, d\boldsymbol{e} \tag{6.67}$$

where $f_e(\boldsymbol{e})$ is the density function of random unit vector $\boldsymbol{E}$ on the unit hypersphere, which is a constant. The realization of the random unit vector can be obtained easily as $\boldsymbol{e} = \boldsymbol{z}'/|\boldsymbol{z}'|$, with $\boldsymbol{z}'$ being a randomly generated vector containing K independent standard normal variates. As can be seen from Eq. (6.67), the reliability of the conditional simulation is the expectation of the conditional reliability, that is, $E_e(p_s|\boldsymbol{e})$. Therefore, similar to the sample-mean Monte Carlo integration, the reliability can be estimated as

$$\hat{p}_s = \frac{1}{n} \sum_{i=1}^{n} (p_s|\boldsymbol{e}_i) = \frac{1}{n} \sum_{i=1}^{n} p_{s,i} = \frac{1}{n} \sum_{i=1}^{n} F_{\chi_K^2}\left(r_i^2\right) \tag{6.68}$$

where n is the total number of repetitions in the simulation, $p_{s,i} = p_s|\boldsymbol{e}_i$, $\boldsymbol{e}_i$ is the unit vector randomly generated in the ith repetition, and r_i is the distance from the origin in the $\boldsymbol{Z}'$-space to the failure surface from solving $W(r_i \boldsymbol{e}_i) = 0$. The directional simulation algorithm can be implemented as follows:

1. Transform stochastic variables in the original $\boldsymbol{X}$-space to the independent standard normal $\boldsymbol{Z}'$-space.

2. Generate K independent standard normal random variates $\boldsymbol{z}' = (z_1', z_2', \ldots, z_K')$, and compute the corresponding directional vector $\boldsymbol{e} = \boldsymbol{z}'/|\boldsymbol{z}'|$

3. Determine the distance r_e from the origin to the failure surface by solving $W(r_e \boldsymbol{e}) = 0$.

4. Compute the conditional reliability $p_{s,i} = F_{\chi_K^2}(r_e)$.

5. Repeat steps 2 through 4 n times, obtaining $\{p_{s,1}, p_{s,2}, \ldots, p_{s,n}\}$.

6. Compute the reliability by Eq. (6.68).

The standard error associated with the reliability estimated by Eq. (6.108) is

$$\text{Var}(\hat{p}_s) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (p_{s,i} - \hat{p}_s)^2 \tag{6.69}$$

If the number of samples n is large, the estimated reliability $\hat{p}_s$ can be treated as a normal random variable (according to the central limit theorem), with the variance given by Eq. (6.69). Then the 95 percent confidence interval for the true reliability p_s can be obtained as

$$\hat{p}_s \pm 1.96[\text{Var}(\hat{p}_s)]^{0.5} \tag{6.70}$$

Since the directional simulation yields the exact solution for the reliability integral when the failure surface is a hypersphere in the $\boldsymbol{Z}'$-space, Bjerager (1988) indicated that the procedure will be particularly efficient for problems where the failure surface is "almost spherical." Furthermore, owing to the analytical evaluation of the conditional reliability in Eq. (6.66), the directional simulation will yield a smaller variance on the reliability estimator for a given sample size n than that of the simple random sampling procedure. Bjerager (1988) demonstrated the directional simulation through several examples and showed that the coefficient of variation of estimated reliability $\hat{p}_s$ for a given sample size depends on the shape of the failure surface and the value of the unknown reliability. For nonspherical failure surfaces, the coefficient of variation increases as the dimensionality of the problem K increases.

Example 6.8 Refer to the slope stability problem in Example 6.4. Use the directional simulation to estimate the probability that the excavation can be performed safely within 40 days.

Solution Referring to Eq. (6.29), the problem is to find the probability that the random drawdown recess time will be less than or equal to 40 days, that is,

$$P(T \leq 40) = P\left[\left(\frac{d}{2 \times 0.477}\right)^2 \frac{S}{K_h h_o} \leq 40\right]$$

in which $d = 50$ m, $h_o = 30$ m, and S and K_h are the random storage coefficient and conductivity, having a bivariate normal distribution. The means and standard deviations of S and K_h are, respectively, $\mu_s = 0.05$, $\mu_{kh} = 0.1$ m/day, $\sigma_s = 0.005$,

$\sigma_{kh} = 0.01$ m/day, and their correlation coefficient is $\rho_{kh,s} = 0.5$. The corresponding performance function can be expressed as

$$W(K_h, S) = S - cK_h$$

where $c = 0.43686$.

By the directional simulation outlined earlier, the stochastic variables involved are transformed to the independent standard normal space. For this example, the random conductivity K_h and storage coefficient S can be written in terms of the independent normal random variables Z_1' and Z_2' by spectral decomposition as

$$K_h = 0.1 + 0.005\left(Z_1' + \sqrt{3}Z_2'\right)$$

$$S = 0.05 - 0.0025\left(Z_1' - \sqrt{3}Z_2'\right)$$

For each randomly generated direction vector, defined by $\boldsymbol{z}' = (z_1', z_2')^t$, the components of the corresponding unit vector $\boldsymbol{e} = (e_1, e_2)^t$ can be computed by normalizing the vector $\boldsymbol{z}'$. Therefore, along the directional vector $\boldsymbol{z}'$, the values of the conductivity and storage coefficient can be expressed in terms of the unit vector $\boldsymbol{e}$ and the length of the vector r_e from the origin to the failure surface in the independent standard normal space as

$$K_h = 0.1 + 0.005r_e\left(e_1 + \sqrt{3}e_2\right)$$

$$S = 0.05 - 0.0025r_e\left(e_1 - \sqrt{3}e_2\right)$$

Substituting the preceding expression for K_h and S into the performance function, the failure surface, defined by $W(k_h, s) = W(r_e\boldsymbol{e}) = 0$, can be explicitly written as

$$s - k_h = \left[0.05 - 0.0025r_e\left(e_1 - \sqrt{3}e_2\right)\right] - c\left[0.1 + 0.005r_e(e_1 + \sqrt{3}e_2)\right] = 0$$

Because the performance function in this example is linear, the distance r_e can be solved easily as

$$r_e = \frac{0.006314432}{0.0046842784e_1 - 0.0005468459e_2}$$

For a more complex, nonlinear performance function, proper numerical root-finding procedures must be applied. Furthermore, a feasible direction $\boldsymbol{e}$ should be the one that yields a positive-valued r_e.

The algorithm $P(T \leq 40)$ by the directional simulation for this example can be summarized as follows:

1. Generate two independent standard normal variates z_1' and z_2'.
2. Compute the elements of the corresponding unit vector $\boldsymbol{e}$.
3. Compute the value of distance variable r_e. If $r_e \leq 0$, reject the current infeasible direction and go back to step 1 for a new direction. Otherwise, go to step 4.
4. Compute $P(T \leq 40|\boldsymbol{e}) = 1 - F_{\chi_2^2}(r_e)$, and store the results.
5. Repeat steps 1 throught 4 a large number of times n.
6. Compute the average conditional probability as the estimate for $P(T \leq 40)$ according to Eq. (6.68). Also calculate the associated standard error of the estimate by Eq. (6.69) and the confidence interval.

Based on $n = 400$ repetitions, the directional simulation yields an estimation of $P(T \leq 40) \approx 0.026141$ associated with a standard error of 0.001283. By the normality assumption, the 95 percent confidence interval is $(0.023627, 0.028655)$.

6.6.4 Efficiency of the Monte Carlo algorithm

Referring to Monte Carlo integration, different algorithms yield different estimators for the integral. A relevant issue is which algorithm is more efficient. The efficiency issue can be examined from the statistical properties of the estimator from a given algorithm and its computational aspects. Rubinstein (1981) showed a practical measure of the efficiency of an algorithm by $t \times \text{Var}(\hat{\Theta})$, with t being the computer time required to compute $\hat{\Theta}$, which estimates Θ. Algorithm 1 is more efficient than algorithm 2 if

$$\frac{t_1 \times \text{Var}(\hat{\Theta}_1)}{t_2 \times \text{Var}(\hat{\Theta}_2)} < 1 \tag{6.71}$$

If the computational times for the two algorithms are approximately equal, comparison of efficiency can be made by examining the relative magnitude of the variances. When the true variances are not known, which is generally the case, sample variances can be used. Without considering the computational time, it can be shown that the sample-mean algorithm using $X \sim \text{U}(a, b)$ is more efficient than the hit-and-miss algorithm (see Problem 6.25).

6.7 Variance-Reduction Techniques

Since Monte Carlo simulation is a sampling procedure, results obtained from the procedure inevitably involve sampling errors, which decrease as the sample size increases. Increasing the sample size to achieve a higher precision generally means an increase in computer time for generating random variates and data processing. *Variance-reduction techniques* aim at obtaining high accuracy for the Monte Carlo simulation results without having to substantially increase the sample size. Hence variance-reduction techniques enhance the statistical efficiency of the Monte Carlo simulation. When applied properly, variance-reduction techniques sometimes can make the difference between an impossible, expensive, simulation study and a feasible, useful one.

Variance-reduction techniques attempt to reduce the error associated with the Monte Carlo simulation results by using known information about the problem at hand. Naturally, such an objective cannot be attained if the analyst is completely ignorant about the problem. On the other extreme, the error is zero if the analyst has complete knowledge about the problem. Rubinstein (1981) stated that "variance reduction cannot be obtained from nothing; it is merely a way of not wasting information." Therefore, for a problem that is not known at the initial stage of the study, pilot simulations can be performed for the purpose of gaining useful insight into the problem. The insight, then, can be incorporated later into the variance-reduction techniques for a more efficient

simulation study. Therefore, most of the variance-reduction techniques require additional effort on the part of analysts.

6.7.1 Importance sampling technique

The *importance sampling technique* concentrates the distribution of sampling points in the part of the domain that is most "important" for the task rather than spreading them out evenly (Marshall, 1956). Refer to the problem of evaluating an integral in Eq. (6.49) by the sample-mean method. The importance sampling technique attempts to generate M sampling points to reduce the variance of $\hat{G}$ given by Eq. (6.61).

Rubinstein (1981) showed that the PDF $f_x(x)$ that minimizes Eq. (6.61) can be obtained as

$$f_x(x) = \frac{|g(x)|}{\int |g(x)|\,dx} \qquad (6.72)$$

Although Eq. (6.72) indicates that the weighing function $f_x(x)$ is a function of $\int |g(x)|\,dx$, which is practically equivalent to the integral sought, however, it is not completely useless. Equation (6.72) implies that if is $f_x(x)$ is chosen to have a similar shape as $|g(x)|$, considerable reduction in variance of the simulation results can be achieved. However, in practical implementations of this technique, consideration must be given to the tradeoff between the desired error reduction and the difficulties of sampling from $f_x(x)$, especially when $|g(x)|$ is not well behaved.

Example 6.9 Repeat Example 6.6 using the importance sampling technique with $n = 2000$. The PDF selected is a trapezoidal distribution with a PDF defined as

$$h_t(t) = 0.006 - b \times t \qquad \text{for } 0 \le t \le 200$$

where b is the slope of the dashed line shown in Fig. 6.12. Compare the efficiency of the technique under this distribution with that in Examples 6.6 and 6.7.

Solution Using the importance sampling method with the trapezoidal PDF indicated earlier, evaluation of the pump failure probability can be expressed as

$$p_f = \int_0^{200} \left[\frac{f_t(t)}{h_t(t)} \right] h_t(t)\,dt = E_T \left[\frac{f_t(T)}{h_t(T)} \right]$$

in which t is a dummy variable, and the random variable T has the trapezoidal distribution. Therefore, the pump failure probability p_f can be estimated by computing the expectation of $f_t(t)/h_t(t)$ within $0 \le T \le 200$.

In this case, it is necessary to first determine the coefficient b in $h_t(t)$ such that it is a legitimate PDF over $0 \le t \le 200$. Based on the two conditions for a PDF presented in Sec. 2.3.1, the coefficient b should satisfy the following two conditions:

(i) $$\int_0^{200} h_t(t)\,dt = \int_0^{200} (0.006 - bt)\,dt = 1.0$$

(ii) $$h_t(t) \ge 0 \qquad \text{for } 0 \le t \le 200$$
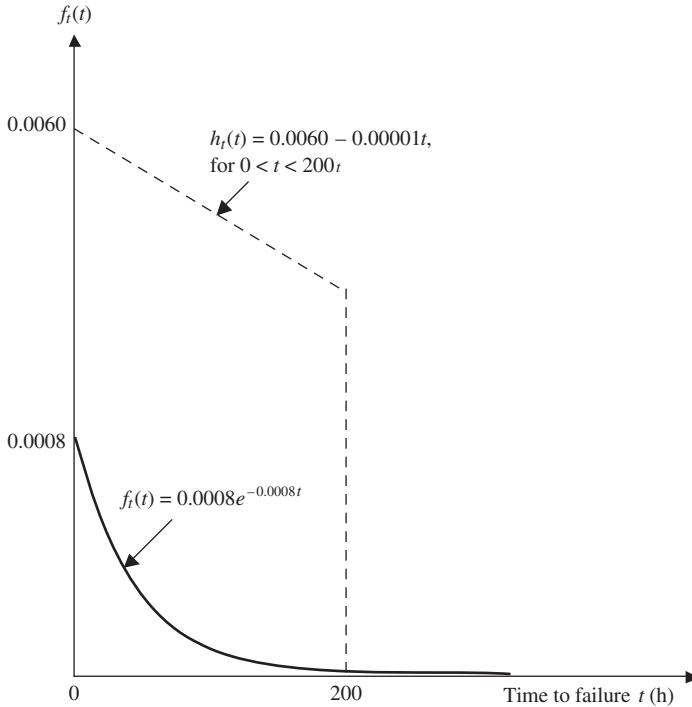
Figure 6.12 Importance sampling for integration in Example 6.9.

From condition (i), the coefficient b can be obtained easily as $b = 0.00001$. Substituting $b = 0.00001$ into $h_t(t)$, one can verify that condition (ii) is also satisfied. Therefore, $h_t(t) = 0.006 - 0.00001 \times t$ is a legitimate PDF in $0 \leq t \leq 200$. To generate random variates from $h_t(t)$, the CDF-inverse method can be used because the CDF of $h_t(t)$ can be obtained easily as

$$H_t(T) = U = \int_0^T (0.006 - 0.00001t)\, dt = 0.006T - \left(\frac{0.00001}{2}\right) T^2$$

Solving for T in terms of U, one obtains

$$T = \frac{2a \pm \sqrt{4a^2 - 8bU}}{2b} \tag{6.73}$$

in which $a = 0.006$ and $b = 0.00001$. Based on the boundary conditions of a CDF, that is, $H_t(t = 0) = 0$ and $H_t(t = 200) = 1$, the only valid expression for T from $h_t(t)$ is

$$T = \frac{2a - \sqrt{4a^2 - 8bU}}{2b}$$

Hence the preceding equation can be used to generate random variates from $h_t(t)$. The algorithm for this example can be outlined as follows:

1. Generate n standard uniform random variates u_i from U(0, 1).
2. Compute t_i according to Eq. (6.73), yielding n random variates from $h_t(t) = 0.006 - 0.00001t, 0 \leq t \leq 200$.

3. Estimate the pump failure probability as

$$\hat{p}_f = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{f_t(t_i)}{h_t(t_i)} \right] = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{0.0008 e^{-0.0008 t_i}}{0.006 - 0.00001 t_i} \right)$$

4. To assess the error associated with the estimated pump failure probability by the preceding equation, compute the following quantity:

$$s_{\hat{p}_f} = \sqrt{\frac{\mathrm{Var}\left[\frac{f_t(T)}{h_t(T)} \right]}{M}} = \sqrt{\langle \hat{p}_f{}^2 \rangle} \tag{6.74}$$

in which $\mathrm{var}[f_t(T)/h_t(T)]$ is computed as

$$\mathrm{Var}\left[\frac{f_t(T)}{h_t(T)} \right] = \frac{1}{M} \sum_{i=1}^{M} \left[\frac{f_t(t_i)}{h_t(t_i)} \right]^2 - \hat{p}_f^2 = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{0.0008 e^{-0.0008 t_i}}{0.006 - 0.0001 t_i} \right)^2 - \hat{p}_f^2$$

Using this algorithm, the estimated pump failure probability is $\hat{p}_f = 0.14767$, whereas the exact failure probability is 0.147856. The associated standard error can be computed as

$$s_{\hat{p}_f} = \sqrt{\langle \hat{p}_f^2 \rangle - \left(\hat{p}_f^2 \right)} = 0.00023$$

Assuming normality for the estimated pump failure probability, the 95 percent confidence interval containing the exact risk p_f is

$$\hat{p}_f \pm 1.96 s_{\hat{p}_f} = (0.14722, 0.14813)$$

Comparing the solutions with those of Examples 6.6 and 6.7, it is observed that with the same $n = 2000$, the importance sampling method has a 0.126 percent error. The magnitude of error and the accuracy level are somewhat worse than the sample-mean procedure but are still far better than the hit-and-miss algorithm. The accuracy of the importance sampling technique depends on how $h_t(t)$ is specified. As can be observed from Problem 6.29, use of the exponential function for $f_t(t)$ improves the accuracy tremendously.

6.7.2 Antithetic-variates technique

The *antithetic-variates technique* (Hammersley and Morton, 1956) achieves the variance-reduction goal by attempting to generate random variates that would induce a negative correlation for the quantity of interest between separate simulation runs. Consider that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of an unknown quantity θ to be estimated. The two estimators can be combined together to form another estimator as

$$\hat{\Theta}_a = \frac{1}{2}(\hat{\Theta}_1 + \hat{\Theta}_2) \tag{6.75}$$

The new estimator $\hat{\Theta}_a$ also is unbiased and has a variance as

$$\mathrm{Var}(\hat{\Theta}_a) = \frac{1}{4}[\mathrm{Var}(\hat{\Theta}_1) + \mathrm{Var}(\hat{\Theta}_2) + 2\mathrm{Cov}(\hat{\Theta}_1, \hat{\Theta}_2)] \tag{6.76}$$

If the two estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ were computed by Monte Carlo simulation through generating two independent, sets of random variates, they would be independent, and the variance for $\hat{\Theta}_a$ would be

$$\text{Var}(\hat{\Theta}_a) = \frac{1}{4}[\text{Var}(\hat{\Theta}_1) + \text{Var}(\hat{\Theta}_2)] \tag{6.77}$$

From Eq. (6.76) one realizes that the variance associated with $\hat{\Theta}_a$ could be reduced if the Monte Carlo simulation can generate random variates, which result in a strong negative correlation between $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

In a Monte Carlo simulation, the values of estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are functions of the generated random variates, which, in turn, are related to the standard uniform random variates. Therefore, $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are functions of the two standard uniform random variables U_1 and U_2. The objective to produce negative $\text{Cov}[\hat{\Theta}_1(U_1), \hat{\Theta}_2(U_2)]$ can be achieved by producing U_1 and U_2, which are negatively correlated. However, it would not be desirable to complicate the computational procedure by generating two sets of uniform random variates subject to the constraint of being negatively correlated. One simple approach to generate negatively correlated uniform random variates with minimal computation is to let $U_1 = 1 - U_2$. It can be shown that $\text{Cov}(U, 1-U) = -1/12$ (see Problem 6.31). Hence a simple antithetic-variates algorithm is the following:

1. Generate u_i from U(0, 1), and compute $1 - u_i$, for $i = 1, 2, \ldots, n$.
2. Compute $\hat{\theta}_1(u_i)$, $\hat{\theta}_2(1 - u_i)$, and then $\hat{\theta}_a$ according to Eq. (6.75).

Example 6.10 Develop a Monte Carlo algorithm using the antithetic-variates technique to evaluate the integral G defined by

$$G = \int_a^b g(x)\,dx$$

in which $g(x)$ is a given function.

Solution Applying the Monte Carlo method to estimate the value of G, the preceding integral can be rewritten as

$$G = \int_a^b \left[\frac{g(x)}{f_x(x)}\right] f_x(x)\,dx = E\left[\frac{g(X)}{f_x(X)}\right]$$

where $f_x(x)$ is the adopted distribution function based on which random variates are generated. As can be seen, the original integral becomes the calculation of the expectation of the ratio of $g(X)$ and $f_x(X)$. Hence the two estimators for G using the antithetic-variates technique can be formulated as

$$\hat{G}_1 = \frac{1}{n}\sum_{i=1}^n \frac{g(X_{1i})}{f_x(X_{1i})} \tag{6.78a}$$

$$\hat{G}_2 = \frac{1}{n}\sum_{i=1}^n \frac{g(X_{2i})}{f_x(X_{2i})} \tag{6.78b}$$

in which $X_{1i} = F_x^{-1}(U_i)$ and $X_{2i} = F_x^{-1}(1 - U_i)$, with $F_x(\cdot)$ being the CDF of the random variable X. The algorithm for the Monte Carlo integral using the antithetic-variates technique is

1. Generate n uniform random variates u_i from U(0, 1), and compute the corresponding $1 - u_i$.
2. Compute $g(x_{1i})$, $f_x(x_{1i})$, $g(x_{2i})$, and $f_x(x_{2i})$, with $x_{1i} = F_x^{-1}(u_i)$ and $x_{2i} = F_x^{-1}(1 - u_i)$.
3. Calculate the values of $\hat{G}_1$ and $\hat{G}_2$ by Eqs. (6.78a) and (6.78b), respectively. Then estimate G by $\hat{G}_a = (\hat{G}_1 + \hat{G}_2)/2$.

In the case that X has a uniform distribution as $f_x(x) = 1/(b - a), a \leq x \leq b$, the estimate of G by the antithetic-variates technique can be expressed as

$$\hat{g}_a = \frac{b - a}{2n} \sum_{i=1}^{n} [g(x_{1i}) + g(x_{2i})] \tag{6.79}$$

Rubinstein (1981) showed that the antithetic-variates estimator, in fact, is more efficient if $g(x)$ is a continuous monotonically increasing or decreasing function with continuous first derivatives.

Example 6.11   Referring to pump reliability Example 6.6, estimate the pump failure probability using the antithetic-variates technique along with the sample-mean Monte Carlo algorithm with $n = 1000$. The PDF selected is a uniform distribution U(0, 200). Also, compare the results with those obtained in Examples 6.6, 6.7, and 6.8.

Solution   Referring to Example 6.7, uniform distribution U(0, 200) has a height of 0.005 (see Fig. 6.8). The antithetic-variate method along with the sample-mean Monte Carlo algorithm for evaluating the pump failure probability can be outlined as follows:

1. Generate n pairs standard uniform random variates $(u_i, 1 - u_i)$ from U(0, 1).
2. Let $t_{1i} = 200 \, u_i$ and $t_{2i} = 200(1 - u_i)$. Compute $f_t(t_{1i})$ and $f_t(t_{2i})$.
3. Estimate the pump failure probability, according to Eq. (6.79), as

$$\hat{p}_{f,a} = \frac{200}{2n} \sum_{i=1}^{n} [f_t(t_{1i}) + f_t(t_{2i})]$$

Using this algorithm, the estimated pump failure probability is $\hat{p}_f = 0.14785$. Comparing with the exact failure probability $p_f = 0.147856$, the estimated failure probability by the antithetic-variates algorithm with $n = 1000$ and the simple uniform distribution is accurate within 0.00406 percent. The standard deviation s associated with the $2n$ random samples is 0.00669. According to Eq. (6.74), the standard error associated with $\hat{p}_{f,a}$ can be computed as $s/\sqrt{2n} = 0.00015$. The skewness coefficient from the $2n$ random samples is 0.077, which is close to zero. Hence, by the normality approximation, the 95 percent confidence interval containing the exact failure probability p_f is (0.14756, 0.14814).

Comparing the solutions with those of Examples 6.6, 6.7, and 6.9, it is observed that the antithetic-variate algorithm is very accurate in estimating the probability.

6.7.3 Correlated-sampling techniques

Correlated-sampling techniques are especially effective for variance reduction when the primary objective of the simulation study is to evaluate small changes in system performance or to compare the difference in system performances between two specific designs (Rubinstein, 1981; Ang and Tang, 1984). Consider that one wishes to estimate

$$\Delta\Theta = \Theta_1 - \Theta_2 \tag{6.80}$$

in which

$$\Theta_1 = \int g_1(x) f_1(x)\, dx = E[g_1(X)]$$

$$\Theta_2 = \int g_2(y) f_2(y)\, dy = E[g_2(Y)] \tag{6.81}$$

with $f_1(x)$ and $f_2(y)$ being two different PDFs. By Monte Carlo simulation, $\Delta\Theta$ can be estimated as

$$\widehat{\Delta\Theta} = \hat{\Theta}_1 - \hat{\Theta}_2 = \frac{1}{n}\left[\sum_{i=1}^{n} g_1(X_i) - \sum_{i=1}^{n} g_2(Y_i)\right] = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Delta\Theta}_i \tag{6.82}$$

in which X_i and Y_i are random samples generated from $f_1(x)$ and $f_2(y)$, respectively, and $\widehat{\Delta\Theta}_i = g_1(X_i) - g_2(Y_i)$.

The variance associated with $\widehat{\Delta\Theta}$ is

$$\text{Var}(\widehat{\Delta\Theta}) = \text{Var}(\hat{\Theta}_1) + \text{Var}(\hat{\Theta}_2) - 2\text{Cov}(\hat{\Theta}_1, \hat{\Theta}_2) \tag{6.83}$$

In the case that random variates X_i and Y_i are generated independently in the Monte Carlo algorithm, $\hat{\Theta}_1$ and $\hat{\Theta}_2$ also would be independent random variables. Hence $\text{Var}(\widehat{\Delta\Theta}) = \text{Var}(\hat{\Theta}_1) + \text{Var}(\hat{\Theta}_2)$.

Note that from Eq. (6.83), $\text{Var}(\widehat{\Delta\hat{\Theta}})$ can be reduced if positively correlated random variables $\hat{\Theta}_1$ and $\hat{\Theta}_2$ can be produced to estimate $\widehat{\Delta\Theta}$. One easy way to obtain positively correlated samples is to use the same sequence of uniform random variates from U(0, 1) in both simulations. That is, the random sequences $\{X_1, X_2, \ldots, X_n\}$ and $\{Y_1, Y_2, \ldots, Y_n\}$ are generated through $X_i = F_1^{-1}(U_i)$ and $Y_i = F_2^{-1}(U_i)$, respectively.

The correlated-sampling techniques are especially effective in reducing variance when the performance difference between two specific designs for a system involve the same or similar random variables. For example, consider two designs A and B for the same system involving a vector of K random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)$, which could be correlated with a joint PDF $f_x(\boldsymbol{x})$ or be independent of each other with a marginal PDF $f_k(x_k), k = 1, 2, \ldots, K$. The performance of the system under the two designs can be expressed as

$$\Theta_A = g(\boldsymbol{a}, \boldsymbol{X}) \qquad \Theta_B = g(\boldsymbol{b}, \boldsymbol{X}) \tag{6.84}$$

in which $g(\cdot)$ is a function defining the system performance, and $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors of design parameters corresponding to designs A and B, respectively. Since the two performance measures Θ_A and Θ_B are dependent on the same random variables through the same performance function $g(\cdot)$, their estimators will be positively correlated. In this case, independently generating two sets of K random variates, according to their probability laws for designs A and B, still would result in a positive correlation between $\hat{\Theta}_A$ and $\hat{\Theta}_B$. To further reduce $\text{Var}(\widehat{\Delta\Theta})$, an increase in correlation between $\hat{\Theta}_A$ and $\hat{\Theta}_B$ can be achieved using a common set of standard uniform random variates for both designs A and B by assuming that system random variables are independent, that is,

$$\theta_{A,i} = g\left[\boldsymbol{a}, F_1^{-1}(u_{1i}), F_2^{-1}(u_{2i}), \ldots, F_K^{-1}(u_{Ki})\right] \qquad i = 1, 2, \ldots, n \qquad \text{(6.85a)}$$

$$\theta_{B,i} = g\left[\boldsymbol{b}, F_1^{-1}(u_{1i}), F_2^{-1}(u_{2i}), \ldots, F_K^{-1}(u_{Ki})\right] \qquad i = 1, 2, \ldots, n \qquad \text{(6.85b)}$$

in which $x_{ki} = F_k^{-1}(u_{ki})$ is the inverse CDF for the kth random variable X_k operating on the kth standard uniform random variate for the ith simulation.

Example 6.12 Refer to the pump reliability problem that has been studied in previous examples. Now consider a second pump the time-to-failure PDF of which also is an exponential distribution but has a different parameter of $\beta = 0.0005$/h. Estimate the difference in the failure probability between the two pumps over the time interval [0, 200 h] using the correlated-sampling technique with $n = 2000$.

Solution Again, the sample-mean Monte Carlo method with a uniform distribution U(0, 200) is applied as in Example 6.7. In this example, the same set of standard uniform random variates $\{u_1, u_2, \ldots, u_{2000}\}$ from U(0, 1) is used to estimate the failure probabilities for the two pumps as

$$\hat{p}_{f,A} = \frac{200}{n} \sum_{i=1}^{n} \left(0.0008e^{-0.0008t_i}\right)$$

$$\hat{p}_{f,B} = \frac{200}{n} \sum_{i=1}^{n} \left(0.0005e^{-0.0005t_i}\right)$$

in which $t_i = 200u_i$, for $i = 1, 2, \ldots, 2000$. The difference in failure probabilities can be estimated as

$$\widehat{\Delta p_f} = \hat{p}_{f,A} - \hat{p}_{f,B} = 0.05276$$

which is within 0.125 percent of the exact solution $e^{-0.0005(200)} - e^{-0.0008(200)} = e^{-0.1} - e^{-0.16} = 0.0526936$.

The standard deviation of the 2000 differences in failure probability $\Delta_i = 200[\hat{f}_A(t_i) - \hat{f}_B(t_i)]$, $i = 1, 2, \ldots, 2000$, is 0.00405. Hence the standard error associated with the estimated difference in failure probability is $0.00405/\sqrt{2000} = 0.00009$.

For the sake of examining the effectiveness of the correlated-sampling technique, let us separately generate a set of independent standard uniform random variates $\{u'_1, u'_2, \ldots, u'_{2000}\}$ and use them in calculating the failure probability for pump B. Then the estimated difference in failure probability between the two pumps is 0.05256, which is slightly larger than that obtained by the correlated-sampling technique. However, the standard error associated with $\Delta_i = 200[\hat{f}_A(t_i) - \hat{f}_B(t_i)]$ then is 0.00016, which is larger than that from the correlated-sampling technique.

6.7.4 Stratified sampling technique

The *stratified sampling* technique is a well-established area in statistical sampling (Cochran, 1966). Variance reduction by the stratified sampling technique is achieved by taking more samples in important subregions. Consider a problem in which the expectation of a function $g(X)$ is sought, where X is a random variable with a PDF $f_x(x), x \in \Xi$. Referring to Fig. 6.13, the domain Ξ for the random variable X is divided into M disjoint subregions $\Xi_m, m = 1, 2, \ldots, M$. That is,

$$\Xi = \bigcup_{m=1}^{M} \Xi_m \qquad \emptyset = \Xi_m \cap \Xi_{m'} \qquad m \neq m'$$
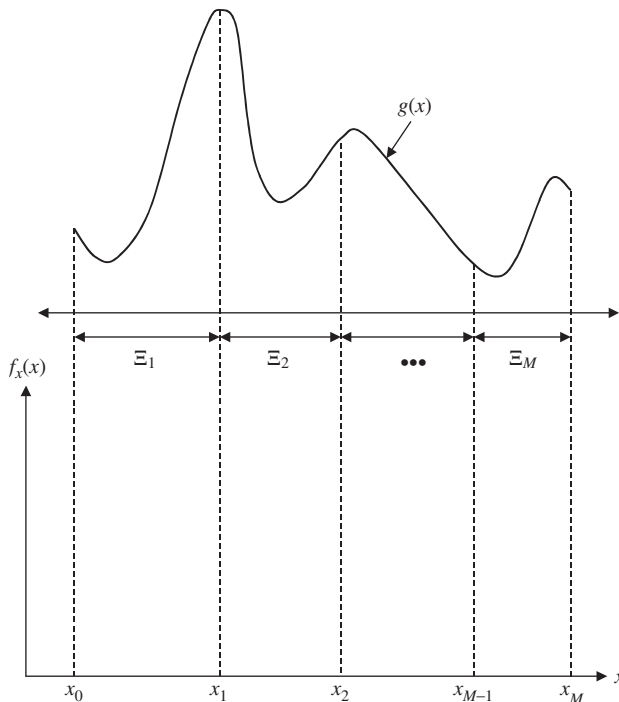
Figure 6.13 Schematic diagram of stratified sampling.

Let p_m be the probability that random variable X will fall within the sub-region Ξ_m, that is, $\int_{x \in \Xi_m} f_x(x)\,dx = p_m$. Therefore, it is true that $\Sigma_m p_m = 1$. The expectation of $g(X)$ can be computed as

$$G = \int_{\Xi} g(x) f_x(x)\,dx = \sum_{m=1}^{M} \int_{\Xi_m} g(x) f_x(x)\,dx = \sum_{m=1}^{M} G_m \qquad (6.86)$$

where $G_m = \int_{\Xi_m} g(x) f_x(x)\,dx$.

Note that the integral for G_m can be written as

$$G_m = p_m \int_{\Xi_m} g(x) \left[\frac{f_x(x)}{p_m} \right] dx = p_m E[g_m(X)] \qquad (6.87)$$

and it can be estimated by the Monte Carlo method as

$$\hat{G}_m = \frac{p_m}{n_m} \sum_{m=1}^{n_m} g(X_m) \qquad m = 1, 2, \ldots, M \qquad (6.88)$$

where n_m is the number of sample points in the mth subregion, and $\Sigma_m n_m = n$, the total number of random variates to be generated. Therefore, the estimator for G in Eq. (6.86) can be obtained as

$$\hat{G} = \sum_{m=1}^{M} \hat{G}_m = \sum_{m=1}^{M} \frac{p_m}{n_m} \left[\sum_{i=1}^{n_m} g(X_{mi}) \right] \qquad (6.89)$$

After the number of subregions M and the total number of samples n are determined, an interesting issue for the stratified sampling is how to allocate the total n sample points among the M subregions such that the variance associated with $\hat{G}$ by Eq. (6.89) is minimized. A theorem shows that the optimal n_m^* that minimizes $\text{Var}(\hat{G})$ in Eq. (6.89) is (Rubinstein, 1981)

$$n_m^* = n \left(\frac{p_m \sigma_m}{\sum_{m'=1}^{M} p_{m'} \sigma_{m'}} \right) \qquad (6.90)$$

where σ_m is the standard deviation associated with the estimator $\hat{G}_m$ in Eq. (6.88).

In general, information about σ_m is not available in advance. It is suggested that a pilot simulation study be made to obtain a rough estimation about the value of σ_m, which serves as the basis in the follow-up simulation investigation to achieve the variance-reduction objective.

A simple plan for sample allocation is $n_m = np_m$ after the subregions are specified. It can be shown that with this sampling plan, the variance associated with $\hat{G}$ by Eq. (6.89) is less than that from the simple random-sample technique. One efficient stratified sampling technique is *systematic sampling* (McGrath, 1970), in which $p_m = 1/M$ and $n_m = n/M$. The algorithm of the systematic sampling can be described as follows:

1. Divide interval [0, 1] into M equal subintervals.
2. Within each subinterval, generate n/M uniform random numbers $u_{mi} \sim$ U[$(m - 1)/n, m/n$], $m = 1, 2, \ldots, M$; $i = 1, 2, \ldots, n/m$.
3. Compute $x_{mi} = F_x^{-1}(u_{mi})$.
4. Calculate $\hat{G}$ according to Eq. (6.89).

Example 6.13 Referring to Example 6.7, apply the systematic sampling technique to evaluate the pump failure probability in the time interval [0, 200 h].

Solution Again, let us adopt the uniform distribution U(0, 200) and carry out the computation by the sample-mean Monte Carlo method. In the systematic sampling, the interval [0, 200] is divided into 10 equal-probability subintervals, each having a probability content of 0.1. Since $h(t) = 1/200, 0 \leq t \leq 200$, the end points of each subinterval can be obtained easily as

$$t_0 = 0, t_1 = 20, t_2 = 40, \ldots, t_9 = 180, t_{10} = 200$$

Furthermore, let us generate $n_m = 200$ random variates from each subinterval so that $\Sigma_m n_m = 2000$. This can be achieved by letting

$$U_{mi} \sim \mathrm{U} \left(\frac{20(m - 1)}{10}, \frac{20m}{10} \right) \qquad \text{for } i = 1, 2, \ldots, 200; m = 1, 2, \ldots, 10$$

The algorithm for estimating the pump failure probability is the following:

1. Initialize subinterval index $m = 0$.
2. Let $m = m + 1$. Generate $n_m = 200$ standard uniform random variates $\{u_{m1}, u_{m2}, \ldots, u_{m,200}\}$, and transform them into the random variates from the corresponding subinterval by $t_{mi} = 20(m - 1) + 20 u_{mi}$, for $i = 1, 2, \ldots, 200$.
3. Compute $\hat{p}_{f,m}$ as

$$\hat{p}_{f,m} = \frac{0.1}{200} \sum_{m_i=1}^{200} f_t(t_{m_i})$$

and the associated variance as

$$\mathrm{Var}(\hat{p}_{f,m}) = \frac{p_m^2 s_m^2}{n_m} = \frac{0.1^2 s_m^2}{200}$$

in which s_m is the standard deviation of 200 $f_t(t_{mi})$ for the mth subinterval.
4. If $m < 10$, go to step 2; otherwise, compute the pump failure probability as

$$\hat{p}_f = \frac{1}{10} \sum_{m=1}^{10} \hat{p}_{f,m}$$

and the associated standard error as

$$s_{\hat{p}_f} = \frac{1}{10} \left[\sum_{m=1}^{10} \mathrm{Var}(\hat{p}_{f,m}) \right]^{1/2}$$

The results from the numerical simulation are shown below:

m	$\hat{p}_{f,m}$	s_m	m	$\hat{p}_{f,m}$	s_m
1	0.15873	0.00071102	6	0.14659	0.00066053
2	0.15626	0.00069358	7	0.14423	0.00064361
3	0.15374	0.00069298	8	0.14194	0.00064993
4	0.15121	0.00072408	9	0.13968	0.00066746
5	0.14887	0.00065434	10	0.13742	0.00067482
	All	0.14787	0.15154×10^{-5}		

The value of p_f is extremely close to the exact solution of 0.147856.

6.7.5   Latin hypercube sampling technique

The *Latin hypercube sampling (LHS) technique* is a special method under the umbrella of stratified sampling that selects random samples of each random variable over its range in a stratified manner. Consider a multiple integral involving K random variables

$$G = \int_{\boldsymbol{x} \in \Xi} g(\boldsymbol{x}) f_x(\boldsymbol{x}) \, d\boldsymbol{x} = E[g(\boldsymbol{X})] \tag{6.91}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^t$ is an K-dimensional vector of random variables, and $f_x(\boldsymbol{x})$ is their joint PDF.

The LHS technique divides the plausible range of each random variable into $M(M \geq K$ in practice) equal-probability intervals. Within each interval, a single random variate is generated resulting in M random variates for each random variable. The expected value of $g(\boldsymbol{X})$, then, is estimated as

$$\hat{G} = \frac{1}{M} \sum_{m=1}^{M} g(X_{1m}, X_{2m}, \ldots, X_{Km}) \tag{6.92}$$

where X_{km} is the variate generated for the kth random variable X_k in the mth set.

More specifically, consider a random variable X_k over the interval of $[\underline{x}_k, \bar{x}_k]$ following a specified PDF $f_k(x_k)$. The range $[\underline{x}_k, \bar{x}_k]$ is partitioned into M intervals, that is,

$$\underline{x}_k = x_{k0} < x_{k1} < x_{k2} < \cdots < x_{k,M-1} < x_{kM} = \bar{x}_k \tag{6.93}$$

in which $P(x_{km} \leq X_k \leq x_{k,m+1}) = 1/M$ for all $m = 0, 1, 2, \ldots, M-1$. The end points of the intervals are determined by solving

$$F_k(x_{km}) = \int_{\underline{x}_k}^{x_{km}} f_k(x_k) \, dx_k = \frac{m}{M} \tag{6.94}$$

where $F_k(\cdot)$ is the CDF of the random variable X_k. The LHS technique, once the end points for all intervals are determined, randomly selects a single value

in each of the intervals to form the M samples set for X_k. The sample values can be obtained by the CDF-inverse or other appropriate method.

To generate M values of random variable X_k from each of the intervals, a sequence of probability values $\{p_{k1}, p_{k2}, \ldots, p_{k,M-1}, p_{kM}\}$ is generated as

$$p_{km} = \frac{m-1}{M} + \zeta_{km} \qquad m = 1, 2, \ldots, M \tag{6.95}$$

in which $\{\zeta_{k1}, \zeta_{k2}, \ldots, \zeta_{k,M-1}, \zeta_{kM}\}$ are independent uniform random numbers from $\zeta \sim U(0, 1/M)$. After $\{p_{k1}, p_{k2}, \ldots, p_{k,M-1}, p_{kM}\}$ are generated, the corresponding M random samples for X_k can be determined as

$$x_{km} = F_k^{-1}(p_{km}) \qquad m = 1, 2, \ldots, M \tag{6.96}$$

Note that p_{km} determined by Eq. (6.96) follows

$$p_{k1} < p_{k2} < \cdots < p_{km} < \cdots < p_{k,M-1} < p_{kM} \tag{6.97}$$

and accordingly,

$$x_{k1} \leq x_{k2} \leq \cdots \leq x_{km} \leq \cdots \leq x_{k,M-1} \leq x_{kM} \tag{6.98}$$

To make the generated $\{x_{k1}, x_{k2}, \ldots, x_{k,M-1}, x_{kM}\}$ a random sequence, random permutation can be applied to randomize the sequence. Alternatively, Latin hypercube samples for K random variables with size M can be generated by (Pebesma and Heuvelink, 1999), that is,

$$x_{km} = F_k^{-1}\left(\frac{s_{km} - u_{km}}{M}\right) \tag{6.99}$$

where s_{km} is a random permutation of 1 to M, and u_{km} is a uniformly distributed random variate in [0, 1]. Figure 6.14 shows the allocation of six samples by the LHS technique for a problem involving two random variables. It is seen that in each row or column of the 6×6 matrix only one cell contains a generated sample. The LHS algorithm can implemented as follows:

1. Select the number of subintervals M for each random variable, and divide the plausible range into M equal-probability intervals according to Eq. (6.94).
2. Generate M standard uniform random variates from U(0, 1/M).
3. Determine a sequence of probability values p_{km}, for $k = 1, 2, \ldots, K; m = 1, 2, \ldots, M$, using Eq. (6.95).
4. Generate random variates for each of the random variables using an appropriate method, such as Eq. (6.96).
5. Randomly permutate generated random sequences for all random variables.
6. Estimate G by Eq. (6.92).

Using the LHS technique, the usual estimators of G and its distribution function are unbiased (McKay, 1988). Moreover, when the function $g(\boldsymbol{X})$ is
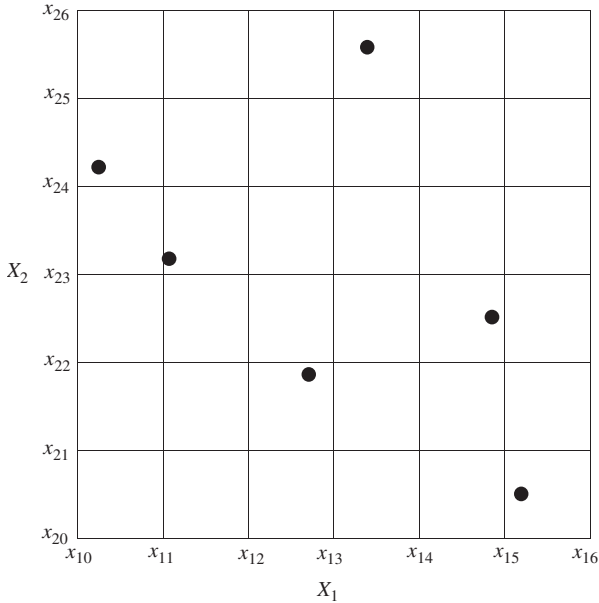
Figure 6.14 Schematic diagram of the Latin hypercube sampling (LHS) technique.

monotonic in each of the X_k, the variances of the estimators are no more than and often less than the variances when random variables are generated from simple random sampling. McKay (1988) suggested that the use of twice the number of involved random variables for sample size ($M \geq 2K$) would be sufficient to yield accurate estimation of the statistics model output. Iman and Helton (1985) indicated that a choice of M equal to 4/3K usually gives satisfactory results. For a dynamic stream water-quality model over a 1-year simulation period, Manache (2001) compared results from LHS using $M = 4/3K$ and $M = 3K$ and found reasonable convergence in the identification of the most sensitive parameters but not in calculation of the standard deviation of model output. Thus, if it is computationally feasible, the generation of a larger number of samples would further enhance the accuracy of the estimation. Like all other variance-reduction Monte Carlo techniques, LHS generally would require fewer samples or model evaluations to achieve an accuracy level comparable with that obtained from a simple random sampling scheme. In hydrosystems engineering, the LHS technique has been applied widely to sediment transport (Yeh and Tung, 1993; Chang et al., 1993), water-quality modeling (Jaffe and Ferrara, 1984; Melching and Bauwens, 2001; Sohrabi et al., 2003; Manache and Melching, 2004), and rainfall-runoff modeling (Melching, 1995; Yu et al., 2001; Christiaens and Feyen, 2002; Lu and Tung, 2003).

Melching (1995) compared the results from LHS with $M = 50$ with those from Monte Carlo simulation with 10,000 simulations and also with those from FOVE and Rosenbleuth's method for the case of using HEC-1 (U.S. Army Corps

of Engineers, 1991) to estimate flood peaks for a watershed in Illinois. All methods yielded similar estimates of the mean value of the predicted peak flow. The variation of standard deviation estimates among the methods was much greater than that of the mean value estimates. In the estimation of the standard deviation of the peak flow, LHS was found to provide the closest agreement to Monte Carlo simulation, with an average error of 7.5 percent and 10 of 16 standard deviations within 10 percent of the value estimated with Monte Carlo simulation. This indicates that LHS can yield relatively accurate estimates of the mean and standard deviation of model output at a far smaller computational burden than Monte Carlo simulation. A detailed description of LHS, in conjunction with the regression analysis for uncertainty and sensitivity analysis, can be found elsewhere (Tung and Yen, 2005, Sec. 6.8).

Example 6.14 Referring to Example 6.7, apply the Latin hypercube sampling technique to evaluate the pump failure probability in the time interval [0, 200 h].

Solution Again, the uniform distribution U(0, 200) is selected along with the sample-mean Monte Carlo method for carrying out the integration. In Latin hypercube sampling, the interval [0, 200] is divided into 1000 equal-probability subintervals, with each having a probability of 0.001. For U(0, 200), the end points of each subinterval can be obtained easily as

$$t_0 = 0, t_1 = 0.2, t_2 = 0.4, \ldots, t_{999} = 199.8, t_{1000} = 200$$

By the LHS, one random variate for each subinterval is generated. In other words, generate a single random variate from

$$U_m \sim U[0.2(m-1), 0.2m] \qquad m = 1, 2, \ldots, 1000$$

The algorithm for estimating the pump failure probability involves the following steps:

1. Initialize the subinterval index $m = 0$.
2. Let $m = m + 1$. Generate one standard uniform random variate u_m, and transform it into the random variate from the corresponding subinterval by $t_m = 0.2(m-1) + u_m$.
3. If $m < 1000$, go to step 2; otherwise, compute the pump failure probability as

$$\hat{p}_f = \frac{1}{1000} \sum_{m=1}^{1000} f_t(t_m)$$

and the associated standard deviation as

$$s_{\hat{p}_f} = \frac{s_m}{\sqrt{1000}}$$

with s_m representing the standard deviation of 1000 computed function values $f_t(t_m)$.

The results from the numerical simulation are

$$\hat{p}_f = 0.14786 \qquad s_{\hat{p}f} = 0.000216$$

The 95 percent confidence interval is (0.14743, 0.14828). The value of $\hat{p}_f$ is extremely close to the exact solution of 0.147856, and only 1000 simulations were used.

6.7.6 Control-variate method

The basic idea behind the *control-variate method* for variance reduction is to take advantage of the available information for the selected variables related to the quantity to be estimated. Referring to Eq. (6.91), the quantity G to be estimated is the expected value of the output of the model $g(\boldsymbol{X})$. The value of G can be estimated directly by those techniques described in Sec. 6.6. However, a reduction in estimation error can be achieved by indirectly estimating the mean of a surrogate model $\hat{g}(\boldsymbol{X}, \zeta)$ as (Ang and Tang, 1984)

$$\hat{g}(\boldsymbol{X}, \zeta) = g(\boldsymbol{X}) - \zeta\{g'(\boldsymbol{X}) - E[g'(\boldsymbol{X})]\} \qquad (6.100)$$

in which $g'(\boldsymbol{X})$ is a control variable with the known expected value $E[g'(\boldsymbol{X})]$, and ζ is a coefficient to be determined in such a way that the variance of $g(\boldsymbol{X}, \zeta)$ is minimized. The control variable $g'(\boldsymbol{X})$ is also a model, which is a function of the same stochastic variables $\boldsymbol{X}$ as in the model $g(\boldsymbol{X})$. It can be shown that $\hat{g}(\boldsymbol{X}, \zeta)$ is an unbiased estimator of the random model output $g(\boldsymbol{X})$, that is, $E[\hat{g}(\boldsymbol{X}, \zeta)] = E[g(\boldsymbol{X})] = G$. The variance of $\hat{g}(\boldsymbol{X}, \zeta)$, for any given ζ, can be obtained as

$$\mathrm{Var}(\hat{g}) = \mathrm{Var}(g) + \zeta^2 \mathrm{Var}(g') - 2\zeta\,\mathrm{Cov}(g, g') \qquad (6.101)$$

The coefficient ζ that minimizes $\mathrm{Var}(\hat{g})$ in Eq. (6.101) is

$$\zeta_* = \frac{\mathrm{Cov}(g, g')}{\mathrm{Var}(g')} \qquad (6.102)$$

and the corresponding variance of $\hat{g}(\boldsymbol{X}, \zeta)$ is

$$\mathrm{Var}(\hat{g}) = \left(1 - \rho_{g,g'}^2\right)\mathrm{Var}(g) \le \mathrm{Var}(g) \qquad (6.103)$$

in which $\rho_{g,g'}$ is the correlation coefficient between the model output $g(\boldsymbol{X})$ and the control variable $g'(\boldsymbol{X})$. Since both model output $g(\boldsymbol{X})$ and the control variable $g'(\boldsymbol{X})$ depend on the same stochastic variables $\boldsymbol{X}$, correlation to a certain degree exists between $g(\boldsymbol{X})$ and $g'(\boldsymbol{X})$. As can be seen from Eq. (6.103), using a control variable $g'(\boldsymbol{X})$ could result in a variance reduction in estimating the expected model output. The degree of variance reduction depends on how large the value of the correlation coefficient is. There exists a tradeoff here. To attain a high variance reduction, a high correlation coefficient is required, which can be achieved by making the control variable $g'(\boldsymbol{X})$ a good approximation to the model $g(\boldsymbol{X})$. However, this could result in a complex control variable for which the expected value may not be derived easily. On the other hand, the use of a simple control variable $g'(\boldsymbol{X})$ that is a poor approximation of $g(\boldsymbol{X})$ would not result in an effective variance reduction in estimation.

The attainment of variance reduction, however, cannot be achieved from total ignorance. Equation (6.103) indicates that variance reduction for estimating G is possible only through the correlation between $g(\boldsymbol{X})$ and $g'(\boldsymbol{X})$. However, the correlation between $g(\boldsymbol{X})$ and $g'(\boldsymbol{X})$ is generally not known in real-life situations. Consequently, a sequence of random variates of $\boldsymbol{X}$ must be produced to compute the corresponding values of the model output $g(\boldsymbol{X})$ and the control variable $g'(\boldsymbol{X})$ to estimate the optimal value of ζ_* by Eq. (6.102). The general algorithm of the control-variate method can be stated as follows.

1. Select a control variable $g'(\boldsymbol{X})$.
2. Generate random variates for $\boldsymbol{X}^{(i)}$ according to their probabilistic laws.
3. Compute the corresponding values of the model $g(\boldsymbol{X}^{(i)})$ and the control variable $g'(\boldsymbol{X}^{(i)})$.
4. Repeat steps 2 and 3 n times.
5. Estimate the value ζ_*, according to Eq. (6.102), by

$$\hat{\zeta}_* = \frac{\sum_{i=1}^n (g^{(i)} - \bar{g})[g'^{(i)} - E(g')]}{n \operatorname{Var}(g')} \tag{6.104}$$

or

$$\hat{\zeta}_* = \frac{\sum_{i=1}^n [g^{(i)} - \bar{g}][g'^{(i)} - E(g')]}{\sum_{i=1}^n [g'^{(i)} - E(g')]^2} \tag{6.105}$$

depending on whether the variance of the control variable $g'(\boldsymbol{X})$ is known or not.

6. Estimate the value of G, according to Eq. (6.100), by

$$\hat{G} = \frac{1}{n} \sum_{i=1}^n (g^{(i)} - \hat{\zeta}_* \, g'^{(i)}) + \hat{\zeta}_* E(g') \tag{6.106}$$

Further improvement in accuracy could be made in step 2 of this above algorithm by using the antithetic-variate approach to generate random variates.

This idea of the control-variate method can be extended to consider a set of J control variates $\boldsymbol{g}'(\boldsymbol{X}) = [g_1'(\boldsymbol{X}), g_2'(\boldsymbol{X}), \ldots, g_J'(\boldsymbol{X})]^t$. Then Eq. (6.100) can be modified as

$$\hat{g}(\boldsymbol{X}, \zeta) = g(\boldsymbol{X}) - \sum_{j=1}^J \zeta_j \{g_j'(\boldsymbol{X}) - E[g_j'(\boldsymbol{X})]\} \tag{6.107}$$

The vector of optimal coefficients $\zeta_* = (\zeta_{*1}, \zeta_{*2}, \ldots, \zeta_{*J})^t$ that minimizes the variance of $\hat{g}(\boldsymbol{X}, \zeta)$ is

$$\zeta_* = \boldsymbol{C}^{-1}\boldsymbol{c} \tag{6.108}$$

in which c is a $J \times 1$ cross-covariance vector between J control variates $g'(X)$ and the model $g(X)$, that is, $c = \{\text{Cov}[g(X), g'_1(X)], \text{Cov}[g(X), g'_2(X)], \ldots, \text{Cov}[g(X), g'_J(X)]\}$, and C is the covariance matrix of the J control variates, that is, $C = [\sigma_{ij}] = [g'_i(X), g'_j(X)]$, for $i, j = 1, 2, \ldots, J$. The corresponding minimum variance of the estimator $\hat{g}(X, \zeta)$ is

$$\text{Var}(\hat{g}) = \text{Var}(g) - c^t C c = \left(1 - \rho^2_{g,g'}\right) \text{Var}(g) \qquad (6.109)$$

in which $\rho_{g,g'}$ is the *multiple correlation coefficient* between $g(X)$ and the vector of control variates $g'(X)$. The squared multiple correlation coefficient is called the *coefficient of determination* and represents the percentage of variation in the model outputs $g(X)$ explained by the J control variates $g'(X)$.

6.8   Resampling Techniques

Note that the Monte Carlo simulation described in preceding sections is conducted under the condition that the probability distribution and the associated population parameters are known for the random variables involved in the system. The observed data are not used directly in the simulation. In many statistical estimation problems, the statistics of interest often are expressed as functions of random observations, that is,

$$\hat{\Theta} = \hat{\Theta}(X_1, X_2, \ldots, X_n) \qquad (6.110)$$

The statistics $\hat{\Theta}$ could be estimators of unknown population parameters of interest. For example, consider that random observations Xs are the annual maximum floods. The statistics $\hat{\Theta}$ could be the distribution of the floods; statistical properties such as mean, standard deviation, and skewness coefficient; the magnitude of the 100-year event; a probability of exceeding the capacity of a hydraulic structure; and so on.

Note that the statistic $\hat{\Theta}$ is a function of the random variables. It is also a random variable, having a PDF, mean, and standard deviation like any other random variable. After a set of n observations $\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$ is available, the numerical value of the statistic $\hat{\Theta}$ can be computed. However, along with the estimation of $\hat{\Theta}$ values, a host of relevant issues can be raised with regard to the accuracy associated with the estimated $\hat{\Theta}$, its bias, its confidence interval, and so on. These issues can be evaluated using the Monte Carlo simulation in which many sequences of random variates of size n are generated from each of which the value of the statistic of interest is computed $\hat{\Theta}$. Then the statistical properties of $\hat{\Theta}$ can be summarized.

Unlike the Monte Carlo simulation approach, resampling techniques are developed that reproduce random data exclusively on the basis of observed data. Tung and Yen (2005, Sec. 6.7) described two resampling techniques, namely, the *jackknife method* and the bootstrap method. A brief description of the latter is given below because the bootstrap method is more versatile and general than the jackknife method.

The *bootstrap technique* was first proposed by Efron (1979a, 1979b) to deal with the variance estimation of sample statistics based on observations. The technique intends to be a more general and versatile procedure for sampling distribution problems without having to rely heavily on the normality condition on which classical statistical inferences are based. In fact, it is not uncommon to observe nonnormal data in hydrosystems engineering problems. Although the bootstrap technique is computationally intensive—a price to pay to break away from dependence on the normality theory—such concerns will be diminished gradually as the calculating power of the computers increases (Diaconis and Efron, 1983). An excellent overall review and summary of bootstrap techniques, variations, and other resampling procedures are given by Efron (1982) and Efron and Tibshirani (1993). In hydrosystems engineering, bootstrap procedures have been applied to assess the uncertainty associated with the distributional parameters in flood frequency analysis (Tung and Mays, 1981), optimal risk-based hydraulic design of bridges (Tung and Mays, 1982), and unit hydrograph derivation (Zhao et al., 1997).

The basic algorithm of the bootstrap technique in estimating the standard deviation associated with any statistic of interest from a set of sample observations involves the following steps:

1. For a set of sample observations of size n, that is, $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, assign a probability mass $1/n$ to each observation according to an empirical probability distribution $\hat{f}$,

$$\hat{f} : P(X = x_i) = 1/n \qquad \text{for } i = 1, 2, \ldots, n \tag{6.111}$$

2. Randomly draw n observations from the original sample set using $\hat{f}$ to form a *bootstrap sample* $\boldsymbol{x}_{\#} = \{x_{1\#}, x_{2\#}, \ldots, x_{n\#}\}$. Note that the bootstrap sample $x_{\#}$ is a subset of the original samples $\boldsymbol{x}$.

3. Calculate the value of the sample statistic $\hat{\Theta}_{\#}$ of interest based on the bootstrap sample $\boldsymbol{x}_{\#}$.

4. Independently repeat steps 2 and 3 a number of times M, obtaining *bootstrap replications* of $\hat{\boldsymbol{\theta}}_{\#} = \{\hat{\theta}_{\#1}, \hat{\theta}_{\#2}, \ldots, \hat{\theta}_{\#M}\}$, and calculate

$$\hat{\sigma}_{\hat{\theta}_{\#}} = \left[\frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_{\#m} - \hat{\theta}_{\#.})^2 \right]^{0.5} \tag{6.112}$$

where $\hat{\theta}_{\#.}$ is the average of the bootstrap replications of $\hat{\Theta}$, that is,

$$\hat{\theta}_{\#.} = \sum_{m=1}^{M} \hat{\theta}_{\#m}/M \tag{6.113}$$

A flowchart for the basic bootstrap algorithm is shown in Fig. 6.15. The bootstrap algorithm described provides more information than just computing the
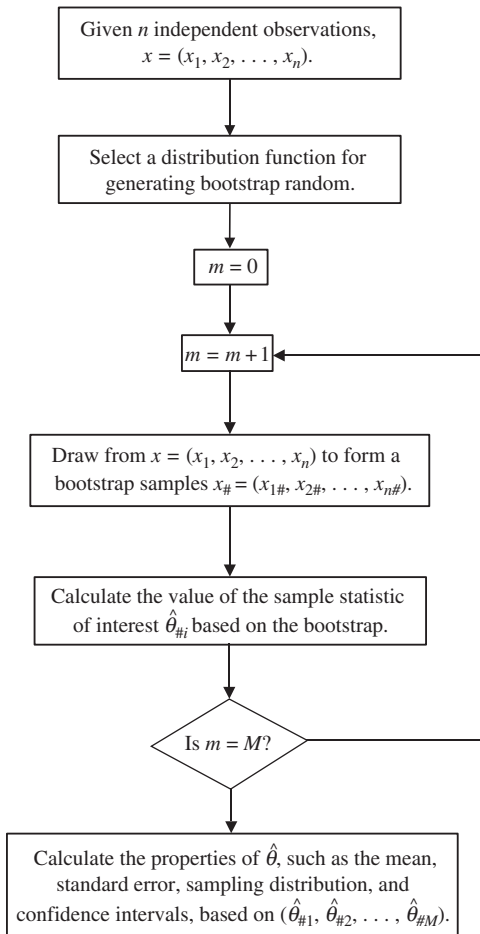
Figure 6.15 Flowchart of basic bootstrap resampling algorithm.

standard deviation of a sample statistic. The histogram constructed on the basis of M bootstrap replications $\hat{\theta}_\# = \{\hat{\theta}_{\#1}, \hat{\theta}_{\#2}, \ldots, \hat{\theta}_{\#M}\}$ gives some ideas about the sampling distribution of the sample statistic $\hat{\Theta}$, such as the failure probability. Furthermore, based on the bootstrap replications $\hat{\theta}_\#$, one can construct confidence intervals for the sample statistic of interest. Similar to Monte Carlo simulation, the accuracy of estimation increases as the number of bootstrap samples gets larger. However, a tradeoff exists between computational cost and the level of accuracy desired. Efron (1982) suggested that $M = 200$ is generally sufficient for estimating the standard errors of the sample statistics. However, to estimate the confidence interval with reasonable accuracy, one would need at least $M = 1000$.

This algorithm is called *nonparametric, unbalanced bootstrapping*. Its parametric version can be made by replacing the nonparametric estimator $\hat{f}$ by a

parametric distribution in which the distribution parameters are estimated by the maximum-likelihood method. More specifically, if one judges that on the basis of the original data set the random observations $x = \{x_1, x_2, \ldots, x_n\}$ are from, say, a lognormal distribution, then the resampling of x's from x using the parametric mechanism would assume that $\hat{f}$ is a lognormal distribution.

Note that the theory of the unbalanced bootstrap algorithm just described only ensures that the expected number to be resampled for each individual observation is equal to the number of bootstrap samples M generated. To improve the estimation accuracy associated with a statistical estimator of interest, Davison et al. (1986) proposed *balanced bootstrap* simulation, in which the number of appearances of each individual observation in the bootstrap data set must be exactly equal to the total number of bootstrap replications generated. This constrained bootstrap simulation has been found, in both theory and practical implementations, to be more efficient than the unbalanced algorithm in that the standard error associated with $\hat{\Theta}$ by the balanced algorithm is smaller. This implies that fewer bootstrap replications are needed by the balanced algorithm than the unbalanced approach to achieve the same accuracy level in estimation. Gleason (1988) discussed several computer algorithms for implementing the balanced bootstrap simulation.

Example 6.15 Based on the annual maximum flood data listed in Table 6.4 for Miller Creek, Los Molinos, California, use the unbalanced bootstrap method to estimate the mean, standard errors, and 95 percent confidence interval associated with the annual probability that the flood magnitude exceeds 20,000 ft^3/s.

Solution In this example, $M = 2000$ bootstrap replications of size $n = 30$ from $\{y_i = \ln(x_i)\}, i = 1, 2, \ldots, 30$, are generated by the unbalanced nonparametric bootstrap procedure. In each replication, the bootstrapped flows are treated as lognormal

TABLE 6.4   Annual Maximum Floods for Mill Creek near Los Molinos, California

Year	Discharge (ft^3/s)	Year	Discharge (ft^3/s)
1929	1,500	1944	3,220
1930	6,000	1945	3,230
1931	1,500	1946	6,180
1932	5,440	1947	4,070
1933	1,080	1948	7,320
1934	2,630	1949	3,870
1935	4,010	1950	4,430
1936	4,380	1951	3,870
1937	3,310	1952	5,280
1938	23,000	1953	7,710
1939	1,260	9154	4,910
1940	11,400	1955	2,480
1941	12,200	1956	9,180
1942	11,000	1957	6,140
1943	6,970	1958	6,880
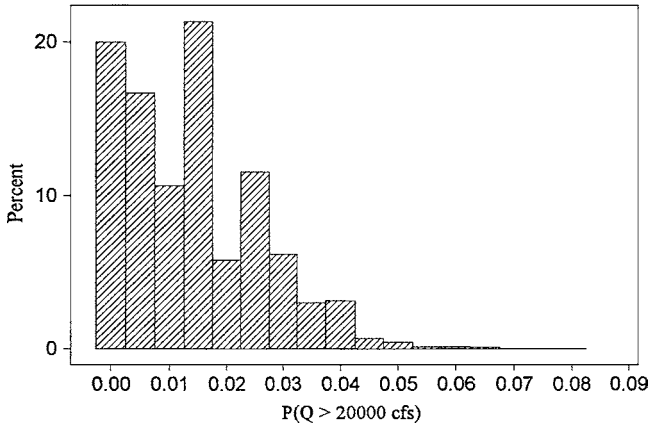
Figure 6.16 Histogram of 2000 bootstrapped replications of $P(Q > 20,000 \text{ ft}^3/\text{s})$ for Example 6.15.

variates based on which the exceedance probability $P(Q > 20,000 \text{ ft}^3/\text{s})$ is computed. The results of the computations are shown below:

Statistic	$P(Q > 20,000 \text{ ft}^3/\text{s})$
Mean	0.0143
Coefficient of variation	0.829
Skewness coefficient	0.900
95 percent confidence interval	(0.000719, 0.03722)

The histogram of bootstrapped replications of $P(Q > 20,000 \text{ ft}^3/\text{s})$ is shown in Fig. 6.16.

Note that the sampling distribution of the exceedance probability $P(Q > 20,000 \text{ ft}^3/\text{s})$ is highly skewed to the right. Because the exceedance probability has to be bounded between 0 and 1, density functions such that the beta distribution may be applicable. The 95 percent confidence interval shown in the table is obtained by truncating 2.5 percent from both ends of the ranked 2000 bootstrapped replications.

Problems

6.1 Generate 100 random numbers from the Weibull distribution with parameters $\alpha = 2.0, \beta = 1.0$, and $\xi = 0$ by the CDF-inverse method. Check the consistency of the sample parameters based on the generated random numbers as compared with the population parameters used.

6.2 Generate 100 random numbers from the Gumbel (extreme type I, max) distribution with parameters $\beta = 3.0$ and $\xi = 1.0$ by the CDF-inverse method. Check the consistency of the sample parameters based on the generated random numbers as compared with the population parameters used.

6.3 Generate 100 random numbers from a triangular distribution with lower bound $a = 2$, mode $m = 5$, and upper bound $b = 10$ by the CDF-inverse method.

Check the consistency of the sample mean, mode, and standard deviation based on the generated random numbers as compared with the population values.

6.4 Prove that $P\{U \leq g(Y)\} = 1/\varepsilon$ for the AR method.

6.5 Consider that the Hazen-William coefficient of a 5-year old, 24-inch cast iron pipe is uncertain, having a triangular distribution with lower bound $a = 115$, mode $m = 120$, and upper bound $b = 125$. Describe an algorithm to generate random numbers by the AR method with $\psi(x) = c$ and $h_x(x) = 1/(b - a)$.

6.6 Refer to Problem 6.5. Determine the efficient constant C and the corresponding acceptance probability for $c = 0.2, 0.3$, and 0.4.

6.7 Refer to Problem 6.5. Develop computer programs to generate 100 random Hazen-Williams coefficients using $c = 0.2, 0.3$, and 0.4. Verify the theoretical acceptance probability for the different c values obtained in Problem 6.6 by your numerical experiment. Discuss the discrepancies, if any exist.

6.8 Generate 100 random variates from

$$f_x(x) = 3x^2 \qquad \text{for } 0 \leq x \leq 1$$

by the AR algorithm delineated in Example 6.3 with $c = 3, a = 0$, and $b = 1$. Also evaluate the theoretical acceptance probability for each random variate to be generated (after Rubinstein, 1981).

6.9 Generate 100 random variates from

$$f_x(x) = \frac{2}{\pi r^2} \sqrt{r^2 - x^2} \qquad \text{for } -r \leq x \leq r$$

by the AR algorithm delineated in Example 6.3 with $c = 2/\pi r$. Also evaluate its theoretical acceptance probability for each random variate to be generated (adopted from Rubinstein, 1981).

6.10 Generate 100 random variates from

$$f_x(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \qquad \text{for } 0 < \alpha < 1; x \geq 0$$

by the general AR algorithm with

$$\psi(x) = Ch_x(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} \qquad \text{for } 0 \leq x \leq 1$$

$$= \frac{e^{-x}}{\Gamma(\alpha)} \qquad \text{for } x > 1$$

$$h_x(x) = \frac{x^{\alpha-1}}{(1/\alpha) + (1/e)} \qquad \text{for } 0 \leq x \leq 1$$

$$= \frac{e^{-x}}{(1/\alpha) + (1/e)} \qquad \text{for } x > 1$$

Also evaluate its theoretical acceptance probability for each random variate to be generated (after Rubinstein, 1981).

6.11   Develop an algorithm to generate random variable $Y = \max\{X_1, X_2, \ldots, X_n\}$, where X_i are independent and identically distributed normal random variables with means μ_x and standard deviations σ_x.

6.12   Develop an algorithm to generate random variable $Y = \min\{X_1, X_2, \ldots, X_n\}$, where X_i are independent and identically distributed lognormal random variables with means μ_x and standard deviations σ_x.

6.13   Based on the algorithm developed in Problem 6.11, estimate the mean, standard deviation, and the magnitude of the 100-year event for a 10-year maximum rainfall ($n = 10$) in which the population for the annual rainfall is normal with a mean of 3 in/h and standard deviation of 0.5 in/h.

6.14   Based on the algorithm developed in Problem 6.12, estimate the mean, standard deviation, and the magnitude of 100-year event for a 10-year minimum water supply ($n = 10$) in which the population for annual water supply is lognormal with mean of 30,000 acre-feet (AF) and standard deviation of 10,000 AF.

6.15   Refer to the strip-mining excavation problem in Example 6.4. Suppose that a decision is made to start the excavation on the fiftieth day ($t = 50$ days). Using the CDF-inverse method, determine the probability that the excavation operation poses no safety threat on the embankment stability. That is, determine the probability that the groundwater drawdown at the excavation point reaches half the original aquifer table depth.

6.16   Resolve Problem 6.15 using the square-root algorithm.

6.17   Resolve Problem 6.15 using the spectral decomposition algorithm.

6.18   Resolve Problem 6.15 assuming that the conductivity and storage coefficient are correlated lognormal random variables. Compare the simulated result with the exact solution.

6.19   Assume that all stochastic model parameters are normal random variables. Develop a Monte Carlo simulation algorithm to solve Problem 4.24, and compare the simulation results with those obtained by the MFOSM and AFOSM reliability methods.

6.20   Assume that all stochastic model parameters are normal random variables. Develop a Monte Carlo simulation algorithm to solve Problem 4.26, and compare the simulation results with those obtained by the MFOSM and AFOSM reliability methods.

6.21   Refer to Problem 4.24, and use the distribution functions specified. Incorporate the normal transform given in Table 4.5 into the Monte Carlo simulation procedure developed in Problem 6.19 to estimate the probability and compare the results with those obtained in Problems 4.24 and 4.34.

6.22   Repeat Problem 6.21 for Problem 4.26 and compare the results with those obtained in Problems 4.26 and 4.36.

6.23 Prove Eq. (6.52)

6.24 Prove Eq. (6.54)

6.25 Show that $\text{Var}(\hat{G})$ by Eq. (6.63) is smaller than that by Eq. (6.52).

6.26 Use directional simulation to solve Problem 6.15, and compare the results with the exact solution and those obtained in Problems 6.15 to 6.17.

6.27 Use directional simulation to solve Problem 6.19, assuming that all stochastic variables are multivariate normal variables. Compare the results with those obtained in Problem 6.19.

6.28 Use directional simulation to solve Problem 6.20, assuming that all stochastic variables are multivariate normal variables. Compare the results with those obtained in Problem 6.20.

6.29 Repeat Example 6.6 using the importance sampling technique with $n = 2000$. The PDF selected has a form of the standard exponential function, that is,

$$f_x(x) = ae^{-x} \qquad \text{for } x \geq 0$$

where a = constant. Compare the results with those obtained in Examples 6.6., 6.7, and 6.8.

6.30 Using the concept of importance sampling, choose $h_x(x) = e^{-ax}$ and estimate the integral

$$G = \int_0^\pi \frac{dx}{x^2 + \cos^2 x}$$

Determine the value of a that minimizes the variance of the integral (after Gould and Tobochnik, 1988).

6.31 Show that $\text{Cov}(U, 1 - U) = -1/12$ in which $U \sim U(0, 1)$.

6.32 Referring to the pump performance in Example 7.6, estimate the failure probability using the antithetic-variates technique along with the sample-mean Monte Carlo algorithm with $n = 1000$. The PDF selected is a standard exponential function, that is,

$$f_x(x) = e^{-x} \qquad \text{for } x \geq 0$$

Also compare the results with those obtained in Examples 6.6, 6.7, 6.8, and 6.11.

6.33 Show that $\text{Var}(\hat{G})$ associated with $\hat{G}$ by Eq. (6.89) is

$$\text{Var}(\hat{G}) = \sum_{m=1}^{M} \frac{P_m^2 \sigma_m^2}{n_m}$$

and derive the corresponding value associated with the optimal sample size allocation n_m^*.

6.34 Refer to the strip mine in Example 6.4. Use the antithetic-variate Monte Carlo technique with $n = 400$ to estimate the first three product-moments of drawdown

recess time corresponding to $s/h_o = 0.5$. Assume that the permeability K_h is the only random variable having a lognormal distribution with the mean $\mu_{kh} = 0.1$ m/day and coefficient of variation $\Omega_{kh} = 10$ percent.

6.35 Refer to the strip mine in Example 6.4. Suppose that engineers are also considering the possibility of starting excavation earlier. Evaluate the difference in expected waiting time between the two options, that is, $s/h_o = 0.5$ and 0.6, by correlated-sampling Monte Carlo simulation with $n = 400$. Assume that the only random variable is the permeability K_h, having a lognormal distribution with the mean 0.1 m/day and coefficient of variation of 0.1.

6.36 Repeat Problem 6.34 using the systematic sampling technique to estimate the first three product-moments of drawdown recess time corresponding to $s/h_o = 0.5$.

6.37 Repeat Problem 6.36 using the LHS technique.

6.38 Resolve Problem 6.15 by incorporating the antithetic-variates method.

6.39 Referring to Problem 6.15, use the correlated-sampling method to determine the difference in probabilities of a safe excavation for $t = 30$ days and $t = 50$ days.

6.40 Resolve Problem 6.15 using the Latin hypercube sampling method.

6.41 Refer to the annual maximum flood data in Table 6.4. Assuming that the flood data follow a lognormal distribution, use the nonparametric unbalanced bootstrap algorithm to estimate $\Theta = P(\text{flood peak} \geq 15{,}000 \text{ ft}^3/\text{s})$ and its associated error. Furthermore, based on the 1000 bootstrap samples generated, assess the probability distribution and 90 percent confidence interval for $\Theta = P$ (flood peak $\geq 15{,}000 \text{ ft}^3/\text{s})$.

6.42 Solve Problem 6.41 using the parametric unbalanced bootstrap algorithm. Compare these results with those obtained from Problem 6.42.

References

Abramowitz, M., and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, Dover Publications, New York.

Ang, A. H.-S., and Tang, W. H. (1984). *Probability Concepts in Engineering Planning and Design, Vol. II: Decision, Risk, and Reliability*, John Wiley and Sons, New York.

Atkinson, A. C. (1979). An easily programmed algorithm for generating gamma random variables, *Journal of Royal Statistical Society* A140:232–234.

Beck, M. B. (1985). Water quality management: A review of the development and application of mathematical models, in *Lecture Notes in Engineering 11*, ed. by C. A. Brebbia and S. A. Orszag, Springer-Verlag, New York.

Borgman, L. E., and Faucette, R. C. (1993). Multidimensional simulation of Gaussian vector random functions in frequency domain, in *Computational Stochastic Mechanics*, ed. by H. D. Cheng and C. Y. Yang. Chapter 3, p. 51–74, Computational Mechanics Publications, Elsevier Applied Science Publishing, Barkingham, UK.

Bjerager, P. (1988). Probability integration by directional simulation, *Journal of Engineering Mechanics*, ASCE, 114(8):1285–1301.

Brown, L. C., and Barnwell, T. O., Jr. (1987). The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual, Report EPA/600/3-87/007, U.S. Environmental Protection Agency, Athens, GA.

Box, G. E. P., and Muller, M. E. (1958). A note on generation of random normal deviates, *Annals of Mathematical Statistics* 29:610–611.

Chang, C. H., Yang, J. C., and Tung, Y. K. (1993). Sensitivity and uncertainty analyses of a sediment transport model: A global approach, *Journal of Stochastic Hydrology and Hydraulics* 7(4):299–314.

Chang, C. H., Tung, Y. K., and Yang, J. C. (1994). Monte Carlo simulation for correlated variables with marginal distributions, *Journal of Hydraulic Engineering*, ASCE, 120(2):313–331.

Chen, X. Y., and Tung, Y. K. (2003). Investigation of polynomial normal transformation, *Journal of Structural Safety*, 25:423–445.

Cheng, S.-T., Yen, B. C., and Tang, W. H. (1982). Overtopping risk for an existing dam, Hydraulic Engineering Series No. 37, Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.

Chilès, J.-P., and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

Christiaens, K., and Feyen, J. (2002). Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model, *Water Resources Research* 38(9):1169.

Cochran, W. (1966). *Sampling Techniques*, 2nd ed., John Wiley and Sons, New York.

Dagpunar, J. (1988). *Principles of Random Variates Generation*, Oxford University Press, New York.

Davison, A. C., Hinkley, D. V., and Schechtman, E. (1986). Efficient bootstrap simulation, *Biometrika* 73(3):555–566.

Der Kiureghian, A., and Liu, P. L. (1985). Structural reliability under incomplete probability information, *Journal of Engineering Mechanics*, ASCE. 112(1):85–104.

Diaconis, P., and Efron, B. (1983). Computer-intensive methods in statistics, *Scientific American Magazine*, May, 116–131.

Efron, B. (1979a). Bootstrap methods: Another look at the jackknife, *Annals of Statistics* 3:1189–1242.

Efron, B. (1979b). Computers and theory of statistics: Thinking the unthinkable, *SIAM Reviews* 21:460–480.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. CBMS 38, SIAM-NSF, Philadelphia, PA.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapmann and Hall, New York.

Gleason, J. R. (1988). Algorithms for balanced bootstrap simulations, *The American Statistician* 42(4):263–266.

Golub, G. H., and Van Loan, C. F. (1989). *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, MD.

Gould, H., and Tobochnik, J. (1988). *An Introduction to Computer Simulation Methods: Applications to Physical Systems*, Part 2, Addison-Wesley, Reading, MA.

Hammersley, J. M., and Morton, K. W. (1956). A new Monte Carlo technique antithetic-variates, *Proceedings of Cambridge Physics Society* 52:449–474.

Hull, T. E., and Dobell, A. R. (1964). Mixed congruential random number generators for binary machines, *Journal of the Association of Computing Machinery*, 11:31–40.

Iman, R. L., and Helton, J. C. (1988). An investigation of uncertainty and sensitivity analysis techniques for computer models, *Risk Analysis* 8(1):71–90.

IMSL (International Mathematical and Statistical Library), IMSL, Inc., Houston, TX, 1980.

Jaffe, P. R., and Ferrara, R. A. (1984). Modeling sediment and water column interactions for hydrophobic pollutants, parameter discrimination and model response to input uncertainty, *Water Research* 18(9):1169–1174.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*, John Wiley & Sons, New York.

Knuth, D. E. (1981). *The Art of Computer Programming: Seminumerical Algorithms*, Vol.2, 2nd ed., Addison-Wesley, Reading, MA.

Laurenson, E. M., and Mein, R. G. (1985). RORB, version 3 Runoff routing program user manual, Department of Civil Engineering, Monash University, Clayton, Victoria, Australia.

Law, A. M., and Kelton, W. D. (1991). *Simulation Modeling and Analysis*, McGraw-Hill, New York.

Lehmer, D. H. (1951). Mathematical methods in large-scale computing units, *Annals Computation Lab.* Harvard University Press, Cambridge, MA, 26:141–146.

Li, S. T., and Hammond, J. L. (1975). Generation of psuedorandom numbers with specified univariate distributions and covariance matrix, *IEEE Transcation on Systems, Man, and Cybernatics*, 5:557–561.

Lu, Z., and Tung, Y. K. (2003). Effects of parameter uncertainties on forecast accuracy of Xinan-jiang model, in *Proceedings*, 1st International Yellow River Forum on River Basin Management, Zhengzhou, China, 12–15 May.

MacLaren, M. D., and Marsaglia, G. (1965). Uniform random number generators, *Journal of the Association of Computing Machinery* 12:83–89.

Manache, G. (2001). Sensitivity of a continuous water-quality simulation model to uncertain model-input parameters, Ph.D. thesis, Chair of Hydrology and Hydraulics, Vrije Universiteit Brussel, Brussels, Belgium.

Manache, G., and Melching, C. S. (2004). Sensitivity analysis of a water-quality model using Latin hypercube sampling, *Journal of Water Resources Planning and Management*, ASCE, 130(3):232–242.

Marsagalia, G., and Bray, T. A. (1964). A convenient method for generating normal variables, *SIAM Review* 6:260–264.

Marshall, A. W. (1956). The use of multistage sampling schemes in Monte Carlo computations, in *Symposium on Monte Carlo Methods*, ed. by M. A. Meyer, John Wiley and Sons, New York.

McGrath, E. I. (1970). *Fundamentals of Operations Research*, West Coast University Press, San Francisco.

McKay, M. D. (1988). Sensitivity and uncertainty analysis using a statistical sample of input values, in *Uncertainty Analysis*, ed. by Y. Ronen, CRC Press, Boca Raton, FL.

Melching, C. S. (1992). An improved first-order reliability approach for assessing uncertainties in hydrologic modeling, *Journal of Hydrology* 132:157–177.

Melching, C. S. (1995). Reliability estimation, in *Computer Models of Watershed Hydrology*, ed. by V. P. Singh, Water Resources Publications, Littleton, CO, pp. 69–118.

Melching, C. S., and Bauwens, W. (2001). Uncertainty in coupled non-point-source and stream water-quality models, *Journal of Water Resources Planning and Management*, ASCE, 127 (6):403–413.

Nataf, A. (1962). Détermination des distributions de probabilités dont les marges sont données, *Computes Rendus de l' Academie des Sciences*, Paris, 255:42–43.

Nguyen, V. U., and Chowdhury, R. N. (1985). Simulation for risk analysis with correlated variables, *Geotechnique* 35(1):47–58.

Nguyen, V. U., and Raudkivi, A. J. (1983). Analytical solution for transient two-dimensional unconfined groundwater flow, *Hydrological Sciences Journal* 28(2):209–219.

Olmstead, P. S. (1946). Distribution of sample arrangements for runs up and down, *Annals of Mathematical Statistics* 17:24–33.

Parrish, R. S. (1990). Generating random deviates from multivariate Pearson distributions, *Computational Statistics and Data Analysis* 9:283–295.

Pebesma, E. J., and Heuvelink, G. B. M. (1999). Latin hypercube sampling of Gaussian random fields, *Technometrics* 41(4):303–312.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989). *Numerical Recipes in Pascal: The Art of Scientific Computing*, Cambridge University Press, New York.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University Press, New York.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, New York.

Ronning, G. (1977). A simple scheme for generating multivariate gamma distributions with non-negative covariance matrix, *Technometrics* 19(2):179–183.

Rosenblatt, M. (1952). Remarks on multivariate transformation, *Annals of Mathematical Statistics* 23:470–472.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, John Wiley and Sons, New York.

Sohrabi, T. M., Shirmohammadi, A., Chu, T. W., Montas, H., and Nejadhashemi, A. P. (2003). Uncertainty analysis of hydrologic and water quality predictions for a small watershed using SWAT2000, *Environmental Forensics* 4(4):229–238.

Tung, Y. K., and Mays, L. W. (1981). Reducing hydrologic parameter uncertainty, *Journal of the Water Resources Planning and Management Division*, ASCE, 107(WR1):245–262.

Tung, Y. K., and Mays, L. W. (1982). Optimal risk-based hydraulic design of bridges, *Journal of the Water Resources Planning and Management Division*, ASCE, 108(WR2):191–203.

Tung, Y. K., and Yen, B. C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, McGraw-Hill, New York.

U. S. Army Corps of Engineers (1990). *HEC-1 Flood Hydrograph Package*, Hydrologic Engineering Center, Davis, CA.

Vale, C. D., and Maurelli, V. A. (1983) Simulating multivariate nonnormal distributions, *Psychometrika* 48(3):465–471.

von Neumann, J. (1951). Various techniques used in connection with random digits, U.S. National Bureau of Standards, *Applied Mathematics Series* 12:36–38.

Wang, Y., and Tung, Y. K. (2005) Stochastic generation of GIUH-based flow hydrograph, in *Environmental Hydraulics and Sustainable Water Management*, Vol. 2, ed. by J. H. W. Lee and K. M. Lam, pp. 1917–1924; Proceedings of the 4th International Symposium on Environmental Hydraulics and 14th Congress of the Asia and Pacific Division of the IAHR 15-18, December 2004, Hong Kong, A. A. Balkema, London.

Yeh, K. C., and Tung, Y. K. (1993). Uncertainty and sensitivity of a pit migration model, *Journal of Hydraulic Engineering*, ASCE, 119(2):262–281.

Young, D. M., and Gregory, R. T. (1973). *A Survey of Numerical Mathematics*, Vol. II, Dover Publications, New York.

Yu, P. S., Yang, T. C., and Chen, S. J. (2001). Comparison of uncertainty analysis methods for a distributed rainfall-runoff model, *Journal of Hydrology* 244:43–59.

Zhao, B., Tung, Y. K., Yeh, K. C., and Yang, J. C. (1997a). Storm resampling for uncertainty analysis of a multiple-storm unit hydrograph, *Journal of Hydrology* 194:366–384.

Zhao, B., Tung, Y. K., Yeh, K. C., and Yang, J. C. (1997b), Reliability analysis of hydraulic structures considering unit hydrograph uncertainty, *Journal of Stochastic Hydrology and Hydraulics* 11(1):33–50.

This page intentionally left blank

Reliability of Systems

7.1 Introduction

Most systems involve many subsystems and components whose performances affect the performance of the system as a whole. The reliability of the entire system is affected not only by the reliability of individual subsystems and components but also by the interactions and configurations of the subsystems and components. Many engineering systems involve multiple failure paths or modes; that is, there are several potential paths and modes of failure in which the occurrence, either individually or in combination, would constitute system failure. As mentioned in Sec. 1.3, engineering system failure can be *structural failure* such that the system can no longer function, or it can be *performance failure*, for which the objective is not achieved but the functioning of the system is not damaged. In terms of their functioning configuration and layout pattern, engineering systems can be classified into series systems or parallel systems, as shown schematically in Figs. 7.1 and 7.2, respectively.

A formal quantitative reliability analysis for an engineering system involves a number of procedures, as illustrated in Fig. 7.3. First, the system domain is defined, the type of the system is identified, and the conditions involved in the problem are defined. Second, the kind of failure is identified and defined. Third, factors that contribute to the working and failure of the system are identified. Fourth, uncertainty analysis for each of the contributing component factors or subsystems is performed. Chapters 4 and 5 of Tung and Yen (2005) and Chap. 6 of this book describe some of the methods that can be used for this step. Fifth, based on the characteristics of the system and the nature of the failure, a logic tree is selected to relate the failure modes and paths involving different components or subsystems. Fault trees, event trees, and decision trees are the logic trees often used. Sixth, identify and select an appropriate method or methods that can combine the components or subsystems following the logic of the tree to facilitate computation of system reliability. Some of the computational
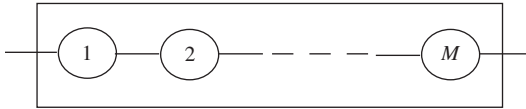
Figure 7.1 Schematic diagram of a series system.

methods are described in Chaps. 4, 5, and 6. Seventh, perform the computation following the methods selected in the sixth step to determine the system failure probability and reliability. Eighth, if the cost of the damage associated with the system failure is desired and the failure damage cost function is known or can be determined, it can be combined with the system failure probability function determined in step 7 to yield the expected damage cost.

The different contributing factors or parameters may have different measurement units. In quantitative combination for reliability analysis, these statistical parameters or factors are normalized through their respective mean or standard deviation to become nondimensional, such as coefficients of variation, to facilitate uncertainty combination.

Real-life hydrosystems engineering infrastructural systems often are so large and complex that teams of experts of different disciplines are required to conduct the reliability analysis and computation. Logic trees are tools that permit division of team work and subsequent integration for the system result. Information on the logic trees and types of systems related to steps 5 and 6 are discussed in this chapter.

7.2 General View of System Reliability Computation

As mentioned previously, the reliability of a system depends on the component reliabilities and interactions and configurations of components. Consequently, computation of system reliability requires knowing what constitutes the system being in a failed or satisfactory state. Such knowledge is essential for system classification and dictates the methodology to be used for system reliability determination.
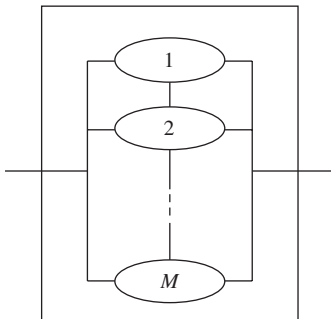
Figure 7.2 Schematic diagram of a parallel system.
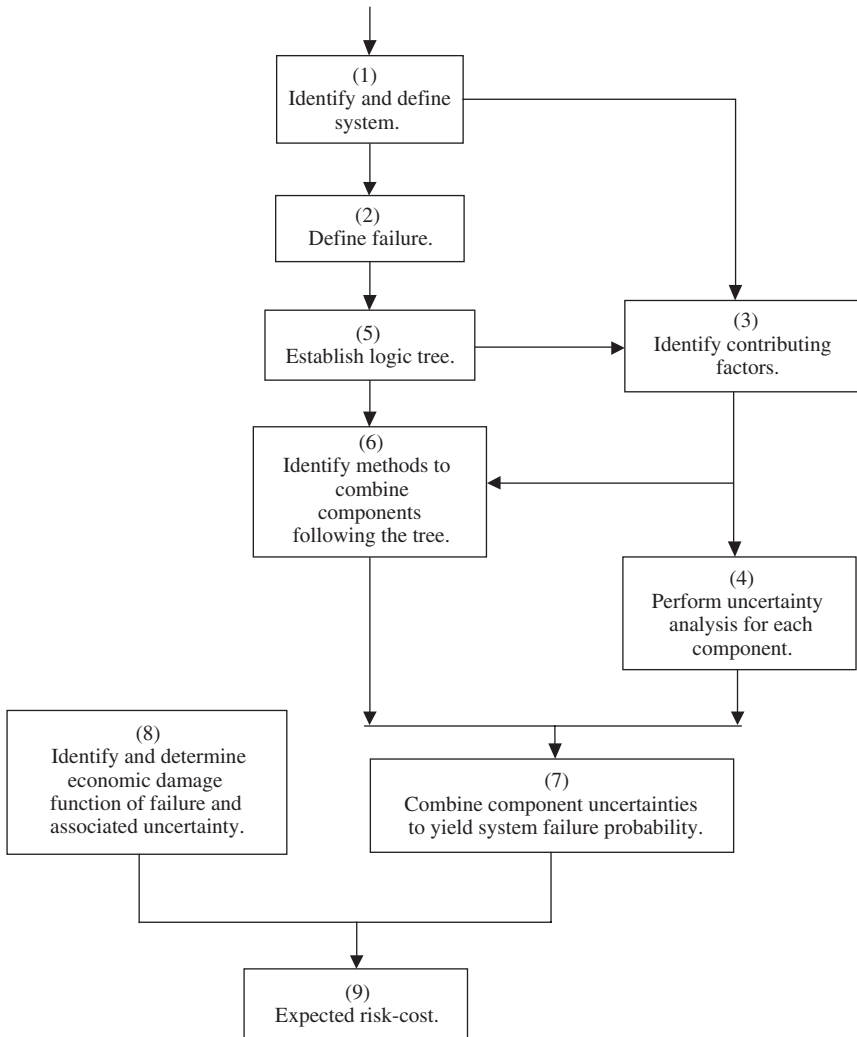
Figure 7.3 Procedure for infrastructural engineering system reliability.

7.2.1 Classification of systems

From the reliability computation viewpoint, classification of the system depends primarily on how system performance is affected by its components or modes of operation. A multiple-component system called a *series system* (see Fig. 7.1) requires that all its components perform satisfactorily to allow satisfactory performance of the entire system. Similarly, for a single-component system involving several modes of operation, it is also viewed as a series system if satisfactory performance of the system requires satisfactory performance of all its different modes of operation.

A second basic type of system is called a *parallel system* (see Fig. 7.2). A parallel system is characterized by the property that the system would serve its intended purpose satisfactorily as long as at least one of its components or modes of operation performs satisfactorily.

For most real-life problems, system configurations are complex, in which the components are arranged as a mixture of series and parallel subsystems or in the form of a loop. In dealing with the reliability analysis of a complex system, the general approach is to reduce the system configuration, based on the arrangement of its components or modes of operation, to a simpler situation for which the reliability analysis can be performed easily. However, this goal may not always be achievable, in which case a special procedure would have to be devised.

7.2.2 Basic probability rules for system reliability

The solution approaches to system reliability problems can be classified broadly into *failure-modes approach* and *survival-modes approach* (Bennett and Ang, 1983). The failure-modes approach is based on identification of all possible failure modes for the system, whereas the survival-modes approach is based on the all possible modes of operation under which the system will be operational. The two approaches are complementary. Depending on the operational characteristics and configuration of the system, a proper choice of one of the two approaches often can lead to significant reduction in efforts needed for the reliability computation.

Consider that a system has M components or modes of operation. Let event F_m indicate that the mth component or mode of operation is in the failure state. If the system is a series system, the failure probability of the system is the probability that *at least* one of the M components or modes of operation fails, namely,

$$p_{f,\text{sys}} = P(F_1 \cup F_2 \cup \cdots \cup F_M) = P\left(\bigcup_{m=1}^{M} F_m\right) \tag{7.1}$$

in which $p_{f,\text{sys}}$ is the failure probability of the system. On the other hand, the system reliability $p_{s,\text{sys}}$ is the probability that all its components or modes of operation perform satisfactorily, that is,

$$p_{s,\text{sys}} = P(F_1' \cap F_2' \cap \cdots \cap F_M') = P\left(\bigcap_{m=1}^{M} F_m'\right) \tag{7.2}$$

in which F_m' is the complementary event of F_m indicating that the mth component or mode of operation does not fail.

In general, failure events associated with system components or modes of operation are not mutually exclusive. Therefore, referring to Eq. (2.4), the failure

probability for a series system can be computed as

$$p_{f,\text{sys}} = P\left(\bigcup_{m=1}^{M} F_m\right) = \sum_{m=1}^{M} P(F_m) - \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} P(F_i, F_j)$$

$$+ \sum_{i<} \sum_{j<} \sum_{k} P(F_i, F_j, F_k) - \cdots + (-1)^M P(F_1, F_2, \ldots, F_M) \quad (7.3)$$

According to Eq. (2.7), the reliability for a series system is

$$p_{s,\text{sys}} = P(F'_1) \times P(F'_2 \,|\, F'_1) \times P(F'_3 \,|\, F'_1, F'_2) \times \cdots \times P(F'_M \,|\, F'_1, F'_2, \ldots, F'_{M-1})$$

$$(7.4)$$

In the case that failure events F'_m's are mutually exclusive or the probability of joint occurrence of multiple failures is negligible, the failure probability of a series system can be easily obtained as

$$p_{f,\text{sys}} = \sum_{m=1}^{M} P(F_m) \qquad (7.5a)$$

with the corresponding system reliability

$$p_{s,\text{sys}} = 1 - p_{f,\text{sys}} = 1 - \sum_{m=1}^{M} P(F_m) \qquad (7.5b)$$

Under the condition that all failure events are statistically independent, the reliability of a series system can be computed easily as

$$p_{s,\text{sys}} = \prod_{m=1}^{M} P(F'_m) = \prod_{m=1}^{M} [1 - P(F_m)] \qquad (7.6a)$$

with the corresponding system failure probability

$$p_{f,\text{sys}} = 1 - \prod_{m=1}^{M} [1 - P(F_m)] \qquad (7.6b)$$

The component probability $P(F_m)$ can be determined by methods described in Chaps. 4, 5, and 6 of this book.

Example 7.1 Consider a series system consisting of M independent components, each with an identical component reliability of p_s. The system reliability, according to Eq. (7.6a), is

$$p_{s,\text{sys}} = p_s^M$$

Figure 7.4 shows the relationship among the reliability of a series system, component reliability, and the number of components. As can be seen, the reliability of a series system decreases as the number of components increases.
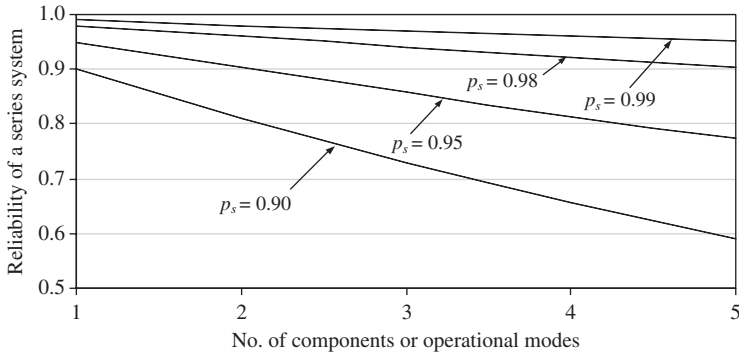
Figure 7.4 Relationship among reliability of a series system, component reliability, and the number of components.

In the case of a parallel system, the system would fail if all its components or modes of operation failed. Hence the failure probability for a parallel system is

$$p_{f,\text{sys}} = P(F_1 \cap F_2 \cap \cdots \cap F_M) = P\left(\mathop{\cap}_{m=1}^{M} F_m\right) \tag{7.7}$$

The reliability of a parallel system, on the other hand, is the probability that at least one of its component or modes of operation is functioning, that is,

$$p_{s,\text{sys}} = P(F_1' \cup F_2' \cup \cdots \cup F_M') = P\left(\mathop{\cup}_{m=1}^{M} F_m'\right) \tag{7.8}$$

Hence, under the condition of independence for all failure events, the failure probability of a parallel system simply is

$$p_{f,\text{sys}} = \prod_{m=1}^{M} P(F_m) \tag{7.9a}$$

with the corresponding system reliability being

$$p_{s,\text{sys}} = 1 - \prod_{m=1}^{M} P(F_m) = 1 - \prod_{m=1}^{M} [1 - P(F_m')] \tag{7.9b}$$

Example 7.2 Consider a parallel system consisting of M independent components, each with an identical component reliability of p_s. The system reliability, according to Eq. (7.9b), is

$$p_{s,\text{sys}} = 1 - (1 - p_s)^M$$

Figure 7.5 shows the relationship among reliability of a parallel system, component reliability, and the number of components. The figure indicates that the reliability of a parallel system increases as the number of components increases.
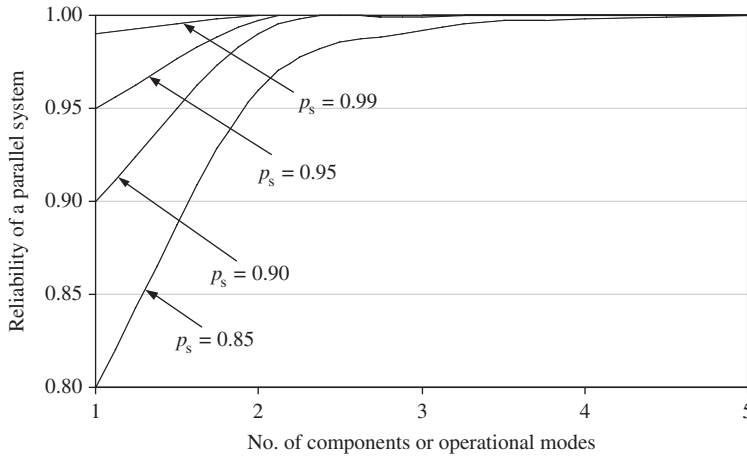
Figure 7.5 Relationship among reliability of a parallel system, component reliability, and the number of components.

For mutually exclusive failure events, reliability of a parallel system can be computed as

$$p_{s,\text{sys}} = \sum_{m=1}^{M} P(F'_m) = M - \sum_{m=1}^{M} P(F_m) \tag{7.10a}$$

with the corresponding system failure probability

$$p_{f,\text{sys}} = 1 + \sum_{m=1}^{M} P(F_m) - M \tag{7.10b}$$

Unfortunately, for a real-life system involving multiple components or modes of operation, the corresponding failure events are neither independent nor mutually exclusive. Consequently, the computation of exact values of system reliability and failure probability would not be a straightforward task. In practical engineering applications, bounds on system reliability are computed based on simpler expressions with less computational effort. As will be seen in the next subsection, to achieve tighter bounds on system reliability or failure probability, a more elaborate computation will be required. Of course, the required precision for the computed system reliability is largely dependent on the importance of the satisfactory performance of the system under consideration.

7.2.3 Bounds for system reliability

Despite the system under consideration being a series or parallel system, the evaluation of system reliability or failure probability involves probabilities of union or intersection of multiple events. Without losing generality, the derivation of the bounds for $P(A_1 \cup A_2 \cup \cdots \cup A_M)$ and $P(A_1 \cap A_2 \cap \cdots \cap A_M)$ are given below.

For example, consider $P(A_1 \cup A_2 \cup \cdots \cup A_M)$. When $A_m = F_m$, the probability is for the failure probability of a series system, whereas when $A_m = F'_m$, the probability is for the reliability of a parallel system. The bounds of system failure probability, that is, can be defined as follows:

$$\underline{p}_{f,\text{sys}} \le p_{f,\text{sys}} \le \overline{P}_{f,\text{sys}} \tag{7.11a}$$

with $\underline{p}_{f,\text{sys}}$ and $\overline{p}_{f,\text{sys}}$ being the lower and upper bounds of system failure probability, respectively. The corresponding bounds for system reliability can be obtained as

$$1 - \overline{P}_{f,\text{sys}} = \underline{p}_{s,\text{sys}} \le p_{s,\text{sys}} \le \overline{P}_{s,\text{sys}} = 1 - \underline{p}_{f,\text{sys}} \tag{7.11b}$$

Similarly, after the bounds on system reliability are obtained, the bounds on system failure reliability can be computed easily.

First-order bounds. These bounds also are called *unimodal bounds* (Ang and Tang, 1984, p. 450). They can be derived as follows. Referring to Eq. (2.7), the probability of the joint occurrence of several events can be computed as

$$P\left(\bigcap_{m=1}^{M} A_m \right) = P(A_1) \times P(A_2 \mid A_1) \times \cdots \times P(A_M \mid A_{M-1}, A_{M-2}, \dots, A_2, A_1) \tag{7.12}$$

Under the condition that all events A_j and A_m are positively correlated, the following inequality relationship holds:

$$P(A_m \mid A_j) \ge P(A_m)$$

Hence

$$P(A_j, A_m) = P(A_m \mid A_j) P(A_j) \ge P(A_m) P(A_j)$$

This can be extended to a multiple-event case as

$$P\left(\bigcap_{m=1}^{M} A_m \right) \ge \prod_{m=1}^{M} P(A_m) \tag{7.13}$$

As can be seen, the lower bound of the probability of an intersection is when all events are as if they are independent. Furthermore, it is also true that

$$\bigcap_{m=1}^{M} A_m \subset A_j \qquad \text{for any } j = 1, 2, \dots, M$$

Therefore,

$$\left(\bigcap_{m=1}^{M} A_m \right) \subset \min\{A_1, A_2, \dots, A_M\}$$

Consequently,

$$P\left(\bigcap_{m=1}^{M} A_m \right) \le \min\{P(A_1), P(A_2), \dots, P(A_M)\} \tag{7.14}$$

Based on Eqs. (7.13) and (7.14), the bounds on probability of joint occurrence of several positively correlated events are

$$\prod_{m=1}^{M} P(A_m) \leq P\left(\bigcap_{m=1}^{M} A_m\right) \leq \min_{m=1,2,\ldots,M}\{P(A_m)\} \tag{7.15}$$

Example 7.3 Consider three standardized normal random variables Z_1, Z_2, and Z_3 with the following correlation matrix:

$$\mathbf{R}_z = \begin{bmatrix} 1.000 & 0.841 & 0.014 \\ 0.841 & 1.000 & 0.536 \\ 0.014 & 0.536 & 1.000 \end{bmatrix}$$

Compute the first-order bounds for $P\{(Z_1 \leq -2.71) \cup (Z_2 \leq -2.88) \cup (Z_3 \leq -3.44)\}$.

Solution The three events corresponding to the preceding trivariate normal probability are

$$A_1 = \{Z_1 \leq -2.71\} \qquad A_2 = \{Z_2 \leq -2.88\} \qquad A_3 = \{Z_3 \leq -3.44\}$$

Since,

$$P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1' \cap A_2' \cap A_3')$$

the first-order bounds for $P(A_1 \cup A_2 \cup A_3)$ can be obtained from those of $P(A_1' \cap A_2' \cap A_3')$.

To derive the first-order bounds for $P(A_1' \cap A_2' \cap A_3')$, individual probabilities are required, which can be obtained from Table 2.2 or Eq. (2.63) as

$$P(A_1') = P(Z_1 \geq -2.71) = 0.99664$$
$$P(A_2') = P(Z_2 \geq -2.88) = 0.99801$$
$$P(A_3') = P(Z_3 \geq -3.44) = 0.99971$$

Furthermore, because all A_m''s are positively correlated, all A_m's are also positively correlated. According to Eq. (7.15), the first-order bounds for $P(A_1' \cap A_2' \cap A_3')$ is

$$(0.99664)(0.99801)(0.99971) \leq P\left(\bigcap_{m=1}^{3} A_m'\right) \leq \min\{0.99664, 0.99801, 0.99971\}$$

which can be reduced to

$$0.99437 \leq P\left(\bigcap_{m=1}^{3} A_m'\right) \leq 0.99664$$

Therefore, the corresponding first-order bounds for $P(A_1 \cup A_2 \cup A_3)$ is

$$1 - 0.99664 \leq P\left(\bigcup_{m=1}^{3} A_m\right) \leq 1 - 0.99437$$

which can be reduced to

$$0.00336 \leq P\left(\bigcup_{m=1}^{3} A_m\right) \leq 0.00563$$

Referring to Eq. (7.2), the first-order bounds for reliability of a series system with positively correlated nonfailure events can be computed as

$$\prod_{m=1}^{M} P(F'_m) \leq p_{s,\text{sys}} \leq \min_{m}\{P(F'_m)\} \tag{7.16a}$$

or in terms of failure probability as

$$\prod_{m=1}^{M} [1 - P(F_m)] \leq p_{s,\text{sys}} \leq \min_{m}\{1 - P(F_m)\} \tag{7.16b}$$

Similarly, referring to Eq. (7.7), by letting $A_m = F_m$, the first-order bounds on the failure probability of a parallel system with positively correlated failure events can be immediately obtained as

$$\prod_{m=1}^{M} P(F_m) \leq p_{f,\text{sys}} \leq \min_{m}\{P(F_m)\} \tag{7.17a}$$

and the corresponding bounds for the system reliability, according to Eq. (7.11b), is

$$1 - \min_{m}[P(F_m)] \leq p_{s,\text{sys}} \leq 1 - \left[\prod_{m=1}^{M} P(F_m)\right] \tag{7.17b}$$

Example 7.4 Consider that the M identical components in a system are positively correlated and the component reliability is p_s. Determine the reliability bounds for the system.

Solution If the system is a series system, the bounds on system reliability, according to Eq. (7.16a), are

$$p_s^M \leq p_{s,\text{sys}} \leq p_s$$

When the system is in parallel, the bounds on the system reliability, according to Eq. (7.17b), are

$$p_s \leq p_{s,\text{sys}} \leq 1 - (1 - p_s)^M$$

The bounds for system reliabilities for different M, with the component reliability of 0.95, for series and parallel systems are shown in Fig. 7.6. As can be observed, the bounds for the system widen as the number of components increases.

In the case that all events A_m's are negatively correlated, the following relationships hold:

$$P(A_m \mid A_j) \leq P(A_j) \qquad \text{for all } j, m$$
$$P(A_j, A_m) \leq P(A_j)P(A_m)$$
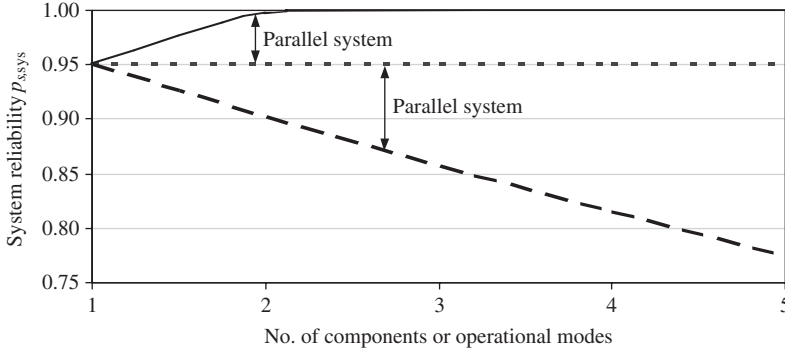
Figure 7.6 First-order bounds for series and parallel systems with component reliability $p_s = 0.95$.

Hence the first-order bounds for the probability of joint occurrence of several negatively correlated events is

$$0 \le P \left(\bigcap_{m=1}^{M} A_m \right) \le \prod_{m=1}^{M} P(A_m)$$

The bounds for reliability of a series system, with $A_m = F'_m$, containing negatively correlated events are

$$0 \le p_{s,\text{sys}} \le \prod_{m=1}^{M} [1 - P(F_m)]$$

whereas for a parallel system, with $A_m = F_m$,

$$1 - \left[\prod_{m=1}^{M} P(F_m) \right] \le p_{s,\text{sys}} \le 1$$

It should be pointed out that the first-order bounds for system reliability may be too wide to be meaningful. Tighter bounds sometimes are required and can be obtained at the expense of more computations.

Second-order bounds (Bimodal bounds). The second-order bounds are obtained by retaining the terms involving the joint probability of two events. By Eq. (2.4), the probability of the union of several events is

$$P \left(\bigcup_{m=1}^{M} A_m \right) = \sum_{m=1}^{M} P(A_m) - \sum_{i<} \sum_{j} P(A_i, A_j) + \sum_{i<} \sum_{j<} \sum_{k} P(A_i, A_j, A_k) - \cdots$$
$$+ (-1)^M P(A_1, A_2, \dots, A_M) \tag{7.18}$$

Notice the alternating signs in Eq. (7.18) as the order of the terms increases. It is evident that the inclusion of only the first-order terms, that is, $P(A_m)$, produces an upper bound for $P(A_1 \cup A_2 \cup \cdots \cup A_M)$. Consideration of the only the first two order terms yields a lower bound, the first three order terms again an upper bound, and so on (Melchers, 1999).

Simple bounds for the probability of a union are

$$\sum_{m=1}^{M} P(A_m) - \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} P(A_i, A_j) \leq P\left(\underset{m=1}{\overset{M}{\cup}} A_m\right) \leq \min\left[1, \sum_{m=1}^{M} P(A_m)\right] \quad (7.19)$$

It should be pointed out that these bounds produce adequate results only when the values of $P(A_m)$ and $P(A_m, A_j)$ are small. Equation (7.18) alternatively can be written as

$$P\left(\underset{m=1}{\overset{M}{\cup}} A_m\right) = [P(A_1)] + [P(A_2) - P(A_1, A_2)] + [P(A_3) - P(A_1, A_3)$$
$$- P(A_2, A_3) + P(A_1, A_2, A_3)] + [P(A_4) - P(A_1, A_4) - P(A_2, A_4)$$
$$- P(A_3, A_4) + P(A_1, A_2, A_4) + P(A_1, A_3, A_4) + P(A_2, A_3, A_4)$$
$$- P(A_1, A_2, A_3, A_4)] + [P(A_5) \cdots \qquad (7.20)$$

To derive the lower bound, consider each of the terms in brackets in Eq. (7.20). For example, consider the bracket containing the terms associated with event A_4. Note that apart from $P(A_4)$, the remaining terms in the bracket are

$$-P[(A_1, A_4) \cup (A_2, A_4) \cup (A_3, A_4)]$$

Furthermore, event A_4 contains $(A_1, A_4) \cup (A_2, A_4) \cup (A_3, A_4)$, which implies that $P(A_4) \geq P[(A_1, A_4) \cup (A_2, A_4) \cup (A_3, A_4)]$. Consequently, each of the bracketed terms in Eq. (7.20) has a nonnegative probability value. Also notice that

$$P\left[\underset{j=1}{\overset{3}{\cup}} (A_j, A_4)\right] \leq \sum_{j=1}^{3} P(A_j, A_4)$$

and thus the following inequality holds:

$$P(A_4) - P\left[\underset{j=1}{\overset{3}{\cup}} (A_j, A_4)\right] \geq P(A_4) - \sum_{j=1}^{3} P(A_j, A_4)$$

This equation can be generalized as

$$P(A_m) - P\left[\underset{j=1}{\overset{m-1}{\cup}} (A_j, A_m)\right] \geq P(A_m) - \sum_{j=1}^{m-1} P(A_j, A_m) \qquad (7.21)$$

It should be pointed out that the terms on the right-hand-side of Eq. (7.21) could be negative, especially when m is large. Owing to the fact that each of the

bracketed terms should be nonnegative, a better lower bound to Eq. (7.20) can be obtained if the right-hand-side of Eq. (7.21) makes a nonnegative contribution to the lower bound (Ditlevsen, 1979), namely,

$$P\left(\bigcup_{m=1}^{M} A_m\right) \geq P(A_1) + \sum_{m=2}^{M} \max\left\{\left[P(A_m) - \sum_{j=1}^{m-1} P(A_j, A_m)\right], 0\right\} \quad (7.22)$$

Earlier, Kounias (1968) proposed an alternative second-order lower bound by selecting only those combinations in Eq. (7.20) which give the maximum values of the lower bound:

$$P\left(\bigcup_{m=1}^{M} A_m\right) \geq P(A_1) + \max\left\{\sum_{\substack{m=2 \\ j<m}}^{M} [P(A_m) - P(A_j, A_m)]\right\} \quad (7.23)$$

It should be pointed out that both lower bounds for the probability of a union depend on the order in which the events are labeled. Algorithms have been developed for identifying the optimal ordering of events to obtain the best bounds (Dawson and Sankoff, 1967; Hunter, 1977). A useful rule of thumb is to order the events in the order of decreasing importance (Melchers, 1999). In other words, events are ordered such that $P(A_{[1]}) > P(A_{[2]}) > \cdots > P(A_{[M]})$, with $[m]$ representing the rank of the event according to its probability of occurrence. For a given ordering, Ramachandran (1984) showed that the lower bound provided by Eq. (7.22) is better than Eq. (7.23), whereas both bounds are equal if all possible orderings are considered.

To derive the upper bound, attention is focused back to Eq. (7.20) and on each of the terms in brackets. For example, consider the bracket containing the terms associated with event A_4. As discussed earlier, apart from $P(A_4)$, the remaining terms in the bracket are

$$-P[(A_1, A_4) \cup (A_2, A_4) \cup (A_3, A_4)]$$

Using the fact that $P(A \cup B) \geq \max[P(A), P(B)]$, the following inequality holds:

$$P\left[\bigcup_{j<4} (A_j, A_4)\right] \geq \max_{j<4}[P(A_j, A_4)]$$

Hence the probability in the fourth bracket involving event A_4 satisfies

$$P(A_4) - P\left[\bigcup_{j<4} (A_j, A_4)\right] \leq P(A_4) - \max_{j<4}[P(A_j, A_4)] \quad (7.24)$$

This inequality relation is true for all the bracketed terms of Eq. (7.22), and the upper bound can be obtained as

$$P\left(\bigcup_{m=1}^{M} A_m\right) \leq \min\left\{1, \sum_{m=1}^{M} P(A_m) - \sum_{m=2}^{M} \max_{j<m}[P(A_j, A_m)]\right\} \quad (7.25)$$

Again, this upper bound value also is dependent on the order of events. As can be seen from Eqs. (7.22) and (7.25), it is realized that the computation of a second-order bound for the probability of a union requires determination of the joint probability for all combinations of all possible pairs of events involved.

Referring to Eq. (7.1), the second-order bounds for the failure probability of a series system can be obtained as

$$P_{f,\text{sys}}\begin{cases} \leq \sum_{m=1}^{M} P(F_m) - \sum_{m=2}^{M} \max_{j<m}[P(F_j, F_m)] \\ \geq P(F_1) + \sum_{m=2}^{M} \max\left\{\left[P(F_m) - \sum_{j=1}^{m-1} P(F_j, F_m)\right], 0\right\} \end{cases} \tag{7.26}$$

Similarly, the second-order bounds for the reliability of a parallel system, according to Eq. (7.8), are

$$P_{f,\text{sys}}\begin{cases} \leq \sum_{m=1}^{M} P(F'_m) - \sum_{m=2}^{M} \max_{j<m}[P(F'_j, F'_m)] \\ \geq P(F'_1) + \sum_{m=2}^{M} \max\left\{\left[P(F'_m) - \sum_{j=1}^{m-1} P(F'_j, F'_m)\right], 0\right\} \end{cases} \tag{7.27}$$

Example 7.5   Refer to Example 7.3. Compute the second-order bounds for the multivariate normal probability.

Solution   To compute the second-order bounds, the probabilities of individual events as well as the joint probabilities between two different event pairs must be computed. From Table 2.2 or Eq. (2.63), the probabilities of individual events are

$$P(A_1) = P(Z_1 \leq -2.71) = 0.003364$$

$$P(A_2) = P(Z_2 \leq -2.88) = 0.001988$$

$$P(A_3) = P(Z_3 \leq -3.44) = 0.0002909$$

The joint probabilities, according to the procedures described in Sec. 2.7.2, are

$$P(A_1, A_2) = P(Z_1 \leq -2.71, Z_2 \leq -2.88 \mid \rho = 0.841) = 0.0009247$$

$$P(A_1, A_3) = P(Z_1 \leq -2.71, Z_3 \leq -3.44 \mid \rho = 0.014) = 0.000001142$$

$$P(A_2, A_3) = P(Z_2 \leq -2.88, Z_3 \leq -3.44 \mid \rho = 0.536) = 0.00004231$$

The lower bound of $P(A_1 \cup A_2 \cup A_3)$, according to Eq. (7.23), is

$$P(A_1) + \sum_{m=2}^{3} \max\left\{\left[P(A_m) - \sum_{j=1}^{m-1} P(A_j, A_m)\right], 0\right\}$$

$$= P(A_1) + \max\left\{[P(A_2) - P(A_1, A_2)], 0\right\} + \max\left\{[P(A_3) - P(A_1, A_2) - P(A_2, A_3)], 0\right\}$$

$$= 0.003364 + \max\left\{[0.001988 - 0.0009247], 0\right\} + \max\left\{[2.909 \times 10^{-4} - 1.142 \times 10^{-6})\right.$$

$$\left. -4.231 \times 10^{-5}], 0\right\} = 0.003364 + 0.00106633 + 0.00024736$$

$$= 0.004675$$

The upper bound of $P(A_1 \cup A_2 \cup A_3)$, according to Eq. (7.25), can be computed as

$$\sum_{m=1}^{3} P(A_m) - \sum_{m=2}^{3} \max_{j<m}[P(A_j, A_m)]$$

$$= \sum_{m=1}^{3} P(A_m) - \left\{ \max[P(A_1, A_2)] + \max[P(A_1, A_3), P(A_2, A_3)] \right\}$$

$$= (0.003364 + 0.001988 + 0.002909) - \left\{ \max(0.0009247) \right.$$

$$\left. + \max[(1.142 \times 10^{-6}, 4.231 \times 10^{-5})] \right\} = 0.005641 - (0.000924 + 0.0000425)$$

$$= 0.004677$$

In summary, the second-order bounds for the trivariate normal probability $P(A_1 \cup A_2 \cup A_3)$ are

$$0.004675 \leq P(A_1 \cup A_2 \cup A_3) \leq 0.004677$$

Comparing with the first-order bounds obtained in Example 7.3, the second-order bounds are definitely tighter.

7.3   Reliability of Simple Systems

In this section the reliability of some simple systems will be discussed. In the framework of time-to-failure analysis, availability of such systems will be presented. Information such as this is essential to serve as the building blocks for determination of reliability or availability of more complex systems.

7.3.1   Series systems

A series system requires that all its components or modes of operation perform satisfactorily to ensure a satisfactory operation of the entire system. In the context of load-resistance interference, the failure event associated with a mode of operation is

$$F_m = \{W_m < 0\} \qquad \text{for } m = 1, 2, \ldots, M$$

in which W_m is the random performance variable associated with the mth mode of operation. Referring to Chap. 4, the failure probability and reliability associated with the mth mode of operation, respectively, are

$$P(F_m) = P(W_m < 0) = P(Z_m \leq -\beta_m) = \Phi(-\beta_m) \tag{7.28a}$$

$$P(F'_m) = P(W_m > 0) = P(Z_m > -\beta_m) = \Phi(\beta_m) \tag{7.28b}$$

in which Z_m is the standard normal random variable associated with W_m, and β_m is the reliability index associated with the mth mode of operation.

The failure probability of a series system involving M modes of operation, according to Eq. (7.1), can be expressed as

$$p_{f,\text{sys}} = P\left[\bigcup_{m=1}^{M}(W_m < 0)\right] = P\left[\bigcup_{m=1}^{M}(Z_m < -\beta_m)\right] \tag{7.29}$$

Because all the standardized normal random variables Z_m's generally are correlated, computation of the exact system failure probability using Eq. (7.29) may not be practical, especially when the number of modes of operation M is large. For this case, the second-order bounds for $p_{f,\text{sys}}$ could be viable. According to Eq. (7.26), the bounds for system failure probability are

$$p_{f,\text{sys}}\begin{cases} \leq \displaystyle\sum_{m=1}^{M}\Phi(-\beta_m) + \sum_{m=2}^{M}\max_{j<m}[\Phi(-\beta_j, -\beta_m\,|\,\rho_{jm})] \\[2ex] \geq \Phi(-\beta_1) + \displaystyle\sum_{m=2}^{M}\max\left\{\left[\Phi(-\beta_m) - \sum_{j=1}^{m-1}\Phi(-\beta_j, -\beta_m\,|\,\rho_{jm})\right]\right\} \end{cases} \tag{7.30}$$

in which $\Phi(-\beta_j, -\beta_m\,|\,\rho_{jm})$ is the bivariate normal probability, which can be computed by procedures described in Sec. 2.7.2, with ρ_{jm} being the correlation coefficient between the performance variables W_j and W_m for the jth and mth modes of operation. Accordingly, the bounds on reliability of a series system can be obtained easily by using Eq. (7.11b).

Although computation of the exact bivariate normal probability can be obtained through numerical integration, sometimes information about its bounds is sufficient. Under a positively correlated case, narrow bounds of $\Phi(-\beta_j, -\beta_m\,|\,\rho_{jm} > 0)$ that require evaluations of only univariate normal probabilities are

$$\Phi(-\beta_j, -\beta_m\,|\,\rho_{jm} > 0)\begin{cases} \leq \Phi(-\beta_m)\Phi(-\beta_{j\,|\,m}) + \Phi(-\beta_j)\Phi(-\beta_{m\,|\,j}) \\[2ex] \geq \max[\Phi(-\beta_m)\Phi(-\beta_{j\,|\,m}), \Phi(-\beta_j)\Phi(-\beta_{m\,|\,j})] \end{cases} \tag{7.31}$$

where

$$\beta_{j\,|\,m} = \frac{\beta_j - \rho_{jm}\beta_m}{\sqrt{1-\rho_{jm}^2}} \tag{7.32a}$$

$$\beta_{m\,|\,j} = \frac{\beta_m - \rho_{jm}\beta_j}{\sqrt{1-\rho_{jm}^2}} \tag{7.32b}$$

In the case that the pair of performance functions is negatively correlated, the bounds for joint failure probability are

$$0 \leq \Phi(-\beta_j, -\beta_m\,|\,\rho_{jm} < 0) \leq \min[\Phi(-\beta_m)\Phi(-\beta_{j\,|\,m}), \Phi(-\beta_j)\Phi(-\beta_{m\,|\,j})] \tag{7.33}$$

The derivations of Eqs.(7.31) and (7.33) are given in Appendix 7A. Ang and Tang (1984) pointed out that use of an approximation of Eq. (7.31) could improve (tighten) the second-order bound of Eq. (7.30) when the single-mode failure probabilities are small, say, on the order of 10^{-4}. However, if the single-mode

failure probabilities are all large (e.g., 10^{-2}), the bound of Eq. (7.31) will be wide.

In fact, the reliability of a series system can be computed, according to Eq. (7.2), as

$$p_{s,\text{sys}} = P\left(\bigcap_{m=1}^{M} F'_m\right) = P\left[\bigcap_{m=1}^{M} (Z_m \geq -\beta_m)\right] \tag{7.34}$$

It should be pointed out that, in general, $P\left[\cap(Z_m \geq -\beta_m)\right] \neq P\left[\cap(Z_m \leq \beta_m)\right]$ unless for the univariate case. As can be seen, the reliability of a series system is the multivariate normal probability whose determination can be made by Ditlevsen's approach, described in Sec. 2.7.2, or by various bounding approaches discussed in Sec. 2.7.3.

Example 7.6   Consider a system consisting of three modes of operation, each of which is specified by the following linear performance functions:

$$W_1(\boldsymbol{X}) = X_1 + 2X_2$$
$$W_2(\boldsymbol{X}) = X_1 + X_2 + X_3$$
$$W_3(\boldsymbol{X}) = X_2 + 2X_3$$

in which the stochastic basic variables X_1, X_2, and X_3 are multivariate normal random variables with the vector of means

$$\boldsymbol{\mu}_x = (\mu_1, \mu_2, \mu_3)^t = (6, 6, 6)^t$$

and covariance matrix

$$\boldsymbol{C}_x = \begin{bmatrix} 9.00 & 0.00 & 0.00 \\ 0.00 & 9.00 & 0.00 \\ 0.00 & 0.00 & 9.00 \end{bmatrix}$$

The state of the system is such that if any of the three modes of operation fail, the system would fail. Calculate the system reliability.

Solution   From the preceding covariance matrix $\boldsymbol{C}_x$, it is understood that all three stochastic basic variables are uncorrelated, each with a variance of 9, that is, $\text{Var}(X_1) = \text{Var}(X_2) = \text{Var}(X_3) = 9$. The vector of expected values of W_1, W_2, and W_3 is

$$\boldsymbol{\mu}_w = (\mu_{w_1}, \mu_{w_2}, \mu_{w_3})^t = [6 + 2(6), 6 + 6 + 6, 6 + 2(6)]^t = (18, 18, 18)^t$$

The covariance matrix of the three performance functions W's can be computed as

$$\boldsymbol{C}_w = \boldsymbol{S}^t \boldsymbol{C}_x \boldsymbol{S}$$

in which $\boldsymbol{S}$, the sensitivity matrix, is an $K \times M$ matrix, with M and K being the number of performance functions and stochastic basic variables, respectively. The sensitivity matrix $\boldsymbol{S}$ contains, in each column, the vector of sensitivity coefficients for each performance function with respect to individual stochastic basic variable, that is,

$$\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M]$$

with

$$\boldsymbol{s}_m^t = \left[\frac{\partial W_m}{\partial X_1}, \frac{\partial W_m}{\partial X_2}, \ldots, \frac{\partial W_m}{\partial X_K} \right]_{\mu_x}^t$$

for $m = 1, 2, \ldots, M$. In this example, since all performance functions are linear, the sensitivity matrix consists of coefficients in the performance functions, that is,

$$\boldsymbol{S}^t = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Hence the covariance matrix of the three performance functions can be obtained as

$$\boldsymbol{C}_w = \begin{bmatrix} 45 & 27 & 18 \\ 27 & 27 & 27 \\ 18 & 27 & 45 \end{bmatrix}$$

As can be seen, even though the three stochastic basic variables are uncorrelated, the three performance functions are correlated because they are defined by some stochastic basic variables common to the other performance functions. Hence the variances of W_1, W_2, and W_3 appear on the diagonal of $\boldsymbol{C}_w$, namely,

$$\text{Var}(W_1) = 45 \qquad \text{Var}(W_2) = 27 \qquad \text{and} \qquad \text{Var}(W_3) = 47$$

The corresponding correlation matrix of random W's can be obtained easily as

$$\boldsymbol{R}_w = \begin{bmatrix} 1.000 & 0.7746 & 0.4000 \\ 0.7746 & 1.000 & 0.7746 \\ 0.4000 & 0.7746 & 1.000 \end{bmatrix}$$

The system failure probability is defined as

$$p_{f,\text{sys}} = P\left[(W_1 < 0) \cup (W_2 < 0) \cup (W_3 < 0)\right]$$
$$= P\left[(Z_1 < -2.68) \cup (Z_2 < -3.46) \cup (Z_3 < -2.68)\right]$$

The exact system failure probability can be obtained as

$$p_{f,\text{sys}} = P\left[(Z_1 < -2.68) \cup (Z_2 < -3.46) \cup (Z_3 < -2.68)\right]$$

$$= \left[P(Z_1 < -2.68) + P(Z_2 < -3.46) + P(Z_3 < -2.68)\right]$$

$$\quad - \left[P(Z_1 < -2.68, Z_2 < -3.46) + P(Z_1 < -2.68, Z_3 < -2.68)\right.$$

$$\left. + P(Z_2 < -3.46, Z_3 < -2.68)\right] + P(Z_1 < -2.68, Z_2 < -3.46, Z_3 < -2.68)$$

$$= (0.003681 + 0.0002701 + 0.003681) - (0.0001659 + 0.0001987 + 0.0001659)$$

$$\quad + 0.0001556$$

$$= 0.0072572$$

Hence the system reliability $p_{s,\text{sys}} = 1 - 0.0072572 = 0.9927428$. Note that the preceding bivariate normal probabilities are calculated by Eq. (2.121), whereas the trivariate normal probability is computed according to Ditlevsen's algorithm using Taylor expansion described in Sec. 2.7.2.

Alternatively, the second-order bounds for the system failure probability can be computed according to Eq. (7.30). The results are

$$0.007102 \leq p_{f,\text{sys}} \leq 0.007268$$

and the corresponding bounds on the system reliability $p_{s,\text{sys}}$ are

$$0.992732 \leq p_{s,\text{sys}} \leq 0.992898$$

In the framework of time-to-failure analysis, the reliability $p_{s,m}(t)$ and failure probability $p_{f,m}(t)$ of the mth component over the time interval $(0, t]$, according to Eqs. (5.1a) and (5.1b), are

$$P(F'_m) = p_{s,m}(t) = P(T_m \geq t) = \int_t^\infty f_m(\tau)\,d\tau \qquad (7.35a)$$

and

$$P(F_m) = p_{f,m}(t) = \int_0^t f_m(\tau)\,d\tau \qquad (7.35b)$$

respectively, where $f_m(t)$ is the failure density function for the mth component.

In the case that the performance of individual components is independent of each other, the reliability of a series system is

$$p_{s,\text{sys}}(t) = P\left(\bigcap_{m=1}^M F'_m \right) = \prod_{m=1}^M p_{s,m}(t) \qquad (7.36)$$

Similarly, the availability of a series system involving M independent components is

$$A_{\text{sys}}(t) = \prod_{m=1}^M A_m(t) \qquad (7.37)$$

in which $A_{\text{sys}}(t)$ and $A_m(t)$ are availabilities of the entire system and the mth component, respectively, at time t.

According to Eqs. (5.2) and (5.3), the failure density function $f_{\text{sys}}(t)$ and the failure rate $h_{\text{sys}}(t)$ for a series system involving M independent components can be derived, respectively, as

$$f_{\text{sys}}(t) = -\frac{d\,[p_{s,\text{sys}}(t)]}{dt} = \sum_{m=1}^M \left[\prod_{j \neq m} p_{s,m}(t) \right] f_m(t) \qquad (7.38)$$

and

$$h_{\text{sys}}(t) = \frac{f_{\text{sys}}(t)}{p_{s,\text{sys}}(t)} = \sum_{m=1}^M \frac{f_m(t)}{p_{s,m}(t)} \qquad (7.39)$$

For the special case of an exponential failure density function such as

$$f_m(t) = \lambda_m e^{-\lambda_m t} \qquad \text{for } t \geq 0,\, \lambda_m > 0,\, m = 1, 2, \ldots, M$$

the reliability and unreliability of a series system with M independent components, respectively, are

$$p_{s,\text{sys}}(t) = \prod_{m=1}^{M} e^{-\lambda_m t} = \exp\left[-\left(\sum_{m=1}^{M} \lambda_m\right) t\right] \tag{7.40a}$$

$$p_{f,\text{sys}}(t) = 1 - \exp\left[-\left(\sum_{m=1}^{M} \lambda_m\right) t\right] \tag{7.40b}$$

Assuming an exponential repair function for each independent component, the availability and unavailability for a series system, according to Eqs. (7.37) and (5.59), are

$$A_{\text{sys}}(t) = \prod_{m=1}^{M} \left(\frac{\eta_m}{\lambda_m + \eta_m} + \frac{\lambda_m}{\lambda_m + \eta_m} e^{-(\lambda_m + \eta_m)t}\right) \tag{7.41a}$$

and $\qquad\qquad U_{\text{sys}}(t) = 1 - A_{\text{sys}}(t) \tag{7.41b}$

in which η_m is the constant repair rate for the mth component, and $U_{\text{sys}}(t)$ is the system unavailability at time t. The stationary system availability, by Eq. (5.60), can be expressed as

$$A_{\text{sys}}(\infty) = \prod_{m=1}^{M} \left(\frac{\eta_m}{\lambda_m + \eta_m}\right) = \prod_{m=1}^{M} \left(\frac{\text{MTTF}_m}{\text{MTTR}_m + \text{MTTF}_m}\right) \tag{7.42}$$

in which MTTR_m and MTTF_m are, respectively, the mean time to repair and mean time to failure of the mth component.

Example 7.7   As an example of a series system, consider a pumping station consisting of two different pumps in series, both of which must operate to pump the required quantity. The constant failure rates for the pumps are $\lambda_1 = 0.0003$ failures/h and $\lambda_2 = 0.0002$ failures/h. For a 2000-h mission time, the system reliability, according to Eq. (7.40a), is

$$p_{s,\text{sys}}(t = 2000) = \exp[-(0.0003 + 0.0002)(2000)] = 0.90484$$

and the MTTF of the system is

$$\text{MTTF} = \frac{1}{\lambda_1 + \lambda_2} = \frac{1}{0.0003 + 0.0002} = 2000\,\text{h}$$

7.3.2 Parallel systems

For a parallel system, the entire system would perform satisfactorily if any one or more of its components or modes of operation is functioning satisfactorily; the entire system would fail only if all its components or modes of operation fail.

In the framework of load-resistance interference for different modes of operation, the failure probability of a parallel system, according to Eq. (7.9), is

$$p_{f,\text{sys}} = P\left[\bigcap_{m=1}^{M}(W_m < 0)\right] = P\left[\bigcap_{m=1}^{M}(Z_m < -\beta_m)\right] = \Phi(-\boldsymbol{\beta} \mid \boldsymbol{R}_z) \qquad (7.43)$$

which can be computed as the multivariate normal probability discussed in Sec. 2.7.2. The bounds for system failure probability also can be computed if the exact value of $p_{f,\text{sys}}$ is not required. In the case that all performance variables W's are independent, the system failure probability reduces to

$$p_{f,\text{sys}} = \prod_{m=1}^{M}\Phi(-\beta_m) \qquad (7.44)$$

Alternatively, the reliability of a parallel system can be expressed as

$$p_{s,\text{sys}} = P\left[\bigcup_{m=1}^{M}(W_m > 0)\right] = P\left[\bigcup_{m=1}^{M}(Z_m > -\beta_m)\right] \qquad (7.45)$$

The second-order bounds for this system reliability, according to Eq. (7.27), are

$$p_{s,\text{sys}} \begin{cases} \leq \displaystyle\sum_{m=1}^{M-1}\Phi(\beta_m) - \sum_{m=2}^{M}\max_{j<m}[L(-\beta_j, -\beta_m \mid \rho_{jm})] \\[2ex] \geq \Phi(\beta_1) + \displaystyle\sum_{m=2}^{M}\max\left\{0, \left[\Phi(\beta_m) - \sum_{j=1}^{m-1}L(-\beta_j, -\beta_m \mid \rho_{jm})\right]\right\} \end{cases} \qquad (7.46)$$

in which $L(-\beta_j, -\beta_m \mid \rho_{jm}) = P[Z_j \geq -\beta_j, Z_m \geq -\beta_m] = \Phi(-\beta_j, -\beta_m \mid \rho_{jm}) + \Phi(\beta_j) + \Phi(\beta_m) - 1$.

Example 7.8   Referring to Example 7.6, determine the system reliability by considering that the system would fail if all three modes of operation fail.

Solution   Since the system is in parallel, the system failure probability can be calculated as

$$\begin{aligned} p_{f,\text{sys}} &= P(W_1 < 0, W_2 < 0, W_3 < 0) \\ &= P(Z_1 < -2.68, Z_2 < -3.46, Z_3 < -2.68) \\ &= 0.0001556 \end{aligned}$$

which is obtained in Example 7.6. Hence the reliability of the system is 0.9998444.

In the framework of time-to-failure analysis, the unreliability of a parallel system involving M independent components can be computed as

$$p_{f,\text{sys}}(t) = \prod_{m=1}^{M}p_{f,m}(t) \qquad (7.47)$$

in which $p_{f,m}(t) = P(T_m \leq t)$, the unreliability of the mth component within the specified time interval $(0, t]$. Hence the system reliability in time interval $(0, t]$ is

$$p_{s,\text{sys}}(t) = 1 - p_{f,\text{sys}}(t) = 1 - \prod_{m=1}^{M} p_{f,m}(t) \tag{7.48}$$

The failure density function $f_{\text{sys}}(t)$ and failure rate $h_{\text{sys}}(t)$ for a parallel system consisting of M independent components are

$$f_{\text{sys}}(t) = \frac{d\,[p_{f,\text{sys}}(t)]}{dt} = \sum_{m=1}^{M} \left[\prod_{j \neq m}^{M} p_{f,j}(t) \right] f_m(t) \tag{7.49}$$

and

$$h_{\text{sys}}(t) = \frac{f_{\text{sys}}(t)}{p_{s,\text{sys}}(t)} = \frac{\sum\limits_{m=1}^{M} \left[\prod\limits_{j \neq m}^{M} p_{f,j}(t) \right] f_m(t)}{1 - \prod\limits_{m=1}^{M} p_{f,m}(t)} \tag{7.50}$$

For each component having an exponential failure density function with the parameter λ_m, for $m = 1, 2, \ldots, M$, the failure probability of a parallel system can be computed as

$$p_{f,\text{sys}}(t) = \prod_{m=1}^{M} (1 - e^{-\lambda_m t}) \tag{7.51}$$

with the corresponding system reliability

$$p_{s,\text{sys}}(t) = 1 - \prod_{m=1}^{M} (1 - e^{-\lambda_m t}) \tag{7.52}$$

The system failure density function $f_{\text{sys}}(t)$ is

$$f_{\text{sys}}(t) = \sum_{m=1}^{M} \left[\prod_{j \neq m}^{M} (1 - e^{-\lambda_j t}) \right] \lambda_m e^{-\lambda_m t} \tag{7.53}$$

In the case that all components have an identical failure rate, that is, $\lambda_1 = \lambda_2 = \cdots = \lambda_M = \lambda$, the MTTF of the system is

$$\text{MTTF} = \frac{1}{\lambda} \sum_{m=1}^{M} \frac{1}{m} \tag{7.54}$$

The unavailability of a parallel system involving M independent components is

$$U_{\text{sys}}(t) = \prod_{m=1}^{M} U_m(t) \tag{7.55a}$$

and the corresponding system availability is

$$A_{\text{sys}}(t) = 1 - \prod_{m=1}^{M} U_m(t) \tag{7.55b}$$

Under the condition of independent exponential repair functions for the M components, the unavailability of a parallel system is

$$U_{\text{sys}}(t) = \prod_{m=1}^{M} \frac{\lambda_m}{\lambda_m + \eta_m}(1 - e^{-(\lambda_m + \eta_m)t}) \tag{7.56}$$

and the stationary system unavailability is

$$U_{\text{sys}}(\infty) = \prod_{m=1}^{M} \frac{\lambda_m}{\lambda_m + \eta_m} = \prod_{m=1}^{M} \left(\frac{\text{MTTR}_m}{\text{MTTR}_m + \text{MTTF}_m} \right) \tag{7.57}$$

Example 7.9 As an example of a parallel system, consider a pumping station consisting of two identical pumps operating in a redundant configuration so that either pump could fail, and the peak discharge could still be delivered. Both pumps have a failure rate of $\lambda = 0.0005$ failures/h, and both pumps start operating at $t = 0$. The system reliability for a mission time of $t = 1000$ h, according to Eq. (7.52), is

$$p_{s,\text{sys}}(t = 1000) = 1 - \left(1 - e^{-(0.0005)(1000)}\right)\left(1 - e^{-(0.0005)(1000)}\right) = 0.8452$$

The MTTF, according to Eq. (7.53), is

$$\text{MTTF} = \frac{1}{\lambda}\left(\frac{1}{1} + \frac{1}{2}\right) = \frac{1.5}{\lambda} = 1.5\left(\frac{1}{0.0005}\right) = 3000\,\text{h}$$

7.3.3 K-out-of-M parallel systems

This is a parallel system of M component for which the system would function if $K(K < M)$ or more components function. This type of system also is called a *partially redundant system*. The general reliability formula for this system is rather cumbersome. For components having an identical reliability function, that is, $p_{s,m}(t) = p_s(t)$, the system reliability and unreliability, when component performances are independent, are

$$p_{s,\text{sys}}(t) = \sum_{j=K}^{M} C_{M,j}\,[p_s(t)]^j\,[1 - p_s(t)]^{M-j} \tag{7.58a}$$

and
$$p_{f,\text{sys}}(t) = \sum_{j=0}^{K-1} C_{M,j}\,[p_s(t)]^j\,[1 - p_s(t)]^{M-j} \tag{7.58b}$$

in which $C_{M,j} = M!/[j!(M - j)!]$ is a binomial constant. Computationally, whether to calculate $p_{s,\text{sys}}(t)$ or $p_{f,\text{sys}}(t)$ is dictated by the number of terms

involved in the summation. Furthermore, if the failure density function is an exponential distribution, the system reliability can be expressed as

$$p_{s,\text{sys}}(t) = \sum_{j=K}^{M} C_{M,j}(e^{-\lambda t})^{j}(1 - e^{-\lambda t})^{M-j} \tag{7.59}$$

The failure density function for the system $f_{\text{sys}}(t)$ based on the system reliability in Eq. (7.58a) is

$$
\begin{aligned}
f_{\text{sys}}(t) &= -\frac{d\left[p_{s,\text{sys}}(t)\right]}{dt} \\
&= \sum_{j=K}^{M} C_{M,j}\left[p_s(t)\right]^{j}\left[1 - p_s(t)\right]^{M-j}\left[\frac{j}{p_s(t)} - \frac{M-j}{1 - p_s(t)}\right] f_t(t) \tag{7.60}
\end{aligned}
$$

The availability and unavailability of the system can be obtained from substituting component availability for component reliability in Eqs. (7.58a) and (7.58b), respectively.

Example 7.10 As an example of a K-out-of-M system, consider a pumping system with three pumps, one of which is on standby, all with constant failure rates of $\lambda = 0.0005$ failures/h. The system reliability for $t = 1000$ h, $M = 3$, and $K = 2$ is

$$
\begin{aligned}
p_{s,\text{sys}}(t = 1000) &= C_{3,2}\left(e^{-(0.0005)(1000)}\right)^{2}\left(1 - e^{-(0.0005)(1000)}\right) + C_{3,3}\left(e^{-(0.0005)(1000)}\right)^{3} \\
&= 3\left(e^{-(0.0005)(1000)}\right)^{2} - 2\left(e^{-(0.0005)(1000)}\right)^{3} \\
&= 1.1036 - 0.4463 \\
&= 0.6573
\end{aligned}
$$

7.3.4 Standby redundant systems

A *standby redundant system* is a parallel system in which only one component or subsystem is in operation (Fig. 7.7). It is a special case of K-out-of-M system with $K = 1$. If the operating component fails, then another component is operated. This type of system is different than the parallel system described in Sec. 7.3.2, where all components are concurrently operating because standby
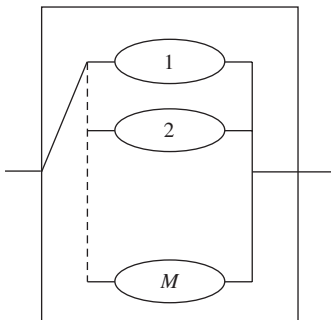
Figure 7.7 Standby redundant systems.

units do not operate. The system reliability for a system with M components out of which $M-1$ units are on standby is the probability that at most $M-1$ components fail. This probability can be expressed by

$$p_{s,\text{sys}}(t) = \sum_{m=0}^{M-1} \frac{(\lambda t)^m e^{-\lambda t}}{m!} \tag{7.61}$$

Note that this equation is valid under the following assumptions: The switching arrangement is perfect, the units are identical, the component failure rates are constant, the standby units are as good as new, and the unit failures are statistically independent. The mean time to failure of the system can be obtained, according to Eq. (5.18), as

$$\text{MTTF} = \int_0^\infty p_{s,\text{sys}}(t)\, dt = \int_0^\infty \sum_{m=0}^{M-1} \frac{(\lambda t)^m e^{-\lambda t}}{m!}\, dt = \frac{M}{\lambda} \tag{7.62}$$

Equation (7.62) is intuitively obvious in that the system's operation is the result of a relay of a series of components. As one component fails, the second one comes to operation until failure occurs. Therefore, the system MTTF is the sum of the MTTFs of individual components.

Example 7.11 As an example of a standby redundant system, assume an exponential failure distribution for two identical pumps, one operating and the second on standby, with identical failure rates of $\lambda = 0.0005$ failures/h. The standby unit is as good as new at time $t = 0$. The system reliability for $t = 1000$ h is

$$p_{s,\text{sys}}(t = 1000) = [1 + (0.0005)(1000)]e^{-(0.0005)(1000)} = 0.9098$$

7.4 Methods for Computing Reliability of Complex Systems

Evaluation of the reliability of simple systems, as described in the preceding section, is generally straightforward. However, many practical hydrosystems engineering infrastructures, such as water distribution systems, have neither series nor parallel configuration. Evaluation of the reliability for such complex systems generally is difficult. For some systems, with their components arranged in a complex configuration, it is possible to combine components into groups in such a manner that it appears as in series or in parallel. For other systems, special techniques have to be developed that require a certain degree of insight and ingenuity from engineers. A great deal of work has been done on developing techniques for evaluating the reliability of complex systems. This section describes some of the potentially useful techniques for hydrosystems reliability evaluation.

7.4.1 State enumeration method

The *state enumeration method* lists all possible mutually exclusive states of the system components that define the state of the entire system. In general,

for a system containing M components, each of which can be classified into K operating states, there will be K^M possible states for the entire system. For example, if the state of each of the M components is classified into failed and operating states, the system has 2^M possible states.

Once all the possible system states are enumerated, the states that result in successful system operation are identified, and the probability of the occurrence of each successful state is computed. The last step is to sum all the successful state probabilities, which yields the system reliability. This method becomes less and less computationally attractive, as one can imagine, when the number of system components and/or the number of states for each component gets larger.

The tree diagram, such as that in Fig. 7.8, is called an *event tree*, and the analysis involving the construction of an event tree is referred to as *event-tree analysis*. As can be seen, an event tree simulates not only the topology of
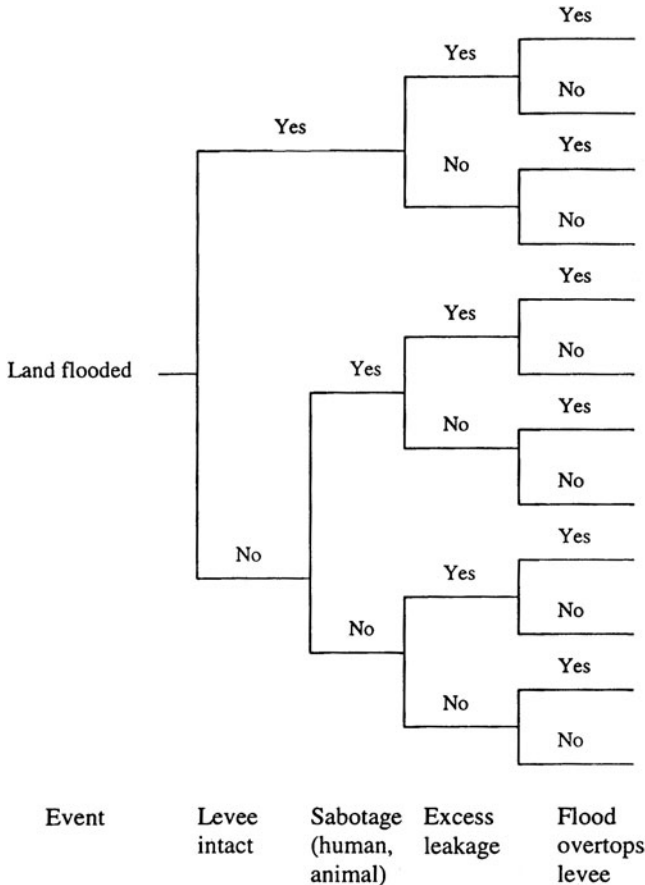
Figure 7.8 An example event tree for land flooding relating to levee performance.

a system but, more important, the sequential or chronologic operation of the system.

Example 7.12 Consider a simple water distribution network consisting of five pipes and one loop, as shown in Fig. 7.9. Node 1 is the source node, and nodes 3, 4, and 5 are demand nodes. The components of this network subject to possible failure are the five pipe sections. Within a given time period, each pipe section has an identical failure probability of 5 percent due to breakage or other causes that require it to be removed from service. The system reliability is defined as the probability that water can reach all three demand nodes from the source. Furthermore, it is assumed that the states of serviceability of each pipe are independent.

Solution Using the state enumeration method for system reliability evaluation, the associated event tree can be constructed to depict all possible combinations of component states in the system, as shown in Fig. 7.10. Since each pipe has two possible states, that is, failure F or nonfailure F', the tree, if fully expanded, would have $2^5 = 32$ branches. However, knowing the role that each pipe component plays in the network connectivity, exhaustive enumeration of all possible states is not necessary.

For example, referring to Fig. 7.10, one realizes that when pipe 1 fails, all demand nodes cannot receive water, indicating a system failure, regardless of the state of the remaining pipe sections. Therefore, branches in the event tree beyond this point do not have to be constructed. Applying some judgment in event-tree construction in this fashion generally can lead to a smaller tree. However, for a complex system, this may not be a trivial task.

The system reliability can be obtained by summing up the probabilities associated with all of the nonfailure branches. In this example, there are five branches, as
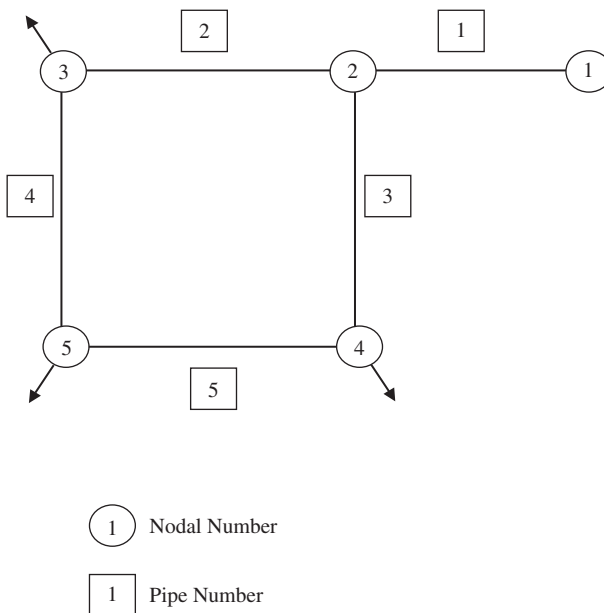
Figure 7.9 Pipe network configuration for Example 7.12.
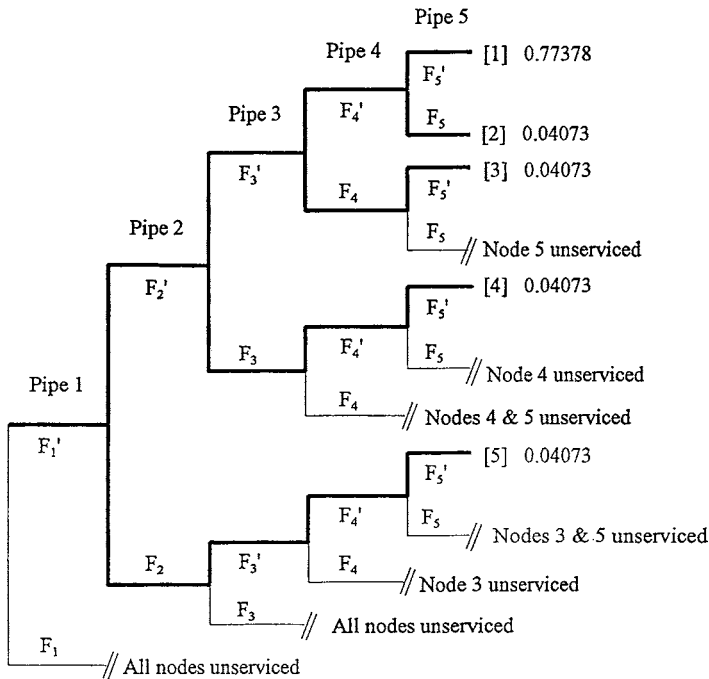
F_i' = Non-failure of pipe i

F_i = Failure of pipe i

// = Branch associated with unserviceability of one or more node

Figure 7.10 Tree diagram showing failure cases for the example pipe network.

indicated by the heavy lines in the tree, for which all users can have the water delivered by the system. Therefore, the system reliability is

$$p_{s,\text{sys}} = P \left(\bigcup_{m=1}^{5} B_{[m]} \right) = \sum_{m=1}^{5} P(B_{[m]})$$

where $P(B_{[m]})$ is the probability that the branch $B_{[m]}$ of the event tree provides full service to all users. The probability associated with each branch resulting in satisfactory delivery of water to all users can be calculated as the following:

$$P(B_{[1]}) = P(F_1')P(F_2')P(F_3')P(F_4')P(F_5')$$

$$= (0.95)(0.95)(0.95)(0.95)(0.95) = 0.77378$$

$$P(B_{[2]}) = P(F_1')P(F_2')P(F_3')P(F_4')P(F_5)$$

$$= (0.95)(0.95)(0.95)(0.95)(0.05) = 0.04073$$

$$P(B_{[3]}) = P(F_1')P(F_2')P(F_3')P(F_4)P(F_5')$$

$$= (0.95)(0.95)(0.95)(0.05)(0.95) = 0.04073$$

$$P(B_{[4]}) = P(F_1')P(F_2')P(F_3)P(F_4')P(F_5')$$

$$= (0.95)(0.95)(0.05)(0.95)(0.95) = 0.04073$$

$$P(B_{[5]}) = P(F_1')P(F_2)P(F_3')P(F_4')P(F_5')$$

$$= (0.95)(0.05)(0.95)(0.95)(0.95) = 0.04073$$

Therefore, the system reliability is the sum of the preceding five probabilities associated with the operating state of the system, which is

$$p_{s,\text{sys}} = 0.77378 + 4(0.04073) = 0.93668$$

7.4.2   Path enumeration method

This is a very powerful method for system reliability evaluation. A *path* is defined as a set of components or modes of operation that leads to a certain outcome of the system. In system reliability analysis, the system outcomes of interest are those of failed state or operational state. A *minimum path* is one in which no component is traversed more than once in going along the path. Under this methodologic category, *tie-set analysis* and *cut-set analysis* are two well-known techniques.

Cut-set analysis.   The *cut set* is defined as a set of system components or modes of operation that, when failed, cause the failure of the system. Cut-set analysis is powerful for evaluating system reliability for two reasons: (1) It can be programmed easily on digital computers for fast and efficient solutions of any general system configuration, especially in the form of a network, and (2) the cut sets are directly related to the modes of system failure and therefore identify the distinct and discrete ways in which a system may fail. For example, in a water distribution system, a cut set will be the set of system components including pipe sections, pumps, storage facilities, etc. that, when failed jointly, would disrupt the service to certain users.

The cut-set method uses the minimum cut sets for calculating the system failure probability. The *minimum cut set* is a set of system components that, when all failed, causes failure of the system but when any one component of the set does not fail does not cause system failure. A minimum cut set implies that all components of the cut set must be in the failure state to cause system failure. Therefore, the components or modes of operation involved in the minimum cut set are effectively connected in parallel, and each minimum cut set is connected in series. Consequently, the failure probability of a system can be expressed as

$$p_{f,\text{sys}} = P\left(\bigcup_{m=1}^{I} C_m\right) = P\left[\bigcup_{m=1}^{I}\left(\bigcap_{j=1}^{J_m} F_{mj}\right)\right] \tag{7.63}$$

in which C_m is the mth of the total I minimum cut sets, J_m is the total number of components or modes of operation in the mth minimum cut set, and F_{mj} represents the failure event associated with the jth components or mode of operation in the mth minimum cut set. In the case that the number of minimum

cut sets I is large, computing the bounds for probability of a union described in Sec. 7.2.3 can be applied. The bounds on the failure probability of the system should be examined for their closeness to ensure that adequate accuracy is obtained.

Example 7.13   Refer to the simple water distribution network shown in Fig. 7.9 in Example 7.12. Evaluate the system reliability using the minimum cut-set method.

Solution   Based on the system reliability as defined, the minimum cut sets for the example pipe network are

$$C_1 : F_1 \qquad C_2 : F_2 \cap F_3 \qquad C_3 : F_2 \cap F_4 \qquad C_4 : F_3 \cap F_5$$
$$C_5 : F_4 \cap F_5 \qquad C_6 : F_2 \cap F_5 \qquad C_7 : F_3 \cap F_4$$

where C_m is the mth cut set, and F_k is the failure state of pipe link k. The seven cut sets for the example network listed above are shown in Fig. 7.11. The system unreliability $p_{f,\text{sys}}$ is the probability of occurrence of the union of the cut set, that is,

$$p_{f,\text{sys}} = P\left(\bigcup_{m=1}^{7} C_m \right)$$

The system reliability can be obtained by subtracting $p_{f,\text{sys}}$ from 1. However, the computation, in general, will be very cumbersome for finding the probability of the
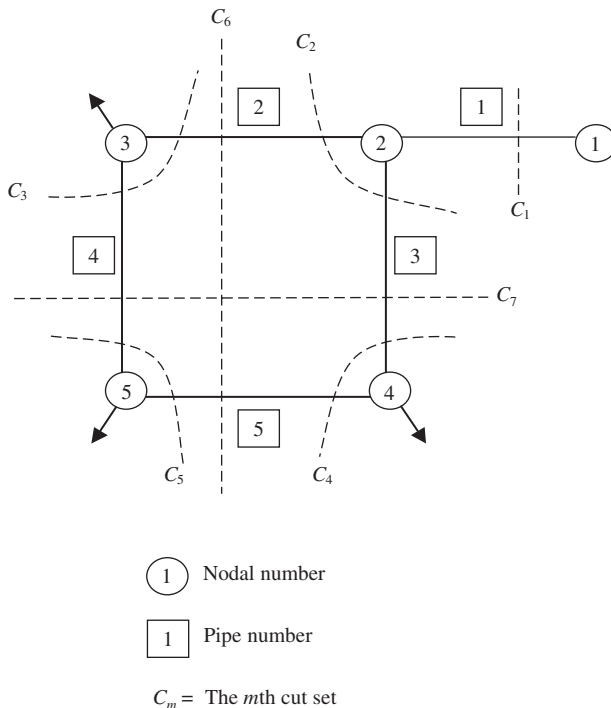
Figure 7.11   Example pipe network with cut sets for Example 7.13.

union of large numbers of events, even if they are independent. In this circumstance, it is computationally easier to compute the system reliability as

$$p_{s,\text{sys}} = 1 - P\left(\bigcup_{m=1}^{7} C_m\right) = P\left(\bigcap_{m=1}^{7} C'_m\right)$$

Since all the cut sets behave independently, all their complements also behave independently. The probability of the intersection of a number of independent events, according to Eq. (2.5), is

$$p_{s,\text{sys}} = P\left(\bigcap_{m=1}^{7} C'_m\right) = \prod_{m=1}^{7} P(C'_m)$$

where
$$P(C'_1) = 0.95 \qquad P(C'_2) = P(C'_3) = \cdots = P(C'_7) = 0.9975$$

Hence the system reliability of the example water distribution network is

$$p_{s,\text{sys}} = (0.95)(0.9975)^6 = 0.9360$$

Tie-set analysis. As the complement of a cut set, a *tie set* is a minimal path of the system in which system components or modes of operation are arranged in series. Consequently, a tie set fails if any of its components or modes of operation fail. All tie sets are effectively connected in parallel; that is, the system will be in the operating state if any of its tie sets are functioning. Therefore, the system reliability can be expressed as

$$p_{s,\text{sys}} = P\left(\bigcup_{m=1}^{I} T_m\right) = P\left[\bigcup_{m=1}^{I}\left(\bigcap_{j=1}^{J_m} F'_{mj}\right)\right] \qquad (7.64)$$

in which T_m is the mth tie set of all I tie sets, J_m is the total number of components or modes of operation in the mth tie set, and F'_{mj} represents the nonfailure state of the jth component in the mth tie set. Again, when the number of tie sets is large, computation of exact system reliability by Eq. (7.64) could be cumbersome. In such a condition, bounds for system reliability could be computed.

The main disadvantage of the tie-set method is that failure modes are not directly identified. Direct identification of failure modes is sometimes essential if a limited amount of a resource is available to focus on a few dominant failure modes.

Example 7.14 Refer to the simple water distribution network as shown in Fig. 7.9. Use tie-set analysis to evaluate the system reliability.

Solution The minimum tie sets (or path), based on the definition of system reliability given previously, for the example network are

$$T_1 : F'_1 \cap F'_2 \cap F'_4 \cap F'_5$$
$$T_2 : F'_1 \cap F'_3 \cap F'_4 \cap F'_5$$
$$T_3 : F'_1 \cap F'_2 \cap F'_3 \cap F'_4$$
$$T_4 : F'_1 \cap F'_2 \cap F'_3 \cap F'_5$$

where T_m is the mth minimum tie set, and F'_j is the nonfailure of the jth pipe link in the network. The four minimum tie sets are shown in Fig. 7.12. The system reliability, based on Eq. (7.64), is

$$
\begin{aligned}
p_{s,\text{sys}} &= P(T_1 \cup T_2 \cup T_3 \cup T_4) \\
&= [P(T_1) + P(T_2) + P(T_3) + P(T_4)] - [P(T_1, T_2) + P(T_1, T_3) \\
&\quad + P(T_1, T_4) + P(T_2, T_3) + P(T_2, T_4) + P(T_3, T_4)] \\
&\quad + [P(T_1, T_2, T_3) + P(T_1, T_2, T_4) + P(T_1, T_3, T_4) + P(T_2, T_3, T_4)] \\
&\quad - P(T_1, T_2, T_3, T_4)
\end{aligned}
$$

Since all pipes in the network behave independently, all minimum tie sets behave independently. In such circumstances, the probability of the joint occurrence of multiple independent events is simply equal to the multiplication of the probability of the individual events. That is,

$$
P(T_1) = P(F'_1)P(F'_2)P(F'_4)P(F'_5) = (0.95)^4 = 0.81451
$$
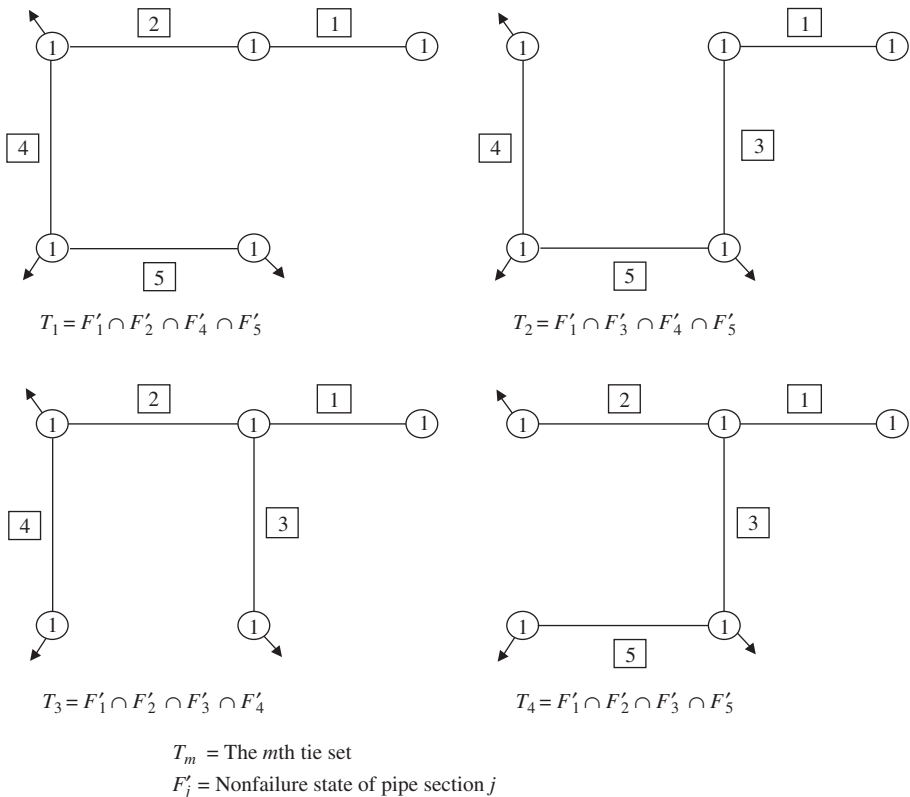
Figure 7.12 Example pipe network with tie sets for Example 7.14.

Similarly,

$$P(T_2) = P(T_3) = P(T_4) = 0.81451$$

Note that in this example the unions of more than two minimum tie sets are the intersections of the nonfailure state of all five pipe sections. For example, $T_1 \cup T_2 = (F_1' \cap F_2' \cap F_4' \cap F_5') \cup (F_1' \cap F_3' \cap F_4' \cap F_5') = (F_1' \cap F_2' \cap F_3' \cap F_4' \cap F_5')$. The system reliability can be reduced to

$$p_{s,\mathrm{sys}} = [P(T_1) + P(T_2) + P(T_3) + P(T_4)] - 3P(F_1' \cap F_2' \cap F_3' \cap F_4')$$

$$= 4(0.81451) - 3(0.95)^5$$

$$= 0.9367$$

In summary, the path enumeration method involves the following steps (Henley and Gandhi, 1975):

1. Find all minimum paths. In general, this has to be done with the aid of a computer when the number of components is large and the system configuration is complex.

2. Find all required unions of the paths.

3. Give each path union a reliability expression in terms of module reliability.

4. Compute the system reliability in terms of module reliabilities.

7.4.3 Conditional probability approach

This approach starts with a selection of key components and modes of operation whose states (operational or failure) would decompose the entire system into simple series and/or parallel subsystems for which the reliability or failure probability can be evaluated easily. Then the reliability of the entire system is obtained by combining those of the subsystems using the conditional probability rule as

$$p_{s,\mathrm{sys}} = p_{s \mid F_m'} \times p_{s,m} + p_{s \mid F_m} \times p_{f,m} \tag{7.65}$$

in which $p_{s \mid F_m'}$ and $p_{s \mid F_m}$ are the conditional system reliabilities given that the mth component is operational F_m' and failed F_m, respectively, and $p_{s,m}$ and $p_{f,m}$ are the reliability and failure probabilities of the mth component, respectively.

Except for very simple and small systems, a nested conditional probability operation is inevitable. Efficient evaluation of system reliability of a complex system hinges entirely on a proper selection of key components, which generally is a difficult task when the scale of the system is large. Furthermore, the method cannot be adapted easily to computerization for problem solving.

Example 7.15 Find the system reliability of the water distribution network in Fig. 7.9 using the conditional probability approach.

Solution Using the conditional probability approach for system reliability evaluation, first select pipe section 1 as the key element that decomposes the system into a simpler

configuration, as shown in Fig. 7.13. After the entire system is decomposed into a simple system configuration, the conditional probability of the decomposed systems can be evaluated easily. For example, the conditional system reliability, after imposing F_1' and F_3 for pipes 1 and 3, respectively, can be expressed as

$$p_{s,\text{sys}|F_1',F_3} = P(F_2' \cap F_4' \cap F_5') = (0.95)^3 = 0.8574$$

where $p_{s|F_1',F_3}$ is conditional system reliability. Conditional system reliabilities for other imposed conditions are shown in Fig. 7.13. After the conditional system reliabilities for the decomposed systems are calculated, the reliability of the entire
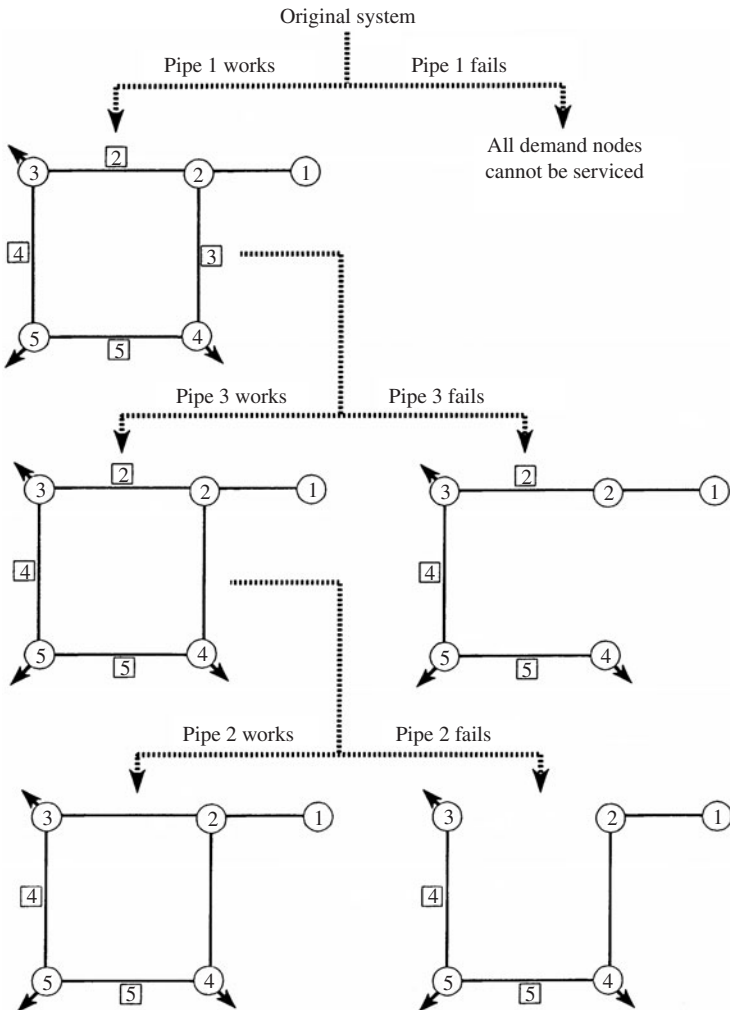
Figure 7.13 Example pipe network configurations under conditionings.

system can be combined using Eq. (7.65). For this particular example, the system reliability is

$$p_{s,\text{sys}} = p_{s \,|\, F'_1, F_3} \times P(F'_1, F_3) + p_{s \,|\, F'_1, F'_3, F'_2} \times P(F'_1, F'_3, F'_2)$$
$$+ \, p_{s \,|\, F'_1, F'_3, F_2} \times P(F'_1, F'_3, F_2)$$
$$= (0.8574)(0.95)(0.05) + (0.9975)(0.95)^3 + (0.9025)(0.95)^2(0.05)$$
$$= 0.9367$$

7.4.4 Fault-tree analysis

Conceptually, *fault-tree analysis*, unlike event-tree analysis, is a backward analysis that begins with a system failure and traces backward, searching for possible causes of the failure. Fault-tree analysis was initiated at Bell Telephone Laboratories and Boeing Aircraft Company (Barlow et al., 1975). Since then, it has been used for evaluating the reliability of many different engineering systems. In hydrosystems engineering designs, fault-tree analysis has been applied to evaluate the risk and reliability of earth dams, as shown in Fig. 7.14 (Cheng, 1982), underground water control systems (Bogardi et al., 1987), and water-retaining structures including dikes and sluice gates (Vrijling, 1987, 1993). Figure 7.15 shows a fault tree for the failure of a culvert as another example.

A fault tree is a logical diagram representing the consequence of the component failures (basic or primary failures) on the system failure (top failure or
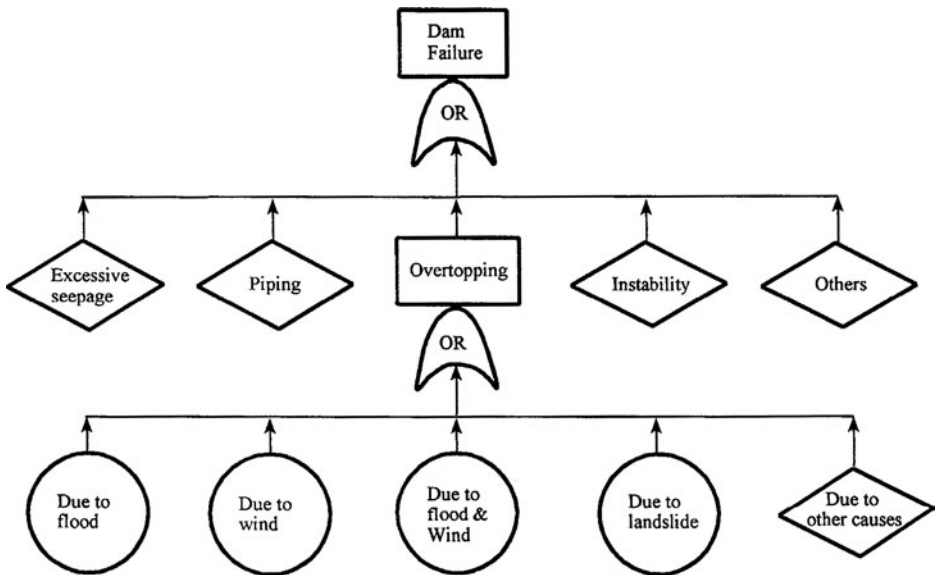
Figure 7.14 Simple fault tree for failure of existing dams. (*After Cheng, 1982.*)
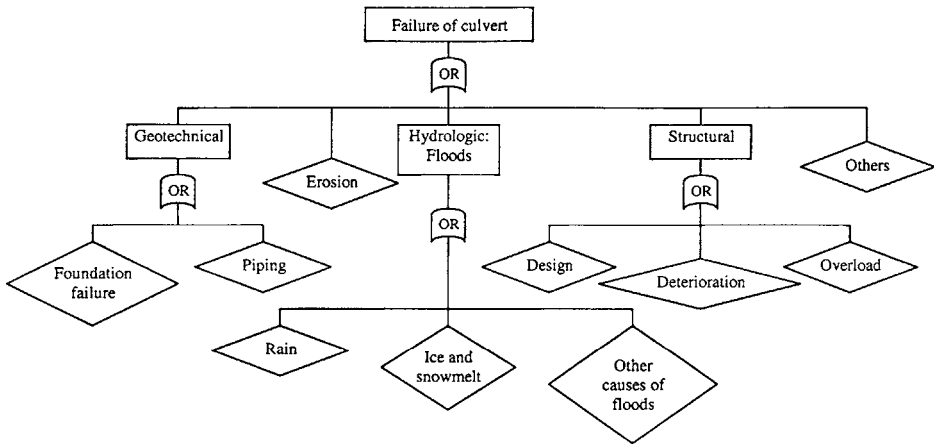
Figure 7.15 A fault tree showing the event of culvert failure.

top event). A simple fault tree is given in the Fig. 7.16a as an example. Two major types of combination nodes (or gates) are used in a fault tree. The *AND node* implies that the output event occurs only if all the input events occur simultaneously, corresponding to the intersection operation in probability theory. The *OR node* indicates that the output event occurs if any one or more of the input events occur, i.e., a union. The two and three other frequently used event notations are shown in Fig. 7.17. *Boolean algebra* operations are used in fault-tree analysis. Thus, for the fault tree shown in Fig. 7.16,

$$B_1 = C_1 \cap C_2 \qquad B_2 = C_3 \cup C_4 \cup C_1$$

Hence the top event is related to the component events as

$$T = B_1 \cup B_2 = (C_1 \cap C_2) \cup (C_3 \cup C_4 \cup C_1) = C_1 \cup C_3 \cup C_4$$

Thus the probability of the top event occurring can be expressed as

$$P(T) = P(C_1 \cup C_3 \cup C_4)$$

If C_1, C_3, and C_4 are mutually exclusive, then

$$P(T) = P(C_1) + P(C_3) + P(C_4)$$

Hence Fig. 7.16a can be reduced to an equivalent but simpler fault tree as Fig. 7.16b. System reliability $p_{s,\text{sys}}(t)$ is the probability that the top event does not occur over the time interval $(0, t]$.

Dhillon and Singh (1981) pointed out the advantages and disadvantages of the fault-tree analysis technique. Advantages include

1. It provides insight into the system behavior.

2. It requires engineers to understand the system thoroughly and deal specifically with one particular failure at a time.
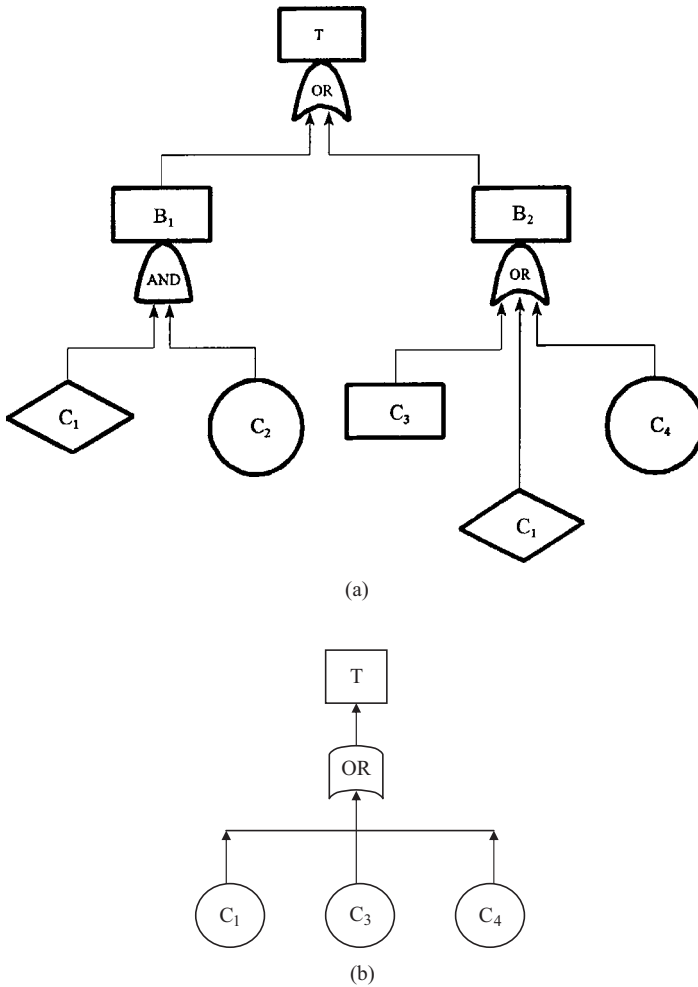
Figure 7.16 An example fault tree: (*a*) original fault tree before simplification; (*b*) reduced fault tree.

3. It helps to ferret out failures deductively.

4. It provides a visible and instructive tool to designers, users, and management to justify design changes and tradeoff studies.

5. It provides options to perform quantitative or qualitative reliability analysis.

6. The technique can handle complex systems.

7. Commercial codes are available to perform the analysis.

Disadvantages include

1. It can be costly and time-consuming.

2. Results can be difficult to check.
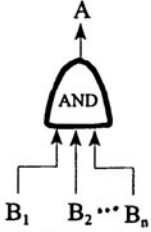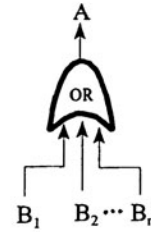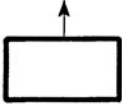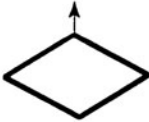
Symbol	Descriptions
Gate Symbols	
	AND node, intersection, output A exists if and only if all of $B_1, B_2, ..., B_n$ exist simultaneously.
	OR node, union, output A exists if any of $B_1, B_2, ..., B_n$, or any combination thereof exists.
Event Symbols	
	Identification of a paraticular event, when contained in the sequence, usually describes the output or input of and AND or OR node
	Basic event or condition, usually a malfunction, describable in terms of a specific component or cause.
	An event purposely not developed further because of lack of information or of insufficient consequence, could also be used to indicated further investigation when additional information becomes available.

Figure 7.17 Some basic node symbols used in fault-tree analysis.

3. The technique normally considers that the system components are in either working or failed state; therefore, the partial failure stats of components are difficult to handle.

4. Analytical solutions for fault trees containing standbys and repairable components are difficult to obtain for the general case.

5. To include all types of common failure causes requires considerable effort.

Fault-tree construction.   Before constructing a fault tree, engineers must thoroughly understand the system and its intended use. One must determine the higher-order functional events and continue the fault event analysis to determine their logical relationships with lower level events. Once this is accomplished, the fault-tree can be constructed. A brief description of fault-tree construction is given in the following paragraphs. The basic concepts of fault-tree analysis are presented in Henley and Kumamoto (1981) and Dhillon and Singh (1981).

The major objective of fault-tree construction is to represent the system condition that may cause system failure in a symbolic manner. In other words, the fault tree consists of sequences of events that lead to system failure. There are actually two types of building blocks: *gate symbols* and *event symbols*.
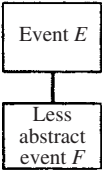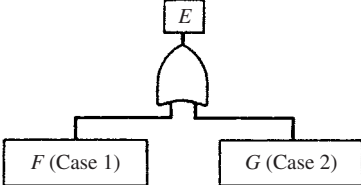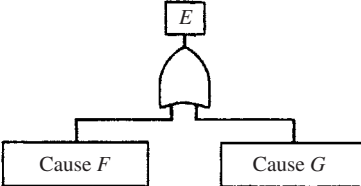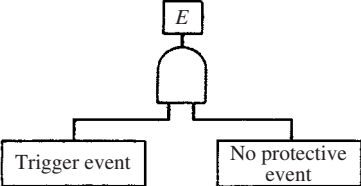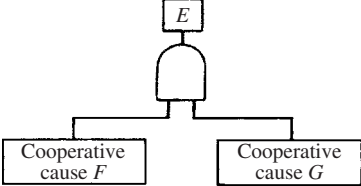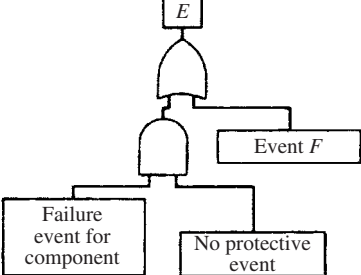
Gate symbols connect events according to their causal relation such that they may have one or more input events but only one output event. Figure 7.17 shows the two commonly used gate symbols and three types of commonly used event symbols. A *fault event*, denoted by a rectangular box, results from a combination of more basic faults acting through logic gates. A circle denotes a basic component failure that represents the limit of resolution of a fault tree. A diamond represents a fault event whose causes have not been fully developed. For more complete descriptions on other types of gate and event symbols, readers are referred to Henley and Kumamoto (1981).

Henley and Kumamoto (1981) presented heuristic guidelines for constructing fault trees, and these are summarized in Table 7.1 and Fig. 7.18 and are listed below:

1. Replace abstract events by less abstract events.

2. Classify an event into more elementary events.

3. Identify distinct causes for an event.

4. Couple trigger events with "no-protection actions."

5. Find cooperative causes for an event.

6. Pinpoint component failure events.

7. Develop component failure using Fig. 7.18.

Figure 7.19 shows a fault tree for the example pipe network of Fig. 7.9.

TABLE 7.1   Heuristic Guidelines for Gault-Tree Construction

	Development policy	Corresponding part of fault tree
1	Equivalent but less abstract event *F*	Event *E* → Less abstract event *F*
2	Classification of event *E*	*E* (OR gate): *F* (Case 1), *G* (Case 2)
3	Distinct causes for event *E*	*E* (OR gate): Cause *F*, Cause *G*
4	Trigger versus no protective event	*E* (AND gate): Trigger event, No protective event
5	Cooperative cause	*E* (AND gate): Cooperative cause *F*, Cooperative cause *G*
6	Pinpoint a component failure event	*E* (OR gate): [AND gate: Failure event for component, No protective event], Event *F*

Source: Henley and Kumomoto (1981).
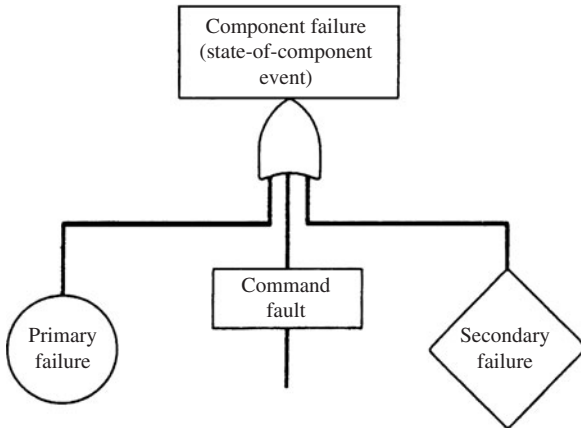
Figure 7.18 Development of component failure. (*Henley and Kumomoto, 1981.*)

Evaluation of fault trees. The basic steps used to evaluate fault trees include (1) construction of the fault tree, (2) determination of the minimal cut sets, (3) development of primary event information, (4) development of cut-set information, and (5) development of top event information.

To evaluate the fault tree, one always should start from the minimal cut sets that in essence, are *critical paths*. Basically, the fault-tree evaluation consists of two distinct processes: (1) determination of the logical combination of events
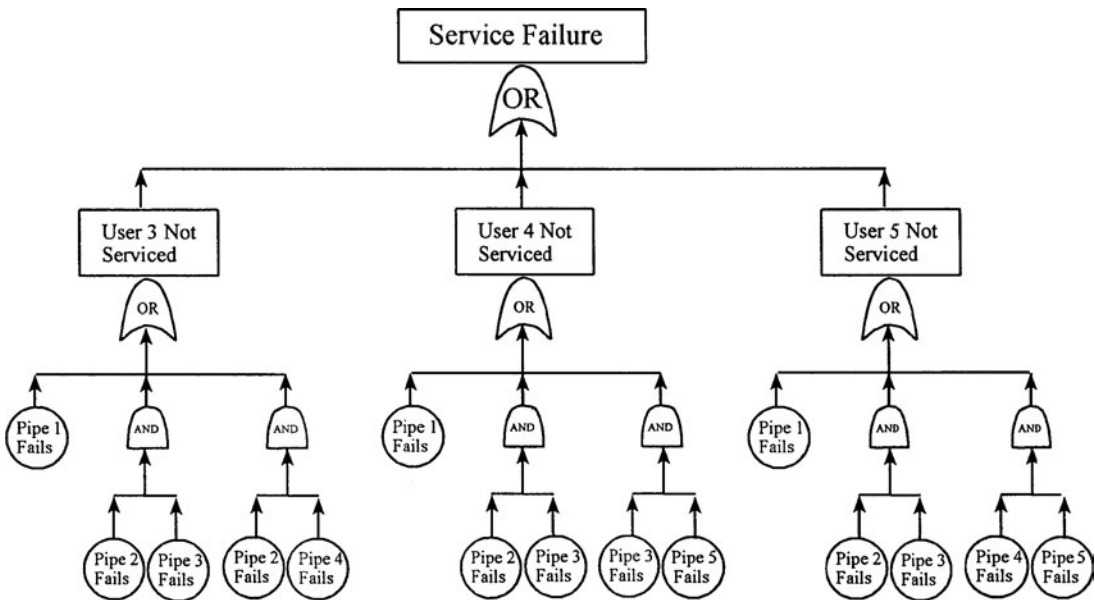
Figure 7.19 Fault tree for reliability analysis of example pipe network in Fig. 7.9.

that cause top event failure expressed in the minimal cut sets and (2) numerical evaluation of the expression.

Cut sets, as discussed previously, are collections of basic events such that if all these basic events occur, then the top event is guaranteed to occur. The tie set is a dual concept to the cut set in that it is a collection of basic events of which if none of the events in the tie set occur, then the top event is guaranteed not to occur. As one could imagine, a large system has an enormous number of failure modes. A minimal cut set is one that if any basic event is removed from the set, the remaining events collectively are no longer a cut set. By the use of minimum cut sets, the number of cut sets and basic events are reduced in order to simplify the analysis.

The system availability $A_{\mathrm{sys}}(t)$ is the probability that the top event does not occur at time t, which is the probability of the systems operating successfully when the top event is an OR combination of all system hazards. System unavailability $U_{\mathrm{sys}}(t)$, on the other hand, is the probability that the top event occurs at time t, which is either the probability of system failure or the probability of a particular system hazard at time t.

System reliability $p_{s,\mathrm{sys}}(t)$ is the probability that the top event does not occur over time interval $(0, t)$. System reliability requires continuation of the nonoccurrence of the top event, and its value is less than or equal to the availability. On other hand, the system unreliability, $p_{f,\mathrm{sys}}(t)$ is the probability that the top event occurs before time t and is complementary to the system reliability. Also, system unreliability, in general, is greater than or equal to system unavailability. From the system unreliability, the system failure density $f_{\mathrm{sys}}(t)$ can be obtained according to Eq. (5.2).

7.5   Summary and Conclusions

Hwang et al. (1981) presented a review of literature related to system reliability evaluation techniques for small to large complex systems. A large system was defined as one that has more than 10 components and a moderate system as one which has more than 6 components and less than 10. Complex systems were defined as ones that could not be reduced to a series-parallel system. Hwang et al. concluded that for a large, complex system, computer programs should be used that provide the minimum cut sets and calculate the minimal cut approximation to system reliability. Minimal paths can be generated from minimum cuts. Based on minimum cut sets, reliability approximations then can be obtained for large, complex networks. Hwang et al. also noted that Monte Carlo methods for system reliability evaluation can be used when component reliabilities are sampled by the Monte Carlo method. They also identified several miscellaneous approaches for evaluating complex systems, including a moment method, a block-diagram method, Bayesian decomposition, and decomposition by Boolean expression.

Hwang et al. (1981) concluded that of all the evaluation techniques in the papers surveyed, only a few had limited success in solving some large, complex

system reliability problems, and few techniques have been completely effective when applied to large system reliability problems. They suggested that a generally efficient graph partitioning technique for reliability evaluation of large, highly interconnected networks should be developed.

Since the 1981 paper by Hwang et al., several other system reliability evaluation techniques have been reported in the literature. Aggarwal et al. (1982) presented a method that uses decomposition of a probabilistic graph using cut sets. The method is applied to a simplified network with five nodes and seven links, and only limited computational results are presented.

Appendix 7A: Derivation of Bounds for Bivariate Normal Probability

Consider two performance functions $W_j(\boldsymbol{Z}') = 0$ and $W_m(\boldsymbol{Z}') = 0$ in a two-dimensional standardized, uncorrelated normal space whose design points are $\boldsymbol{z}'_{j*}$ and $\boldsymbol{z}'_{m*}$, respectively. At each of the design points, the first-order failure hyperplanes can be expressed as

$$W_j(\boldsymbol{Z}') = 0 \approx \sum_{k=1}^{2} \left(\frac{\partial W_j}{\partial Z'_k} \right) (Z'_k - z'_{k,j*}) = a_{0j} + a_{1j}Z'_1 + a_{2j}Z'_2 \qquad (7A.1)$$

$$W_m(\boldsymbol{Z}') = 0 \approx \sum_{k=1}^{2} \left(\frac{\partial W_m}{\partial Z'_k} \right) (Z'_k - z'_{k,m*}) = a_{0m} + a_{1m}Z'_1 + a_{2m}Z'_2 \qquad (7A.2)$$

in which $z_{k,m*}$ is the coordinate of the kth stochastic basic variable at the design point z'_{m*} of the mth performance function, and

$$a_{0m} = -\sum_{k=1}^{2} \left(\frac{\partial W_m}{\partial Z'_k} \right)_{\boldsymbol{z}'_{m*}} z'_{k,m*} \qquad a_{km} = \sum_{k=1}^{2} \left(\frac{\partial W_m}{\partial Z'_k} \right)_{\boldsymbol{z}'_{m*}} \qquad \text{for } k = 1, 2$$

The covariance between the two performance functions can be obtained as

$$\begin{aligned}
\text{Cov}[W_j(\boldsymbol{Z}'), W_m(\boldsymbol{Z}')] &= E\{[(a_{0j} + a_{1j}Z'_1 + a_{2j}Z'_2) - a_{0j}] \\
&\quad \times [(a_{0m} + a_{1m}Z'_1 + a_{2m}Z'_2) - a_{0m}]\} \\
&= E[(a_{1j}Z'_1 + a_{2j}Z'_2)(a_{1m}Z'_1 + a_{2m}Z'_2)] \\
&= a_{1j}a_{1m} + a_{2j}a_{2m} \qquad (7A.3)
\end{aligned}$$

Hence the correlation coefficient between the two performance functions at the design points is

$$\rho_{jm} = \frac{a_{1j}a_{1m} + a_{2j}a_{2m}}{\sqrt{a_{1j}^2 + a_{2j}^2}\sqrt{a_{1m}^2 + a_{2m}^2}} \qquad (7A.4)$$

This can be generalized to multidimensional problems involving M stochastic basic variables as

$$\rho_{jm} = \frac{\sum\limits_{k=1}^{K} a_{kj}a_{km}}{\sqrt{\sum\limits_{k=1}^{K} a_{kj}^2}\sqrt{\sum\limits_{k=1}^{K} a_{km}^2}} \tag{7A.5}$$

Note that the preceding correlation coefficient between the two performance functions ρ_{jm} is exactly equal to the inner product of the corresponding directional derivative vectors

$$\begin{aligned}
\rho_{jm} &= \left(\frac{a_{1j}}{\sqrt{a_{1j}^2 + a_{2j}^2}}, \frac{a_{2j}}{\sqrt{a_{1j}^2 + a_{2j}^2}}\right)^t \left(\frac{a_{1m}}{\sqrt{a_{1m}^2 + a_{2m}^2}}, \frac{a_{2m}}{\sqrt{a_{1m}^2 + a_{2m}^2}}\right) \\
&= (-\boldsymbol{\alpha}_{j*})^t(-\boldsymbol{\alpha}_{m*}) = (\boldsymbol{\alpha}_{j*})^t(\boldsymbol{\alpha}_{m*}) \\
&= |\boldsymbol{\alpha}_{j*}^t||\boldsymbol{\alpha}_{m*}|\cos\theta \\
&= \cos\theta
\end{aligned} \tag{7A.6}$$

in which θ is the angle between the directional derivatives of the two performance functions. Hence, if the two performance functions are positively correlated, the angle θ between $\boldsymbol{\alpha}_{m*}$ and $\boldsymbol{\alpha}_{j*}$ lies in the range $0° \leq \theta \leq 90°$. On the other hand, negative correlation between $W_j(\mathbf{Z})$ and $W_m(\mathbf{Z})$ corresponds to the range $90° < \theta \leq 180°$. Plots for positively and negatively correlated performance functions are show in Figs. 7A.1 and 7A.2, respectively.

When $W_j(\mathbf{Z})$ and $W_m(\mathbf{Z})$ are positively correlated, that is, $\rho_{jm} > 0$, referring to Fig. 7A.1, the shaded area representing the joint failure of the two performance functions satisfies the following relationships:

$$(F_j, F_m) \supset A \qquad \text{and} \qquad (F_j, F_m) \supset B$$

in which (F_j, F_m) represents the joint failure events of the two performance functions, and sets A and B are defined in Fig. 7A.1.

Again, referring to Fig. 7A.1, the following relationship holds:

$$\text{Max}[P(A), P(B)] \leq P(F_j, F_m) \leq P(A) + P(B) \tag{7A.7}$$

By orthogonality, one has

$$P(A) = \Phi(-\beta_m)\Phi(\beta_{j\,|\,m}) \tag{7A.8}$$

$$P(B) = \Phi(\beta_j)\Phi(\beta_{m\,|\,j}) \tag{7A.9}$$

where
$$\beta_{j\,|\,m} = \frac{\beta_j - \rho_{jm}\beta_m}{\sqrt{1 - \rho_{jm}^2}} \qquad \beta_{m\,|\,j} = \frac{\beta_m - \rho_{mj}\beta_j}{\sqrt{1 - \rho_{mj}^2}} \tag{7A.10}$$

which are defined in Fig. 7A.1.
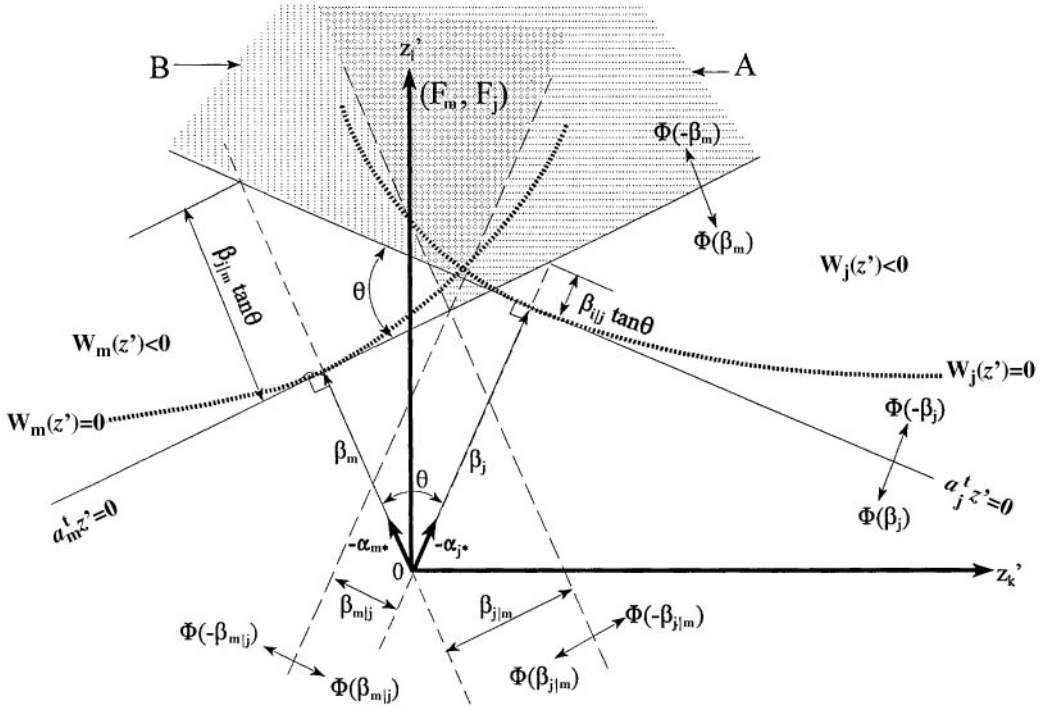
Figure 7A.1 Two intersecting tangent planes with positively correlated failure events. (*Ang and Tang, 1984.*)
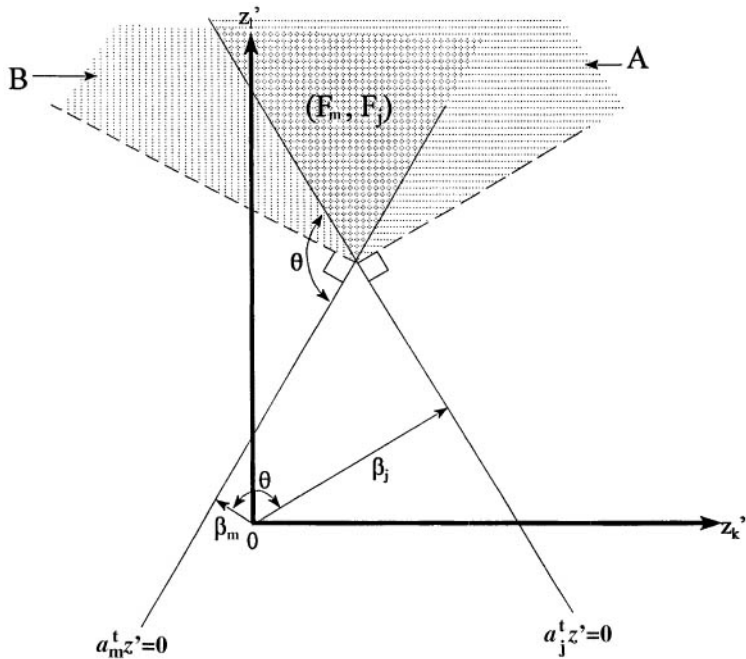
Figure 7A.2 Two intersecting tangent planes with negatively correlated failure events. (*Ang and Tang, 1984.*)

Referring to Fig. 7A.2 for negatively correlated $W_j(\mathbf{Z})$ and $W_m(\mathbf{Z})$, it can be observed that

$$(F_j, F_m) \subset A \qquad \text{and} \qquad (F_j, F_m) \subset B$$

resulting in

$$0 \leq P(F_j, F_m) \leq \min[P(A), P(B)] \tag{7A.11}$$

with $P(A)$ and $P(B)$ given in Eqs.(7A.8) and (7A.9), respectively.

Problems

7.1 Derive the expression of system reliability for the system configurations shown in Fig. 7P.1 under the condition of (a) all units are dependent and different and (b) all units are independent and identical.
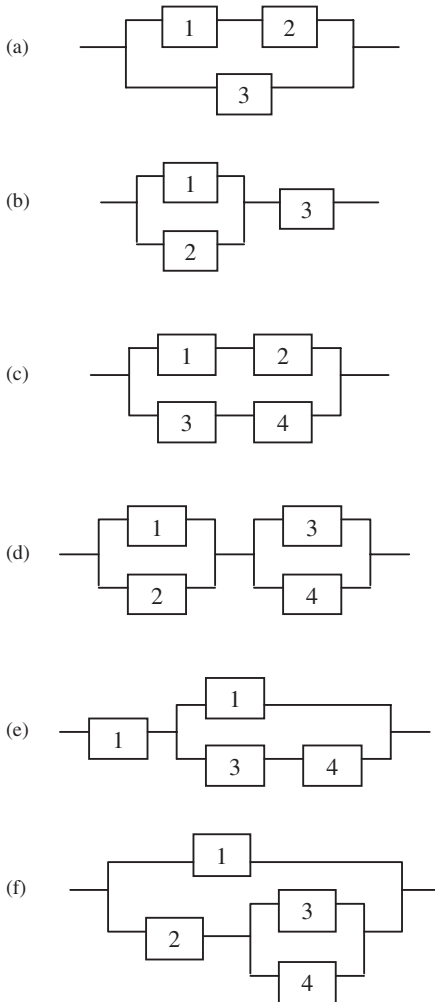
Figure 7P.1 Configuration of various systems for Problem 7.1.

7.2   Consider that two independent, identical units are to be added to an existing unit that would result in three units in the whole system.

(a)   Sketch all possible system configurations according to the arrangement of the three units.

(b)   Rank your system configurations according to the system reliability.

7.3   Consider the two system configurations shown in Fig. 7P.2. Use the cut-set method to determine the system reliability. Assume that all system components are identical and behave independently of each other.

7.4   Consider a hypothetical water distribution network consisting of two loops, as shown in Fig. 7P.3. Let's say that the service failure of the system is when at least one demand node cannot receive water. (a) Construct a tree diagram indicating failure cases for the water distribution network. (b) Determine the system reliability if all pipe sections behave independently, and each has a breakage probability of 0.03.

7.5   Resolve Problem 7.4 by cut-set analysis.

7.6   Resolve Problem 7.4 by tie-set analysis

7.7   Resolve Problem 7.4 by the conditional probability approach.

7.8   A detention basin is designed to accommodate excessive surface runoff temporarily during storm events. The detention basin should not overflow, if possible, to prevent potential pollution of the stream or other receiving water bodies. For simplicity, the amount of daily rainfall is categorized as heavy, moderate, and light (including none). With the present storage capacity, the detention basin is capable of accommodating runoff generated by two consecutive days of heavy rainfall or three consecutive days of at least moderate rainfall. The daily rainfall amounts around the detention basin site are not entirely independent. In other
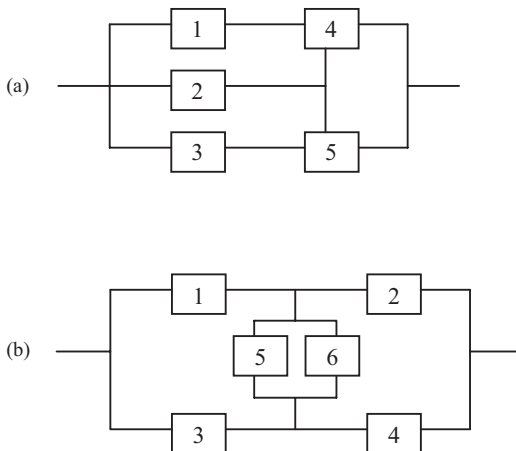
Figure 7P.2   System configurations for Problem 7.3.
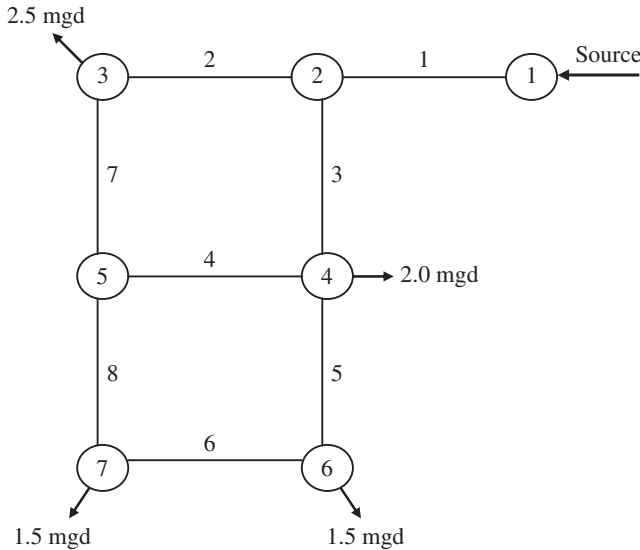
Figure 7.3 Hypothetical water distribution network.

words, the amount of daily rainfall on a given day would affect the daily rainfall amount on the next day. Let random variable X_t represent the amount of rainfall on any day t. The transition probability matrix, indicating the conditional probability of rainfall amount in a given day t conditioned on the rainfall amount of the previous day $t-1$, is shown in the following table (after Mays and Tung 1992).

		X_{t+1}		
		Heavy (H)	Moderate (M)	Light (L)
X_t	Heavy (H)	0.3	0.5	0.2
	Moderate (M)	0.3	0.4	0.3
	Light (L)	0.1	0.3	0.6

Construct a tree diagram to determine, for a given day on which the amount of rainfall is light, the probability that the detention basin will overflow in the next three days.

References

Ang, A. H-S., and Tang, W. H. (1984). *Probability Concepts in Engineering Planning and Design,* Vol. II: *Decision, Risk, and Reliability*, John Wiley & Sons, New York.

Aggarwal, K. K., Chopra, Y. C., and Bajwa, J. S. (1982). Reliability evaluation by network decomposition, IEEE, *Transactions on Reliability*, R-3(4):355–358.

Barlow, R. E., Fussell, J. B., and Singpurwalla, N. D., eds. (1975). *Reliability and Fault Tree Analysis*, SIAM, Philadephia, MD.

Bennett, R. M., and Ang, A. H. S. (1983). Investigation of methods for structural system reliability, Structural Research Series No. 510, University of Illinois, Urbana.

Bogardi, I., Duckstein, L., and Szidarovszky, F. (1987). Reliability estimation of underground water control systems under natural and sample uncertainty, in *Engineering Reliability and Risk in Water Resources*, ed. by L. Duckstein and E. J. Plate, Martinus Nijhoff, Dordrecht, The Netherlands, pp. 115–134.

Cheng, S. T. (1982). Overtopping risk evaluation of an existing dam, Ph.D. thesis, Department of Civil Engineering, University of Illinois, Urbana-Champaign.

Dawson, D. A., and Sankoff, D. (1967). An inequality for probabilities, *Proceedings of the American Mathematics Society* 18:504–507.

Dhillon, B. S., and Singh, C. (1981). *Engineering Reliability: New Techniques and Applications*, John Wiley & Sons, New York.

Ditlevsen, O. (1979). Narrow reliability bounds for structural systems, *Journal of Structural Mechanics* 7(4):435–451.

Henley, E. J., and Gandhi, S. L. (1975). Process reliability analysis, *American Institute of Chemical Engineering Journal* 21(4):677–686.

Henley, E. J., and Kumamoto, H. (1981). *Reliability Engineering and Risk Assessment*, Prentice-Hall, Englewood Cliffs, NJ.

Hunter, D. (1977). Approximate percentage points of statistics expressible as maxima, *Management Science* 7:25–36.

Hwang, C. L., Tillman, F. A., and Lee, M. H. (1981). System reliability evaluation techniques for complex/large systems: A review, IEEE, *Transactions on Reliability* R-30(5):416–422.

Kounias, E. G. (1968). Bounds for the probability of a union, with applications, *American Mathematical Statistics* 39(6):2154–2158.

Mays, L. W., and Tung, Y. K. (1992). *Hydrosystems Engineering and Management*, McGraw-Hill, New York.

Melchers, R. E. (1999). *Structural Reliability Analysis and Prediction*, John Wiley and Sons, New York.

Ramachandran, K. (1984). Systems bounds: A critical study, *Civil Engineering Systems* 1:123–128.

Tung, Y. K., and Yen, B. C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, McGraw-Hill, New York.

Virjling, J. K. (1987). Probabistic design of water retaining structures, in *Engineering Reliability and Risk in Water Resources*, ed. by L. Duckstein and E. J. Plate, Martinus Nijhoff, Dordrecht, The Netherlands, pp. 115–134.

Vrijling, J. K. (1993). Development in probabilistic design of flood defenses in the Netherlands, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. by B. C. and Y. K. Tung, ASCE, New York, pp. 133–178.

This page intentionally left blank

Integration of Reliability in Optimal Hydrosystems Design

8.1 Introduction

All hydrosystems engineering problems involve many interconnected and interrelated components. The analysis of any hydrosystem problem should take those interactions into account so that the overall behavior of the system is modeled properly. In general, problems in hydrosystems engineering can be classified into (1) development problems, (2) design problems, and (3) operational problems (Buras, 1972). In fact, practically all hydrosystems engineering problems encompass these problem types, which involve activities relating to determination of (1) the optimal scale of development of the project, (2) the optimal dimensions of the various components of the system, and (3) the optimal operation of the system.

Frequently, design and analysis of hydrosystems involve the use of models. The primary objectives of modeling exercises are (1) to analyze the behavior of existing systems so as to improve their performance and (2) to identify the "best" structural components and configurations of a system under planning. As discussed in Chap. 1, owing to the existence of various uncertainties in hydrosystems modeling, one cannot be certain that the best solution obtained is indeed truly optimal. The conventional approach when facing uncertainties in engineering design is to conduct sensitivity analysis, by which the influences of variation in model parameters subject to uncertainty on the system responses are assessed quantitatively. Simple sensitivity analyses often are ineffective in providing design, management, or operational guidance because when the various system parameters are changed systematically in sensitivity analysis, no consideration is given to whether the changed values are likely or realistic. It is therefore the objective of this chapter to present some practical approaches that integrate the uncertainties and reliability in an optimization framework for hydrosystems design, management, and operation.

This chapter starts with a brief description of the concepts of some frequently used optimization techniques in hydrosystems engineering design, management, and operation. More detailed descriptions of the various optimization techniques are given by Mays and Tung (1992), along with several specialized textbooks on the different subject matters. In Sec. 8.2, focus is placed on some typical problems in the context of resource allocation to optimize system reliability. Then the concept of risk-based design is described in Sec. 8.3, followed by an example application to hydrosystems engineering in Sec. 8.4. The last two sections, Secs. 8.5 and 8.6, describe a simple way to solve an optimization model in which the parameters are subject to uncertainty.

8.1.1 General framework of optimization models

Optimization models possess algorithms to compare the measures of effectiveness of a system and attempt to yield the optimal solution having the most desirable value of the adopted measures. In other words, an optimization model applies an optimum-seeking algorithm, which enables the search of all alternative solutions to select the best one. The general class of such optimum-seeking algorithms is called *mathematical programming*, which includes linear programming, nonlinear programming, dynamic programming, etc.

The main advantage of optimization models is that the optimal solution to a multidimensional (or multivariate) problem can be found readily by using an efficient search algorithm. The limitation, generally dictated by the solution technique available for model solving, is that sometimes drastic simplifications of real-life systems are inevitable in order to make the model mathematically tractable. Consequently, it is important to recognize that the optimal solution so derived is for a rather restricted case; that is, the solution is optimal to the simplified problem, and it is the optimal solution to the real-life problem of interest only to the extent that the simplifications are not damaging.

All optimization models consist of three basic elements: (1) decision variables and parameters, (2) constraints, and (3) objective functions. *Decision variables* are those unknown variables which are to be determined from the solution of the model, whereas *parameters* describe the characteristics of the system. Decision variables generally are controllable, whereas model parameters may or may not be controllable in real-life problems. *Constraints* describe the physical, economical, legal, and other limitations of the system expressed in mathematical form. The constraints must be included in the model, if they exist, to limit the decision variables to their feasible values. *Objective functions* are measures of the effectiveness of system performance and are also expressed as mathematical functions of decision variables. In a mathematical programming model, the value of the objective function is to be optimized.

The general form of an optimization model can be expressed as

$$\text{Optimize} \quad x_o = f(x_1, x_2, \ldots, x_n) \tag{8.1a}$$

Subject to

$$g_i(x_1, x_2, \ldots, x_n) = 0 \qquad i = 1, 2, \ldots, m \tag{8.1b}$$

$$a_j \leq x_j \leq b_j \qquad j = 1, 2, \ldots, n \tag{8.1c}$$

where $f(\cdot)$ and $g(\cdot)$ are, respectively, the general expressions for the objective function and constraint equations, which can be linear or nonlinear. The constraint set by Eq. (8.1c) is called the *bounding constraint*, indicating that the decision variables x_j's are bounded by their respective lower bound a_j and upper bound b_j. The most commonly seen bounding constraint type is the *nonnegativity constraint*, with the lower bound being 0 and upper bound being infinity.

In the preceding formulation, the decision variables x_j's in the problem generally are controllable inputs. The solution to the problem consists a set of decision variables in the system, each of which has a particular value. A solution can be feasible or infeasible. A *feasible solution* to an optimization problem is the one that satisfies all system constraints simultaneously. That is, a feasible solution is an element in the *feasible solution space* defined by the constraint equations. The solution to an optimization problem is the one in the feasible solution space that yields the most desirable objective function value, which is called the *optimum feasible solution* or simply the *optimum solution*.

The feasible solution space to an optimization model can be classified as either *convex* or *nonconvex*. Schematic sketches of convex and nonconvex feasible solution spaces are shown in Fig. 8.1. The nature of convexity of the feasible region of an optimization problem would dictate whether the optimal solution obtained is a global optimum or a local optimum. A *global optimum* is the best solution to the problem within the entire feasible space, whereas a *local optimum* is the best solution to the problem in the neighborhood of the solution point.

8.1.2   Single-objective versus multiobjective programming

The optimization model defined by Eqs. (8.1a–c) is for a single-objective problem. On the other hand, *multiobjective programming* deals with problems involving several noncommensurable and often conflicting objectives simultaneously. Among the objective functions involved, no single one has an importance that is overwhelmingly dominant over all others. Under this circumstance, the ideological theme of optimality in the single-objective context is no longer appropriate.

Mathematically, a multiobjective programming problem can be expressed in terms of *vector optimization* as

$$\text{Maximize} \quad \boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x})] \tag{8.2a}$$

$$\text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0} \tag{8.2b}$$

in which $\boldsymbol{f}(\boldsymbol{x})$ is a vector of M objective functions, $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ are vector of constraints, and $\boldsymbol{x}$ is a vector of decision variables.

The solution to a multiobjective programming problem is a *best compromisable solution* according to the decisionmaker's preference among the objectives and the noninferior solutions to the problem. A *noninferior solution* to a multiobjective programming problem is a feasible solution to which there is no other

(a)

(b)
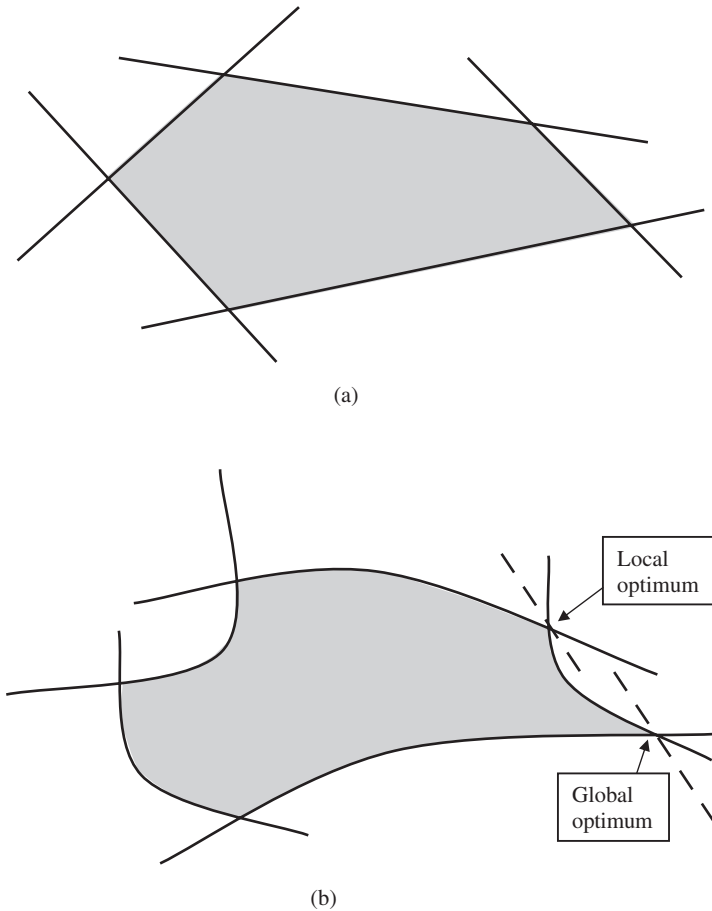
Figure 8.1 Schematic diagrams of (*a*) convex and (*b*) non-convex spaces.

feasible solution that will yield an improvement in one objective without causing degradation to at least one other objective (Cohon, 1978). The collection of such noninferior solutions allows the assessment of tradeoff among conflicting objectives.

To obtain the solution to a multiobjective programming problem, the preference of the decisionmaker among the conflicting objectives must be known. Information concerning the decisionmaker's preference is commonly called the *utility function*, which is a function of the objective function values. Geometrically, the utility function can be depicted as a series of *indifference curves* (Fig. 8.2). The utility of a decision maker will be the same for a combination of solutions that fall on the same indifference curve. The best compromise solution to a multiobjective programming problem is a unique set of alternatives that possesses the property of maximizing the decisionmaker's utility, and the alternatives are elements of the noninferior solution set.
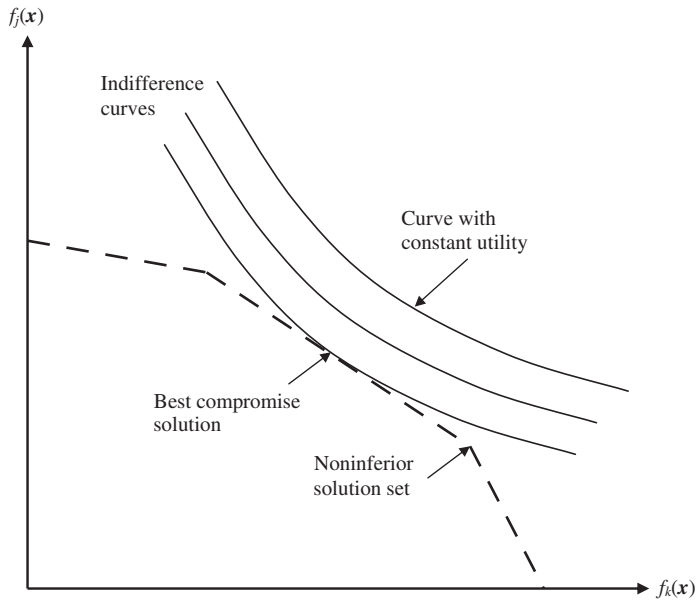
Figure 8.2 Noninferior solution and indifference curves.

Many methods have been proposed to derive the solution to a multiobjective programming problem. Basically, they can be classified into two categories (Cohon, 1978): *generating techniques* and *techniques incorporating knowledge of the decisionmaker's preference*. The purpose of generating techniques is to provide decisionmakers with information about the noninferior solution set or feasible tradeoff to the multiobjective problem. System analysts play the role of information providers and do not actively engage in decision making. On the other hand, techniques in the other category explicitly incorporate the decisionmaker's preference to search for the best compromise solution. Detailed descriptions of various techniques for solving multiobjective problems can be found elsewhere (Cohon, 1978; Goicoechea et al., 1982; Chankong and Haimes, 1983; Steuer, 1986). Regardless of the approach to be used for solving a multiobjective programming problem, the basic solution technique is still one of single-objective optimization algorithms.

8.1.3 Optimization techniques

Having constructed an optimization model, one must choose an optimization technique to solve the model. The way that one solves the problem depends largely on the form of the objective function and constraints, the nature and number of variables, the kind of computational facility available, and personal taste and experiences. Mays and Tung (1992) provide a rather detailed discussion on linear programming, dynamic programming, and nonlinear programming techniques. In this subsection, brief descriptions are given to provide

readers with some background about these techniques. A list of applications of various optimization techniques to hydrosystems engineering problems can be found elsewhere (Mays and Tung, 2005).

Linear programming (LP). LP has been applied extensively to optimal resource allocation problems. When the system under study is linear, LP models also can be used for parameter identification and system operation. By the name of the technique, an LP model has the basic characteristic that both the objective function and constraints are linear functions of the decision variables. The general form of an LP model can be expressed as

$$\text{Max (or min)} \quad x_o = \sum_{j=1}^{n} c_j x_j \tag{8.3a}$$

subject to

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = 1, 2, \ldots, m \tag{8.3b}$$

$$x_j \geq 0 \quad j = 1, 2, \ldots, n \tag{8.3c}$$

where the c_j's are the *objective function coefficients*, the a_{ij}'s are called the *technological coefficients*, and the b_i's are the right-hand-side (RHS) coefficients. An LP model can be expressed concisely in matrix form as

$$\text{Max (or min)} \quad x_o = \boldsymbol{c}^t \boldsymbol{x} \tag{8.4a}$$

$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{8.4b}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{8.4c}$$

where $\boldsymbol{c}$ is an $n \times 1$ column vector of objective function coefficients, $\boldsymbol{x}$ is an $n \times 1$ column vector of decision variables, $\boldsymbol{A}$ is an $m \times n$ matrix of technological coefficients, $\boldsymbol{b}$ is an $m \times 1$ column vector of the RHS coefficients, and the superscript t represents the transpose of a matrix or vector. Excellent books on LP include Winston (1987), Taha (1987), and Hillier and Lieberman (1990).

The commonly used algorithm for solving an LP model is the *simplex method* developed by Dantzig (1963). Since its conception, the method has been used widely and considered as the most efficient approach for solving LP problems. The simplex algorithm, starting at an initial feasible solution to the LP model, searches along the boundary of a problem's feasible space to reach the optimal solution. The method has been applied to solve large problems involving thousands of decision variables and constraints on today's computers. Computer codes based on the simplex method are widely available. Some of the well-known LP software includes GAMS (general algebraic modeling system) by Brooke et al. (1988) and LINDO (linear, interactive, and discrete optimizer) by Schrage (1986), for which PC versions of the programs are available. For LP

models, because of the convexity of the feasible region, the solution obtained is a global optimum.

Two additional methods for solving LP problems have been developed that apply different algorithmic strategies (Khatchian, 1979; Karmarkar, 1984). In contrast to the simplex algorithm, *Khatchian's ellipsoid method* and *Karmarkar's projective scaling method* seek the optimum solution to an LP problem by moving through the interior of the feasible region.

Nonlinear programming (NLP).   A *nonlinear programming* model has the general format of Eqs. (8.1a–c) in which either the objective function $f(\boldsymbol{x})$ or the constraints $\boldsymbol{g}(\boldsymbol{x})$ or both are nonlinear. In an NLP problem, the convexity of the feasible space defined by the nonlinear constraints $\boldsymbol{g}(\boldsymbol{x})$ is generally difficult to assess. As a result, the solution obtained by any NLP algorithm cannot guarantee to be globally optimal. Detailed description for solving NLP problems can be found in Edgar and Himmelblau (1988), Fletcher (1980), McCormick (1983), and Gill et al. (1981).

Basically, algorithms for solving NLP problems are divided into two categories: unconstrained and constrained algorithms. *Unconstrained NLP algorithms* solve Eq. (8.1a) without considering the presence of constraints. They provide the building blocks for the more sophisticated constrained NLP algorithms. Consider an unconstrained NLP problem

$$\text{Minimize } f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n \tag{8.5}$$

in which $\mathbb{R}^n$ is an n-dimensional real space. Assume that $f(\boldsymbol{x})$ is twice differentiable; the necessary conditions for $\boldsymbol{x}^*$ to be a solution to Eq. (8.5) are (1) $\nabla_x f(\boldsymbol{x}^*) = \boldsymbol{0}$ and (2) $\nabla_x^2 f(\boldsymbol{x}^*) = \boldsymbol{H}(\boldsymbol{x}^*)$ is semipositive definite in which $\nabla_x f = (\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_n)^t$, a *gradient vector* of the objective function and $\nabla_x^2 f = (\partial^2 f/\partial x_i \partial x_j)$ is an $n \times n$ *Hessian matrix*. The sufficient conditions for an unconstrained minimum $\boldsymbol{x}^*$ are (1) $\nabla_x f(\boldsymbol{x}^*) = \boldsymbol{0}$ and (2) $\nabla_x f(\boldsymbol{x}^*) = \boldsymbol{H}(\boldsymbol{x}^*)$ is strictly positive definite.

In theory, the solution to Eq. (8.5) can be obtained by solving $\nabla_x f(\boldsymbol{x}^*) = \boldsymbol{0}$, which involves a system of n nonlinear equations with n unknowns. This approach has been regarded as indirect because it backs away from solving the original problem of minimizing $f(\boldsymbol{x})$. Furthermore, numerical iterative procedures are required to solve the system of nonlinear equations which tend to be computationally inefficient. Therefore, the general preference is given to those solution procedures which directly attack the problem of optimizing the objective function. Like the LP solution procedure, *direct solution methods*, during the course of iteration, generate a sequence of points in the solution space that converge to the solution of the problem by the following recursive equation:

$$\boldsymbol{x}^{(r+1)} = \boldsymbol{x}^{(r)} + \beta^{(r)} \boldsymbol{d}^{(r)} \qquad r = 1, 2, \ldots \tag{8.6}$$

in which the superscript (r) represents the iteration number, $\boldsymbol{x}$ is the vector of the solution point, $\boldsymbol{d}$ is the vector of the *search direction* along which the

objective function $f(\pmb{x})$ is to be minimized, and β is a scalar, called the *step size*, that minimizes $f(\pmb{x}^{(r)} + \beta^{(r)}\pmb{d}^{(r)})$. This procedure is called the *line search* or *one-dimensional search*. Several unconstrained NLP algorithms have been developed, and they differ by the way the search directions are determined during the course of iteration. Mays and Tung (1992, p. 136) summarize the search directions of various methods.

Without losing generality, consider a nonlinear constrained problem stated by Eq. (8.1) with no bounding constraints. Note that the constraints Eq. (8.1b) are all equality constraints. Under this condition, the *Lagrangian multiplier method* can be applied, which converts a constrained NLP problem to an unconstrainted one by an augmented objective function called the *Lagragian*. That is, Eqs. (8.1a–b) can be written as

$$\text{Minimize } L(\pmb{x}, \lambda) = f(\pmb{x}) + \pmb{\lambda}^t \pmb{g}(\pmb{x}) \tag{8.7}$$

in which $L(\pmb{x}, \pmb{\lambda})$ is the *Lagrangian function*, $\pmb{\lambda}$ is the vector of m Lagragian multipliers, and $\pmb{g}(\pmb{x})$ is the vector of constraint equations. The new objective function $L(\pmb{x}, \pmb{\lambda})$ involves $n + m$ decision variables. The necessary condition and sufficient conditions for $\pmb{x}^*$ to be the solution for Eq. (8.7) are

1. $f(\pmb{x}^*)$ is convex, and $\pmb{g}(\pmb{x}^*)$ is convex in the vicinity of $\pmb{x}^*$.
2. $\nabla_x L(\pmb{x}^*, \pmb{\lambda}) = \pmb{0}$.
3. $\nabla_\lambda L(\pmb{x}^*, \pmb{\lambda}) = \pmb{g}(\pmb{x}^*) = \pmb{0}$.
4. λ's are unrestricted in sign.

Solving conditions 2 and 3 simultaneously yields the optimal solution to Eq. (8.7).

The most important theoretical results for the NLP problem are the *Kuhn-Tucker conditions*, which can be derived easily by using the Lagrangian method for the general NLP problem, as stated in Eq. (8.1a–c). These conditions must be satisfied at any constrained optimum, local or global, of any LP or NLP problem. They form the basis for the development of many computational algorithms.

Several NLP computer codes are available commercially. They are the GRG2 (generalized reduced gradient 2) developed by Lasdon and his colleagues (Lasdon et al., 1978; Lasdon and Waren, 1978); (2) GINO (Liebman et al., 1986); (3) MINOS (modular in-core nonlinear optimization system) by Mautaugh and Saunders (1980, 1983), and (4) GAMS-MINOS, a link for GAMS and MINOS. Microsoft Excel SOLVER implements GINO in a spreadsheet working environment.

Dynamic programming (DP).   Before performing the optimization, it is sometimes desirable to make some changes of variables and transformations so that the model can be solved more efficiently. However, one must keep in mind that such transformations should completely preserve the properties of the original problem (model) so that the transformed model will yield the optimal solution

to the original problem. Basically, DP is such a transformation that takes a sequential or multistage decision process containing many interrelated decision variables and converts it into a series of single-stage problems, each containing only one or a few variables. In other words, the DP technique decomposes an n-decision problem into a sequence of n separate but interrelated single-decision subproblems. Books that deal with DP are Dreyfus and Law (1977), Cooper and Cooper (1981), and Denardo (1982).

To describe the general philosophy of the DP technique, consider the following resource allocation problem (Tung and Mays, 1992). Suppose that one wishes to allocate funds to three water development projects, A, B, and C, to maximize the total revenue. Each development project consists of several alternative configurations that require different funding levels and yield different revenues. Owing to the budget limitation, the total available funds for the entire development are fixed. If the number of alternatives for each project is not too large, one probably can afford to enumerate all possible combinations of project alternatives exhaustively to identify the optimal alternatives for the entire project development. This brute-force exhaustive enumeration approach possesses three main shortcomings: (1) It would become impractical if the number of alternative combinations is large, (2) the optimal course of action cannot be verified, even it is obtained in the early computations, until all the combinations are examined, and (3) infeasible combinations (or solutions) cannot be eliminated in advance.

Using the DP technique, one considers the selection of alternatives within each project individually without ignoring the interdependence among the projects through the total available budget. Since the total funds are limited, the amount available to each project depends on the allocations to the other projects. Whatever funds are given to project A and project B, the allocation to the remaining project C must be made to optimize its return with respect to the remaining capital. In other words, the optimal allocation to project C is conditioned only on the available funds for project C after allocations to project A and project B are made. Since one does not know the optimal allocations to project A and project B, the optimal allocation and the corresponding revenue from project C must be determined for all possible remaining funds after allocations to project A and project B have been made. Furthermore, whatever amount is allocated to project A, the allocations to project B and project C must be made optimal with respect to the remaining funds after the allocation is made to project A. To find the optimal allocation to project B, one finds the allocation maximizing the revenue from project B together with the optimal revenue from project C as a function of remaining funds from the allocation to project B. Finally, the optimal allocation to project A is determined, to maximize the revenue from project A plus the optimal revenue from both project B and project C, as a function of the funds remaining after the allocation to project A.

This description of the DP algorithm applied to a budget allocation example can be depicted schematically as Fig. 8.3, from which the basic elements and terminologies of a DP formulation are defined.
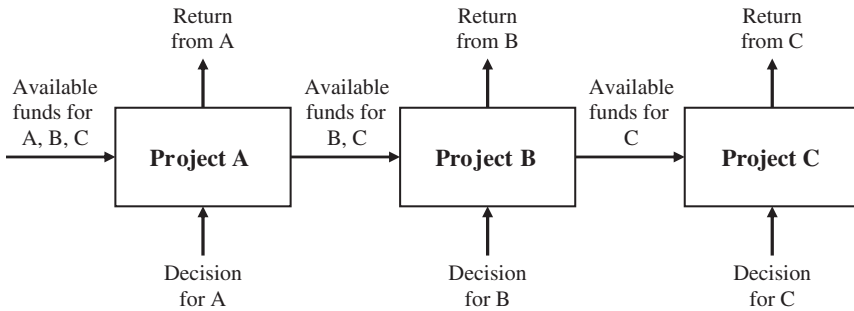
Figure 8.3 Dynamic programming diagram for budget allocation example.

1. *Stages (n)* are the points in the problem where decisions are to be made. In the funds allocation example, each different project represents different stages in the DP model.

2. *Decision variables (d_n)* are courses of action to be taken for each stage. The decision in each stage (project) is the alternative within the project to be selected for funding. The number of decision variables d_n in each stage is not necessarily equal to one.

3. *State variables (S_n)* are variables describing the state of a system at different stages. A state variable can be discrete or continuous, finite or infinite. Referring to Fig. 8.3, at any stage n, there are the input state S_n and the output state S_{n+1}. The state variables of the system in a DP model have the function of linking between succeeding stages so that when each stage is optimized separately, the resulting decision is automatically feasible for the entire problem. Furthermore, it allows one to make optimal decisions for the remaining stages without having to check the effect of future decisions on the decisions made previously. Given the current state, an optimal policy for the remaining stages is independent of the policy adopted in the previous stages. This is called *Bellman's principle of optimality*, which serves as the backbone of the optimization algorithm of the DP technique.

4. *Stage return (r_n)* is a scalar measure of effectiveness of the decision made in each stage. It is a function of input state, output state, and the decision variable of a particular stage. That is, $r_n = r(S_n, S_{n+1}, d_n)$.

5. *State transition function (t_n)* is a single-valued transformation that expresses the relationships between input state, output state, and decision. In general, through the stage transition function, the output state S_{n+1} at any stage n can be expressed as the function of input state S_n and the decision d_n as

$$S_{n+1} = t_n(S_n, d_n) \tag{8.8}$$

The solution begins with finding the optimal decision for each possible state in the last stage (called the *backward recursive*) or in the first stage (called the *forward recursive*). Usually, one can exchange the sequence of the decision-making process. Hence which end to begin will be trivial. A recursive

relationship that identifies the optimal policy for each state at any stage n can be developed, given the optimal policy for each state at the next stage $n + 1$. This backward-recursive equation, referring to Fig. 8.4, can be written as

$$f_n^*(S_n) = \text{opt}_{d_n}\{r_n(S_n, d_n)\} \qquad \text{for } n = N$$
$$= \text{opt}_{d_n}\{r_n(S_n, d_n) \circ f_{n+1}^*[t_n(S_n, d_n)]\} \qquad \text{for } n = 1, 2, \ldots, N - 1 \quad (8.9)$$

in which $\circ$ represents a general algebraic operator that can be $+$, $-$, $\times$, or others.

The efficiency of an optimization algorithm is commonly measured by the computer time and storage required in problem solving. In the DP algorithm, the execution time mainly arises from the evaluation of the recursive formula, whereas the storage is primarily for storing the optimal partial return and the decision for each feasible state in each stage. DP possesses several advantages in solving problems involving the analysis of multiperiod processes; however, there are two major disadvantages of applying DP to many hydrosystems problems, i.e., the computer memory and time requirements. These disadvantages would become especially severe under two situations: (1) when the number of state variables is beyond three or four and (2) when DP is applied in a discrete fashion to a continuous state space. The problem associated with the second situation is that difficulties exist in obtaining the true optimal solution without having to considerably increase discretization of the state space.

Because of the prohibitive memory requirement of DP for multidimensional systems, several attempts have been made to reduce this problem. One such modification is the *discrete differential DP* (DDDP). The DDDP is an iterative DP that is specially designed to overcome the shortcomings of the DP approach. The DDDP uses the same recursive equation as the DP to search for an improved trajectory among discrete states in the stage-state domain in the vicinity of an initially specified or trail trajectory (Heidari et al., 1971). Instead of searching over the entire state-stage domain for the optimal trajectory, as is the case for DP, DDDP examines only a portion of the state-stage domain, saving a considerable amount of computer time and memory (Chow et al., 1975). This optimization procedure is solved through iterations of trial states and decisions to search for optimal returns (maximum or minimum) for a system subject to the constraints that the trial states and decisions should be within the respective admissible domain in the state and decision spaces.
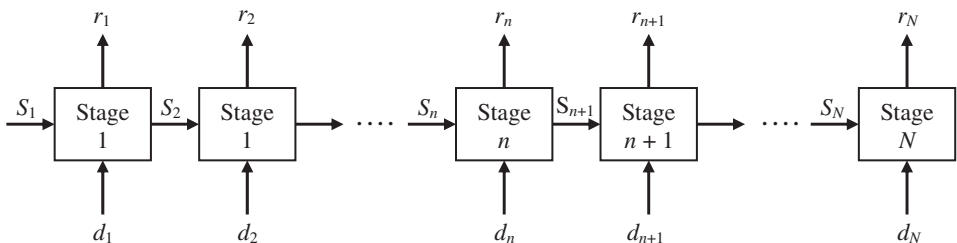
Figure 8.4 Schematic diagram of dynamic programming representation.

General purpose computer codes for solving DP problems are not available commercially because most problems are very specific and cannot be cast into a general framework such as Eqs. (8.1a–c). Therefore, analysts often have to develop a customized compute code for a specific problem under consideration.

Global optimization techniques.   To optimize a complex hydrosystem involving a large number of interrelated decision variables, it is generally difficult to be certain that the optimal solution can be obtained within a reasonable amount of time. In dealing with such "hard" problems, one could opt to obtain the optimal solution with the anticipation to consume a vary large amount of computational time using the optimization techniques previously mentioned or to reach a quick but good solution that is suboptimal through some approximate or heuristic algorithms. *Simulated annealing* (SA) and *genetic algorithm* (GA) are two types of high-quality general algorithms that are independent of the nature of the problem. Both SA and GA, by nature, are randomization algorithms that apply a local search based on stepwise improvement of the objective function through exploring the neighboring domain of the current solution. The quality of the local optimum can be strongly dependent on the initial solution, and in practice, it is difficult to assess the time needed to reach the global optimum. To avoid being trapped in a local optimum, local search algorithms can (1) try a large number of initial solutions, (2) introduce a more complex neighborhood structure to search a larger part of the solution space, and (3) accept limited transitions to explore the solution space in which the solution is inferior (Aarts and Korst, 1989).

Simulated annealing algorithms.   The simulated annealing (SA) algorithms solve optimization problems by using an analogy to physical annealing process of decreasing temperature to lower energy in a solid to a minimum level. The *annealing process* involves two steps: (1) increasing the temperature of a heat bath to a maximum value at which the solid melts, and (2) decrease carefully the temperature of the heat bath to allow particles to arrange themselves in a more structured lattice to achieve minimum energy level. If the temperature of the heat bath decreases too rapidly (called *quenching*), the solid could be frozen into a metastable state in which the solid would be brittle. The connection between optimization and the annealing process was first noted by Pincus (1970) and formalized as an optimization technique by Kirkpatrick et al. (1983).

SA algorithms employ *random search*, which not only accepts solutions that improve the current solution but also accept solutions that are inferior with a specified probability according to the *Metropolis criterion*, that is,

$$P\left(\text{accept } \boldsymbol{x}_j^{(r+1)}\right) = \exp\left\{ -\frac{\left[f\left(\boldsymbol{x}_j^{(r+1)}\right) - f\left(\boldsymbol{x}_*^{(r)}\right)\right]}{T} \right\} \qquad (8.10)$$

where $\boldsymbol{x}_*^{(r)}$ is the optimal solution in the rth iteration, $\boldsymbol{x}_j^{(r+1)}$ is the jth trial solution of the $(r+1)$th iteration, $f(\cdot)$ is the objective function value analogous

to the energy level of a solid, and T is the control parameter analogous to the temperature of the heat bath. As can be seen from Eq. (8.10), in a subsequent iteration a new trial solution yielding an objective function value that is a worse objective function value compared with the current optimum one will have lower (but not zero) probability of being accepted than a solution producing a better objective function value.

Implementation of the SA algorithm is remarkably easy, as shown in Fig. 8.5, which involves the following steps: (1) generation of possible solutions to explore
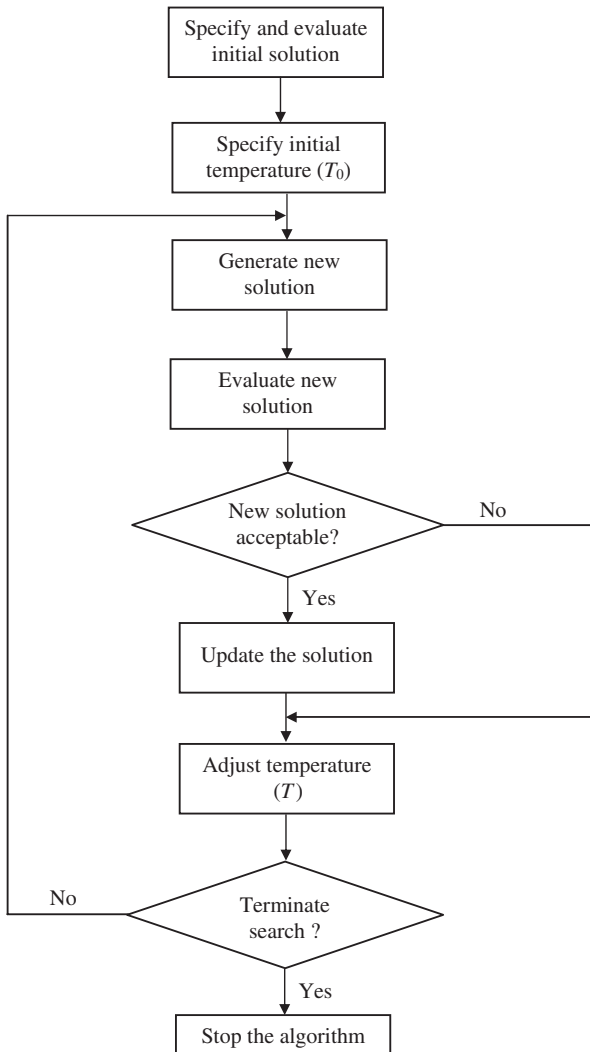
Figure 8.5 Algorithm of simulated annealing.

the solution space, (2) evaluation of the objective function, and (3) definition of the annealing schedule specified by the initial control parameter T_0, decrement of control parameter ΔT, and convergence criteria. Mechanisms for generating trial solutions in each iteration involve introducing random changes with the intent to cover the solution domain to be searched. The domain of search generally changes with the iteration, and there are many ways to implement domain search (Vanderbilt and Louie, 1984; Parks, 1990).

Since SA algorithms only require objective function evaluation at each generated trial solution, computational efficiency of the entire process could become an important issue because implementation of the algorithms anticipates a large number of function evaluations. Many optimization problems in hydrosystems engineering involve objective functions whose values depend on physical constraints defined by complex flow simulations. In such cases, it is worthy of the effort to search and implement more computationally efficient procedures. To handle constraints in an SA algorithm, the simplest way is to reject the trial solution if it leads to a constraint violation. Alternatively, penalty function can be introduced to account for the constraint violations.

In implementation of the SA algorithm, T_0 is generally set to be high enough to allow virtually all trial solutions to be accepted. It is analogous to having the temperature in the heat bath high enough to "melt" the solid. It is equivalent to the acceptance probability for T_0 being close to 1. As the solution improves with the iterations, the control parameter T gradually decreases in value. The SA iteration is terminated when the control parameter T reaches a specified final value or the total number of trial solutions is reached. Alternatively, one can halt the algorithm if lack of improvement in the objective a function value is defined, which can be (1) no improvement can be found in all trial solutions at one control parameter and (2) acceptance ratio falls below a specified value.

Genetic algorithms. *Genetic algorithms* (GAs) are a class of computer-based search procedures that emulate the biologic evolution processes of natural selection and genetics. Since its introduction by Goldberg in 1983, this innovative search procedure has been applied widely for optimization in a wide variety of areas (Goldberg, 1989). GAs have been demonstrated to be robust search procedures that outperform the conventional optimization procedures, in particular for problems with high dimensionality, discontinuities, and noise.

Using GA for optimization, analogues are made between the properties of an optimization problem and the biologic process of gene selection and reproduction. The solution space for an optimization problem can be treated as the environment in which potential solutions can be considered as genes. The degree of fitness of each gene can be measured by the value of the objective function of an optimization problem. In each iteration of a GA search, several genes representing solutions are generated from the population. These genes compete for their survival based on their fitness: The one that is fitter is more likely to survive and influence the next generation. Through this competition, the population evolves to contain high-performance individuals. In a GA, iteration is

represented by a generation, and decision variables are analogous to biologic genes and are represented by coded structures either in the form of a real number or binary codes. A point in the solution space is represented by a collection of genes, and the coded genes are juxtaposed to form an individual or chromosome.

Like any optimum-seeking algorithm, a GA requires the specification of an initial population (the first generation) of potential solutions. This can be done by random generation of a population or by specifying an initial solution. In the initial population, the fitness of an individual is evaluated with respect to the objective function. Those individuals with a high level of fitness will have higher chance being selected to produce offspring than those individuals with a lower level of fitness. The selection can be conducted in many different ways, such as stochastic universal sampling (Baker, 1987), which behaves like a roulette wheel with the slot size apportioned to each individual's relative fitness in the population. The principle of the selection is to prefer better solutions to worse ones. This very feature of the selection procedure in a GA is similar to the Metropolis acceptance criterion of SA algorithms. The implementation of such selection procedure will prevent the solutions from being trapped at the local optimum.

On the completion of selection of individuals in a generation, individuals selected will mate to produce the population of the next generation. During the mating, biologic processes of gene crossover and mutation could take place. Again, fitness of individuals in the population of the new generation will be evaluated based on which selection and mating processes will be repeated. A schematic diagram of a GA is shown in Fig. 8.6. Through this repeated
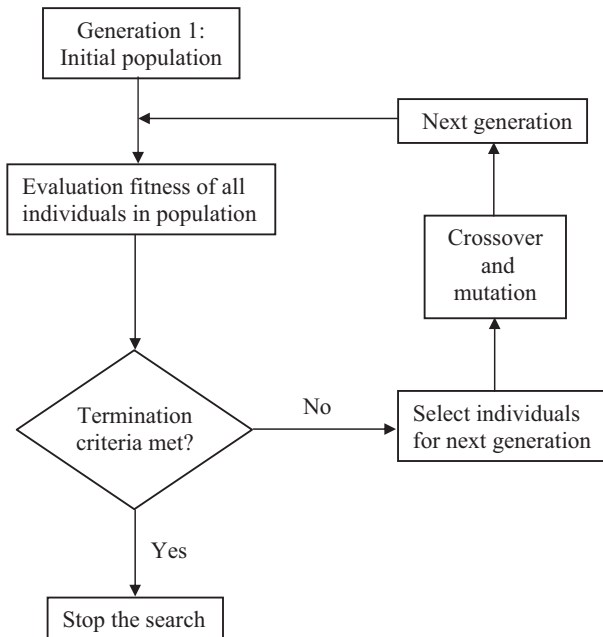
Figure 8.6 Schematic diagram of genetic algorithm.

fitness-selection-reproduction cycle, the population generated by the GA will evolve toward improved solutions.

Two main mechanisms are involved for those selected individuals to generate offspring in the next generation, i.e., crossover and mutation. *Crossover* allows information exchange between two chosen individuals forming the parents. In GA, crossover is done by randomly selecting positions in a chromosome of two individuals to swap their coded genes to produce the offspring. The rate of crossover has been suggested to be from 0.6 to 1.0 (Goldberg, 1989). *Mutation* then is a process by which new genetic materials can be introduced into the population. It is applied on a bit-by-bit basis, with the mutation rate specified by the user.

8.2   Optimization of System Reliability

As described in Chap. 5, reliability of a multicomponent system depends on the component reliability, number of redundant components, and the arrangement of the components. With everything else remaining fixed, the system reliability gets higher as the number of redundant components increases. However, nothing is free—achieving higher system reliability has to be paid for by a higher price for the system. In general, practically all systems exhibit a behavior of diminishing rate of return as the number of redundant components increases (see Example 5.2). In the context in system reliability, this behavior is describable by a strictly concave function relation between the system reliability and number of redundant components (or total system cost). A relevant issue is how much one is willing to invest before the improvement in system reliability is no longer economically justifiable. The problem becomes even more challenging when there are several components with different reliability levels competing for the limited resources, such as budget and space. How would an engineer decide the best allocation of resources to achieve the highest system reliability possible? This section describes several examples showing how system reliability can be optimized by various techniques introduced in Sec. 8.1.

8.2.1   Reliability design with redundancy

Consider the design of a hydrosystem consisting of n subsystems that are arranged in series so that the failure of one subsystem will cause the failure of the entire system. In this case, reliability of the hydrosystem can be improved by installing standby units in each subsystem (see Fig. 8.7). Figure 8.7 consists of a series-parallel configuration that is called *unit redundancy*. Suppose that each subsystem can install up to K standby units and that the total capital available for the hydrosystem is C. Furthermore, the cost functions are known, with $C_i(k_i)$ being the cost associated with installing k_i standby units on the ith subsystem.

Suppose that the engineer is interested in determining the number of standby units for each subsystem to maximize the system reliability α_{sys} without
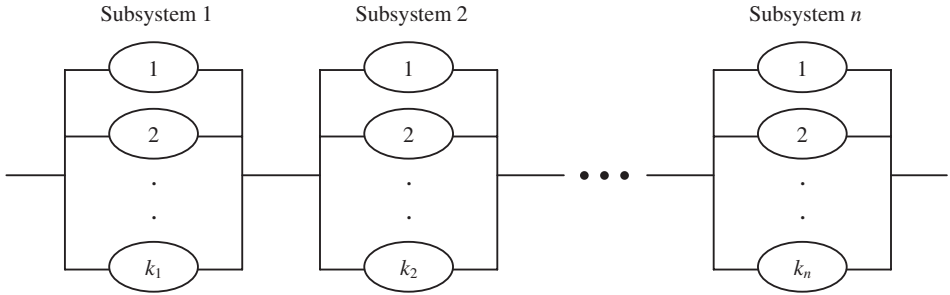
Figure 8.7 Unit redundancy with series-parallel configuration.

exceeding the available capital. The optimization model for the problem can be expressed as

$$\text{Maximize} \qquad \alpha_{\text{sys}} = \prod_{i=1}^{n} \alpha_i(k_i) \tag{8.11a}$$

$$\text{subject to} \qquad \sum_{i=1}^{n} C_i(k_i) \leq C \tag{8.11b}$$

where k_i is the nonnegative integer-valued decision variable, $0 \leq k_i \leq K$, and $\alpha_i(k_i)$ is the reliability of the ith subsystem installed with k_i standby units.

This optimization problem can be solved efficiently by the DP approach described in Sec. 8.1.3. The stages are the subsystems $i = 1, 2, \ldots, n$. The DP backward-recursive equation can be written as

$$f_i(b_i) = \begin{cases} \max_{k_i}[\alpha_i(k_i)] & i = n \\ \max_{k_i}\{\alpha_i(k_i) \times f_{i+1}[b_i - C_i(k_i)]\} & i = 1, 2, \ldots, n-1 \end{cases} \tag{8.12}$$

where b_i is the state variable representing the total capital available for subsystems $i, i + 1, \ldots, n$.

Alternatively, the design engineer may be interested in finding the system configuration associated with the least total capital investment while achieving some acceptable reliability for the system $\alpha_{\text{sys,min}}$. The problem of this type can be expressed by the following model:

$$\text{Minimize} \qquad \sum_{i=1}^{n} C_i(k_i) \tag{8.13a}$$

$$\text{subject to} \qquad \prod_{i=1}^{n} \alpha_i(k_i) \geq \alpha_{\text{sys,min}} \tag{8.13b}$$

The model defined by Eqs. (8.13a–b) also can be solved by the DP approach. To illustrate the application of other optimization procedure, let the reliability

of each standby unit in each subsystem be equal to α_i and the unit cost be $c_i, i = 1, 2, \ldots, n$. Furthermore, the minimum acceptable system reliability is set equal to $\alpha_{\text{sys,min}}$. Then, Eqs. (8.13a–b) can be rewritten as

$$\text{Minimize} \quad \sum_{i=1}^{n} c_i k_i \tag{8.14a}$$

$$\text{subject to} \quad \prod_{i=1}^{n} \left[1 - (1 - \alpha_i)^{k_i} \right] = \alpha_{\text{sys,min}} \tag{8.14b}$$

The model, Eqs. (8.14a–b), involves a linear objective function and a nonlinear equality constraint. To simplify the multiplicative relationship of Eq. (8.14b), a new variable p_i is defined to satisfy the following equation:

$$\alpha_{\text{sys,min}}^{p_i} = 1 - (1 - \alpha_i)^{k_i}$$

In terms of the new variable p_i, the original decision variable k_i and the constraint Eq. (8.14b) can be expressed, respectively, as

$$k_i = \frac{\ln\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)}{\ln(1 - \alpha_i)} \qquad i = 1, 2, \ldots, n \tag{8.15a}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{8.15b}$$

Therefore, the original model, Eqs. (8.14 a–b), can be written as

$$\text{Minimize} \quad \sum_{i=1}^{n} c_i \times \frac{\ln\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)}{\ln(1 - \alpha_i)} \tag{8.16a}$$

$$\text{subject to} \quad \sum_{i=1}^{n} p_i = 1 \tag{8.16b}$$

This constrained minimization problem, Eqs. (8.16a–b), can be solved by minimizing the following Lagrangian function:

$$\underset{p_1, p_2, \ldots, p_n, \lambda}{\text{Minimize}} \ L(p_1, p_2, \ldots, p_n, \lambda) = \sum_{i=1}^{n} c_i \times \frac{\ln\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)}{\ln(1 - \alpha_i)} + \lambda \left(\sum_{i=1}^{n} p_i - 1 \right) \tag{8.17}$$

Solving

$$\frac{\partial L(p_1, p_2, \ldots, p_n, \lambda)}{\partial p_i} = \frac{c_i}{\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)(1 - \alpha_i)} \frac{\partial \left(1 - \alpha_{\text{sys,min}}^{p_i}\right)}{\partial p_i}$$

$$+ \lambda = 0 \qquad i = 1, 2, \ldots, n \tag{8.18}$$

results in

$$\lambda = \frac{c_i \alpha_{\text{sys,min}}^{p_i} \ln\left(\alpha_{\text{sys,min}}^{p_i}\right)}{\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)(1 - \alpha_i)} \tag{8.19}$$

Suppose that the minimum acceptable system reliability $\alpha_{\text{sys,min}}$ is chosen to be close to 1. In such a case,

$$\lim_{\alpha_{\text{sys,min}} \to 1} \frac{\ln\left(\alpha_{\text{sys,min}}^{p_i}\right)}{\left(1 - \alpha_{\text{sys,min}}^{p_i}\right)} = -\frac{1}{p_i} \tag{8.20}$$

Then Eq. (8.19) can be reduced to

$$\lambda = -\frac{c_i}{p_i \ln(1 - \alpha_i)} \tag{8.21}$$

which renders

$$p_i = -\frac{c_i}{\lambda \ln(1 - \alpha_i)} \qquad i = 1, 2, \ldots, n \tag{8.22}$$

Substituting Eq. (8.22) into Eq. (8.16b) yields

$$\lambda = -\sum_{i=1}^{n} \frac{c_i}{\ln(1 - \alpha_i)} \tag{8.23}$$

Then, by Eqs. (8.22) and (8.23), the new variable p_i can be obtained, in terms of the unit cost c_i, and reliability of standby unit α_i, as

$$p_i = \frac{\left[\frac{c_i}{\ln(1 - \alpha_i)}\right]}{\left[\sum_{j=1}^{n} \frac{c_j}{\ln(1 - \alpha_j)}\right]} \qquad i = 1, 2, \ldots, n \tag{8.24}$$

Once the values of p_i's are computed by Eq. (8.24), the number of standby units for each subsystem k_i can be obtained by Eq. (8.15a). Finally, one should realize that the values of k_i so obtained are not guaranteed to be integer-valued. A round-off to the closest integer may be needed.

8.2.2 Determination of optimal maintenance schedule

In Sec. 6.3.4 it was shown that the implementation of scheduled maintenance can increase the mean time to failure (MTTF) of a system having an increasing hazard function. Increasing maintenance frequency would result in a decrease in repair frequency and vice versa. Suppose that an engineer is considering implementing a regular scheduled maintenance for a system. The problem of interest is to determine the optimal maintenance frequency associated with the minimum total cost, which includes the maintenance cost and repair cost. Of course, the issue is worth considering if the cost of maintenance is much

lower than the cost of repair. Otherwise, there will be little economic incentive to conduct scheduled maintenance. The following descriptions show a simple example of determining the optimal maintenance schedule associated with the minimum total cost of repair and maintenance. More sophisticated models for dam safety inspection have been developed by Tang and Yen (1991, 1993).

The total cost per unit time with a maintenance cycle of time interval length t_M can be expressed as

$$TC(t_M) = C_R \times f_R(t_M) + C_M \times f_M(t_M) \tag{8.25}$$

in which C_R and C_M are unit cost per repair and unit cost per maintenance, respectively, and f_R and f_M are the repair and maintenance frequencies, respectively.

Assume that the repair is ideal. The repair frequency (number of repairs per unit time) f_R for a specified maintenance interval t_M can be calculated by

$$f_R(t_M) = \frac{1}{t_M} \int_0^{t_M} f_t(\tau) d\tau \tag{8.26}$$

in which $f_t(t)$ is the failure density function. On the other hand, since there will be one maintenance within each scheduled time interval t_M, the maintenance frequency is $1/t_M$. Therefore, the total cost per unit time is

$$TC(t_M) = \frac{1}{t_M} \left[C_R \times \int_0^{t_M} f_t(\tau) d\tau + C_M \right] \tag{8.27}$$
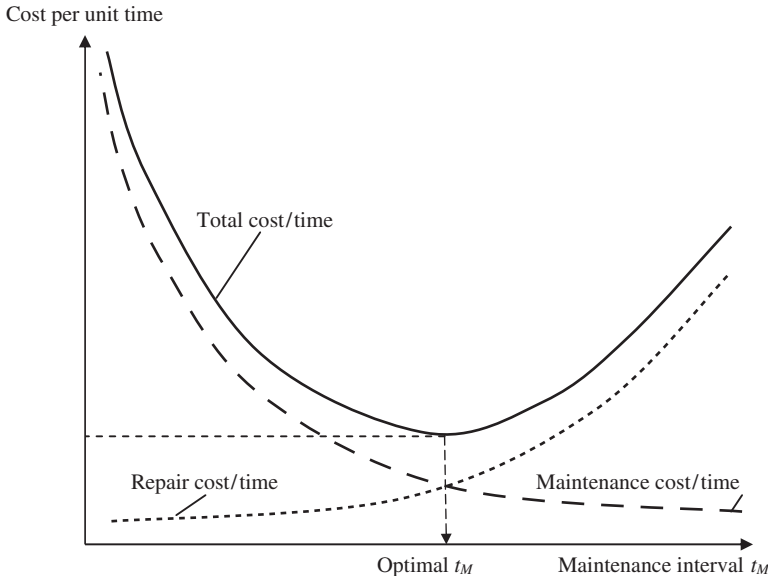
Figure 8.8 Tradeoff between repair cost and maintenance cost.

The relationships between the three cost items and the scheduled maintenance interval are shown in Fig. 8.8, which shows that the repair cost per unit time increases with t_M, whereas the maintenance cost per unit time decreases with t_M. Therefore, there is a tradeoff between the two cost components, and the objective is to determine the optimal scheduled maintenance interval t_M associated with the least total cost.

The optimal scheduled maintenance time interval can be obtained by solving $d[TC(t_M)]/d(t_M) = 0$, that is,

$$C_R \int_0^{t_M} f_t(\tau)d\tau + C_M + t_M f_t(t_M) = 0 \tag{8.28}$$

In general, solving Eq. (8.28) requires the use of numerical root-finding procedures.

8.3 Optimal Risk-Based Design of Hydrosystem Infrastructures

Reliability analysis methods can be applied to design hydrosystem infrastructures with or without considering risk costs. *Risk costs* are those cost items incurred owing to the unexpected failure of structures, and they can be broadly classified into tangible and intangible costs. *Tangible costs* are those measurable in terms of monetary unit, which include damage to property and structures, loss of business, cost of repair, etc. On the other hand, *intangible costs* are not measurable by monetary unit, such as psychological trauma, loss of lives, social unrest, damage to the environment, and others. Without considering risk costs, reliability has been explicitly accounted for in the design of storm sewer systems (Yen and Ang, 1971; Yen et al., 1976; Yen and Jun, 1984), culverts (Yen et al., 1980; Tung and Mays, 1980), and levees (Tung and Mays, 1981a; Lee and Mays, 1986). Cheng et al. (1986) demonstrated how to apply the AFOSM method to calculate the risk reduction associated with freeboard in dam design. Melching et al. (1987) suggested different flood peak-risk curves for forecasting and for design. However, it is the risk-based least cost design of hydrosystem infrastructure that promises to be potentially the most significant application of reliability analysis.

The *risk-based design* procedure integrates the procedures of uncertainty and reliability analyses in the design practice. The procedure considers the tradeoff among various factors such as failure probability, economics, and other performance measures in hydraulic structure design. Plate and Duckstein (1987, 1988) list a number of performance measures, called the *figures of merit*, in the risk-based design of hydraulic structures and water resource systems, which are further discussed by Plate (1992). When risk-based design is embedded into an optimization framework, the combined procedure is called the *optimal risk-based design*.

8.3.1 Basic concept

The basic concept of risk-based design is shown schematically in Fig. 8.9. The risk function accounting for the uncertainties of various factors can be obtained using the reliability computation procedures described in earlier chapters. Alternatively, the risk function can account for the potential undesirable consequences associated with the failure of hydrosystem infrastructures. For simplicity, only the tangible damage cost is presented herein.

Risk costs associated with the failure of a hydrosystem infrastructure cannot be predicted accurately from year to year. A practical way is to quantify risk cost using an expected value on an annual basis. The *total annual expected cost* (TAEC) is the sum of the annual installation cost and annual expected damage cost, which can be expressed mathematically as

$$\text{TAEC}(\boldsymbol{x}) = \text{FC}(\boldsymbol{x}) \times \text{CRF} + \text{E}(D|\boldsymbol{x}) \tag{8.29}$$

where FC is the first or total installation cost, which is the function of decision vector $\boldsymbol{x}$ defined by the size and configuration of the hydraulic structure, $\text{E}(D|\boldsymbol{x})$ is the *annual expected damage cost* associated with structural failure, and CRF is the *capital recovery factor*, which brings the present worth of the installation costs to an annual basis and can be computed as

$$\text{CRF} = \frac{(1+i)^T - 1}{i(1+i)^T} \tag{8.30}$$
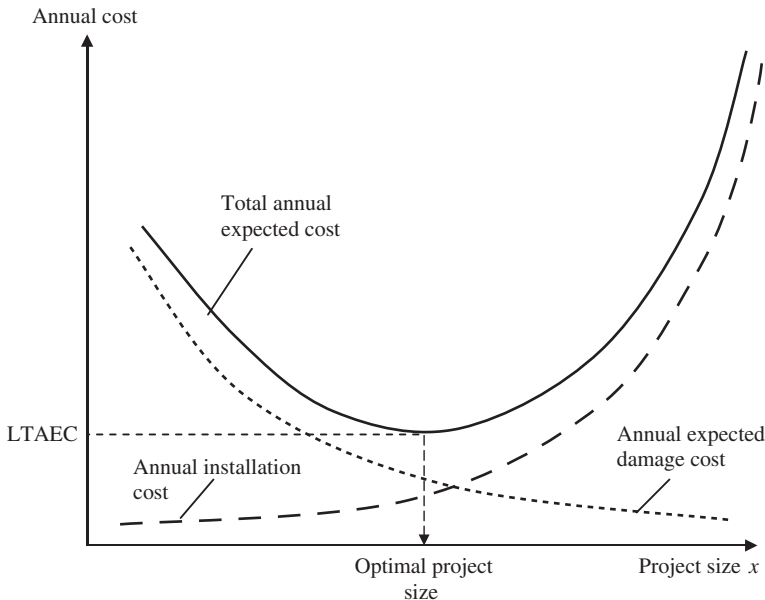
Figure 8.9 Schematic diagram of optimal risk-based design.

with T and i being the expected service life of the structure and the interest rate, respectively.

In practice, the *optimal risk-based design* determines the optimal structural size, configuration, and operation such that the annual total expected cost is minimized. Referring to Fig. 8.6, as the structural size increases, the annual installation cost increases, whereas the annual expected damage cost associated with the failure decreases. The optimal risk-based design procedure attempts to determine the lowest point on the total annual expected cost curve. Mathematically, the optimal risk-based design problem can be stated as

$$\text{Minimize} \quad \text{TAEC}(\boldsymbol{x}) = \text{FC}(\boldsymbol{x}) \times \text{CRF} + \text{E}(D|\boldsymbol{x}) \quad (8.31a)$$

$$\text{subject to} \quad g_i(\boldsymbol{x}) = 0 \quad i = 1, 2, \ldots, m \quad (8.31b)$$

where $g_i(\boldsymbol{x}) = 0$, $i = 1, 2, \ldots, m$, are constraints representing the design specifications that must be satisfied.

In general, the solution to Eqs. (8.31a–b) could be acquired through the use of appropriate optimization algorithms. The selection or development of the solution algorithm is largely problem-specific, depending on the characteristics of the problem to be optimized. Section 8.4 describes an application of the optimal risk-based design to pipe culverts for roadway drainage.

8.3.2 Historical development of hydrosystem design methods

The evolution of hydrosystem design methods can be roughly classified into four stages: (1) historical event–based design, (2) return-period design, (3) conventional risk-based design, and (4) optimal risk-based design with consideration given to a variety of uncertainties.

Historical event–based design. In the design of hydrosystem infrastructures and the establishment of land-use management practices to prevent and/or reduce damages resulting from natural disasters, the risk (damage) assessment typically has been implicit. The earliest structures and land-use management approaches for flood protection were designed or established on the basis of their ability to withstand previous disastrous floods. For example, Chow (1962) noted that the Dun waterway table used to design railroad bridges in the early 1900s was primarily determined from channel areas corresponding to high-water marks studied during and after floods. Thus previous large floods of unknown frequency could pass through the designed bridges safely. Also, after a devastating flood on the Mississippi River in 1790, a homeowner in Saint Genieve, Missouri, rebuilt his house outside the boundary of that flood. Similar rules were applied in the design of coastal-protection works in The Netherlands at the time the Zuiderzee was closed (1927–1932) (Vrijling, 1993).

Rules based on previous experience work well in some cases. For example, the house in Missouri was not flooded until the 1993 flood on the Mississippi River,

and the Zuiderzee protection works survived the 1953 storm that devastated the southwestern part of The Netherlands. However, in most cases these methods are inadequate because human experience with floods and other natural hazards do not include a broad enough range of events. As noted by Vrijling (1993), "One is always one step behind when a policy is only based on historical facts."

Return-period design.   In the early part of the twentieth century, the concept of frequency analysis began to emerge as a method to extend limited data on extreme events to probabilistically estimate the magnitude of rarely occurring events. Frequency analysis of observed events is a key aspect of meteorologic, hydrologic, and seismic hazard analyses. Thus, using frequency-analysis methods, it is possible to estimate events with magnitudes beyond those that have been observed. This necessitates the selection of a societally acceptable hazard frequency (see Sec. 8.3.6).

Using the return-period design approach, a hydraulic engineer first determines the design discharge from a frequency-discharge relationship by selecting an appropriate design frequency or return period. The design discharge then is used to determine the size and layout of the hydrosystem that has a satisfactory hydraulic performance. In the return-period design method, selection of the design return period is crucial to the design. Once the design return period is determined, it remains fixed throughout the whole design process. In the past, the design return period was selected subjectively on the basis of an individual's experience or the societally acceptable hazard frequency (Sec. 8.3.6). Selection of the design return period is a complex procedure that involves considerations of economic, social, legal, and other factors. However, the procedure does not account for these factors explicitly.

Conventional risk-based design.   Risk-based design is a procedure that evaluates among alternatives by considering the tradeoff between the investment cost and the expected economic losses due to failures. Specifically, the conventional risk-based design considers the inherent hydrologic uncertainty in calculation of the expected economic losses. In the risk-based design procedure, the design return period is a decision variable instead of being a preselected design parameter value, as with the return-period design procedure.

The concept of risk-based design has been recognized for many years. As early as 1936, Congress passed the Flood Control Act (U.S. Statutes 1570), in which consideration of failure consequences in the design procedure was advocated. The economic risks or the expected flood losses were not considered explicitly until the early 1960s. Pritchett's work (1964) was one of the early attempts to apply the risk-based hydraulic design concept to highway culverts. At four actual locations, Pritchett calculated the investment costs and expected flood damage costs on an annual basis for several design alternatives, among which the most economical one was selected. The results indicated that a more economical solution could be reached by selecting smaller culvert sizes compared with the

traditional return-period method used by the California Division of Highways. The conventional approach has been applied to the design of highway drainage structures such as culverts (Young et al., 1974; Corry et al., 1980) and bridges (Schneider and Wilson, 1980). Inherent randomness of hydrologic processes is integrated with reliability analysis in seismic, structural, and geotechnical aspects in the design of new dams (Pate-Cornell and Tagaras, 1986) and evaluation of alternatives for rehabilitating existing dams (McCann et al., 1984; Bureau of Reclamation, 1986; National Research Council, 1983).

Risk-based design considering other uncertainties.   In the conventional risk-based hydraulic design procedure, economic risks are calculated considering only the randomness of hydrologic events. In reality, there are various types of uncertainties, as described in Sec. 1.2, in a hydrosystem infrastructure design. Advances have been made to incorporate other aspects of uncertainty in the design of various hydrosystem infrastructures. For example, both hydrologic and hydraulic uncertainties were considered in the design of highway drainage structures (Mays, 1979; Tung and Mays, 1980, 1982; Tung and Bao, 1990), storm sewer systems (Yen and Ang, 1971; Yen and Jun, 1984; Tang and Yen, 1972; Tang et al., 1975, 1976), levee systems (Tung and Mays, 1981b), riprap design of stable channels (Tung, 1994), and river diversion (Afshar et al., 1994). Inherent hydrologic uncertainty, along with parameter and model uncertainties, was considered in design of levee systems (Wood, 1977; Bodo and Unny, 1976). Economic uncertainty, along with hydrologic and hydraulic uncertainties, has been considered in flood-damage-reduction projects (U.S. Army Corps of Engineers, 1996).

8.3.3   Tangible costs in risk-based design

Design of a hydrosystem infrastructure, by nature, is an optimization problem consisting of an analysis of the hydraulic performance of the structure to convey flow across or through the structure and a determination of the most economical design alternative. The objective function is to minimize the sum of capital investment cost, the expected flood damage costs, and operation and maintenance costs. The relevant variables and parameters associated with the investment cost and the expected damage costs of a hydraulic structure, i.e., a highway drainage structure, are listed in Tables 8.1 and 8.2, respectively. The maintenance costs over the service life of the structure generally are treated as a yearly constants. Based on Tables 8.1 and 8.2, the information needed for the risk-based design of hydraulic structures can be categorized into four types:

1. *Hydrologic/physiographic data* include flood and precipitation data, drainage area, channel bottom slope, and drainage basin slope. These are needed to predict the magnitude of hydrologic events such as streamflow and rainfall by frequency analysis and/or regional analysis.

TABLE 8.1 **Variables and Parameters Relevant in Evaluating Capital Investment Cost of Highway Drainage Structures**

	Pipe culverts	Box culverts	Bridges
Parameters	Unit cost of culvert	Unit cost of concrete Unit cost of steel	Unit cost of bridge
Variables	Number of pipes Pipe size Pipe length Pipe material	Number of barrels Length of barrel Width of barrel Quantity of concrete Quantity of steel	Bridge length Bridge width

SOURCE: After Tung et al. (1993).

2. *Hydraulic data* include flood-plain slopes, geometry of the channel cross section, roughness coefficients, size of the structural opening, and height of the embankment. These are needed to determine the flow-carrying capacities of hydraulic structures and to perform hydraulic analysis.

TABLE 8.2 **Damage Categories with Related Economic Variables and Site Characteristics in Risk-Based Design of Highway Drainage Structures**

Damage category	Economic variables	Site characteristics

Floodplain property damage:
  Losses to crops
  Losses to buildings
	Types of crops
	Economic value of crops
	Types of buildings
	Economic values of buildings
	  and contents
		Location of crop fields
		Location of buildings
		Physical layout of drainage
		  structures
		Roadway geometry
		Flood characteristics
		Stream valley cross-section
		Slope of channel profile
		Channel and floodplain roughness

Damage to pavement and
embankment:
  Pavement damage
  Embankment damage
	Material cost of pavement
	Material cost of embankment
	Equipment costs
	Labor costs
	Repair rate for pavement and
	  embankment
		Flood magnitude
		Flood hydrograph
		Overtopping duration
		Depth of overtopping
		Total area of pavement
		Total volume of embankment
		Types of drainage structure
		  and layout
		Roadway geometry

Traffic-related losses:
  Increased travel cost due
    to detour
  Lost time of vehicle occupants
  Increased risk of accidents on
    a flooded highway
	Rate of repair
	Operational cost of vehicle
	Distribution of income for vehicle
	  occupants
	Cost of vehicle accident
	Rate of accident
	Duration of repair
		Average daily traffic volume
		Composition of vehicle types
		Length of normal detour paths
		Flood hydrograph
		Duration and depth of overtopping

SOURCE: After Tung and Bao (1990).

3. *Structural data* include material of substructures and layout of structure.

4. *Economic data* include (a) type, location, distribution, and economic value of upstream properties such as crops and buildings, (b) unit costs of structural materials, equipment, operation of vehicles, accidents, occupancy, and labor fee, (c) depth and duration of overtopping, rate of repair, and rate of accidents, and (d) time of repair, and length of detour.

In the design of hydrosystem infrastructures, the installation cost often depends on environmental conditions, such as the location of the structure, geomorphic and geologic conditions, the soil type at the structure site, type and price of construction material, hydraulic conditions, flow conditions, recovery factor of the capital investment, and labor and transportation costs. In reality, these cost-affecting factors would result in uncertainties in the cost functions used in the analysis. However, a practical way to incorporate economic uncertainties in the risk-based design of hydrosystem infrastructures remains to be developed.

8.3.4   Evaluations of annual expected flood damage cost

In risk-based and optimal risk-based designs of hydrosystem infrastructures, the thrust of the exercise, after uncertainty and risk analyses are performed, is to evaluate $E(D|\boldsymbol{x})$ as the function of the probability density functions (PDFs) of loading and resistance, damage function, and the types of uncertainty considered.

Conventional approach. In conventional risk-based design, where only inherent hydrologic uncertainty is considered, the structural size $\boldsymbol{x}$ and its corresponding flow-carrying capacity q_c, in general, have a one-to-one monotonically increasing relationship. Consequently, the design variables $\boldsymbol{x}$ alternatively can be expressed in terms of design discharge of the hydrosystem infrastructure. The annual expected damage cost, in the conventional risk-based hydraulic design, can be computed as

$$E_1(D|\boldsymbol{x}) = \int_{q_c^*}^{\infty} D(q|q_c^*) f_q(q)\, dq \qquad (8.32)$$

where q_c^* is the deterministic flow capacity of a hydraulic structure subject to random floods following a PDF $f_q(q)$, and $D(q|q_c^*)$ is the damage function corresponding to the flood magnitude of q and hydraulic structural capacity q_c^*. A graphic representation of Eq. (8.32) is shown in Fig. 8.10, and $E_1(D|\boldsymbol{x})$ corresponds to the shaded area under the damage-frequency curve. Owing to the complexity of the damage function and the form of the PDF of floods, the analytical integration of Eq. (8.31), in most real-life applications, is difficult, if not impossible. Hence the evaluation of annual expected damage cost by Eq. (8.32) is done numerically.
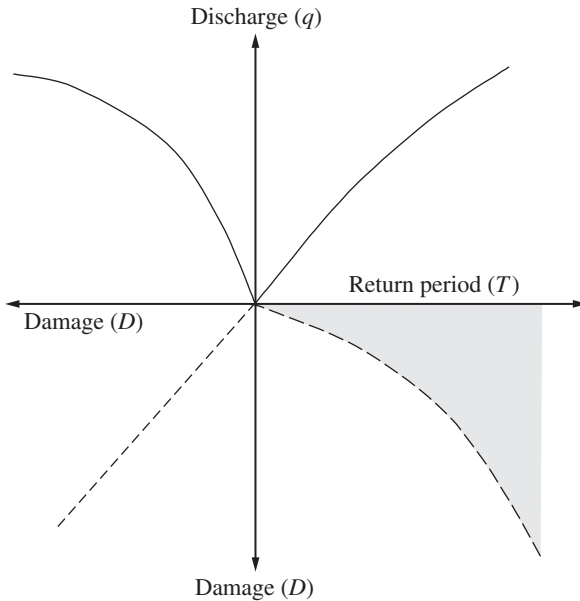
Figure 8.10   Computation of annual expected damage.

Note that Eq. (8.32) considers only the inherent hydrologic uncertainty owing to the random occurrence of flood events, represented by the PDF $f_q(q)$. It does not consider hydraulic and economic uncertainties. Furthermore, a perfect knowledge about the probability distribution of flood flow is assumed. This is generally not the case in reality.

Incorporation of hydraulic uncertainties.   As described in Sec. 1.2, uncertainties also exist in the process of hydraulic computations for determining the flow-carrying capacity of the hydraulic structure. In other words, q_c is a quantity subject to uncertainty. From the uncertainty analysis of q_c (Tung and Yen, 2005), the statistical properties of q_c can be estimated. Hence, to incorporate uncertainty feature of q_c into the risk-based design, the annual expected damage can be calculated as

$$E_2(D) = \int_0^\infty \left[\int_{q_c}^\infty D(q|q_c) f_q(q) dq \right] g_{qc}(q_c) dq_c = \int_0^\infty E_1(D|q_c) g_{qc}(q_c) dq_c$$

(8.33)

in which $g_{qc}(q_c)$ is the PDF of random flow-carrying capacity q_c. Again, in practical problems, the annual expected damage estimated by Eq. (8.33) would have to evaluated through the use of appropriate numerical integration schemes.

Considering hydrologic inherent and parameter uncertainties.   Since the occurrence of streamflow is random by nature, the statistical properties such as the mean, standard deviation, and skewness coefficient of the distribution calculated from a finite sample are also subject to sampling errors. In hydrologic frequency analysis, a commonly used frequency equation (Eq. 3.5) for determining the magnitude of a hydrologic event, say, a flood, of a specified return period T years is

$$q_T = \mu_q + K_T \, \sigma_q \qquad\qquad (8.34)$$

in which q_T is the magnitude of the hydrologic event of the T year, μ_q and σ_q are the population mean and standard deviation of floods, respectively, and K_T is the frequency factor depending on the skewness coefficient and distribution of the flood event.

   Consider floods being the hydrologic event that could potentially cause the failure of the hydraulic structure. Owing to the uncertainty associated with the estimated values of μ_q, σ_q, and K_T in Eq. (8.34), the flood magnitude of a specified return period q_T is also a random variable associated with its probability distribution (see Fig. 8.11) instead of being a single-valued quantity presented by its "average," as commonly done in practice. Section 3.8 describes the sample distributions for some of the probability distributions frequently
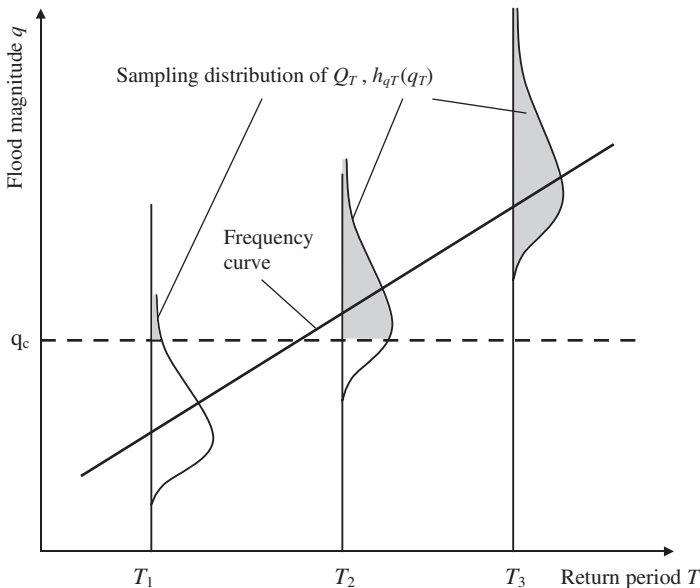
Figure 8.11 Schematic sketch of sampling distribution of flood magnitude estimator.

used in hydrologic flood frequency analysis. Hence there is an expected damage corresponding to the T-year flood magnitude that can be expressed as

$$E(D_T \,|q_c^*) = \int_{q_c^*}^{\infty} D(q_T \,|q_c^*) h_{q_T}(q_T)\, dq_T \tag{8.35}$$

where $E(D_T \,|q_c^*)$ is the expected damage corresponding to a T-year flood given a known flow capacity of the hydraulic structure q_c^*, $h_{qT}(q_T)$ is the sampling PDF of the flood magnitude estimator of a T-year return period, and q_T is the dummy variable for a T-year flood. Equation (8.35) represents an integration of flood damage over the shaded area associated with the sample distribution of a T-year flood. To combine the inherent hydrologic uncertainty, represented by the PDF of annual flood $f_q(q)$, and the hydrologic parameter uncertainty, represented by the sampling PDF for a flood sample of a given return period $h_{qT}(q_T)$, the annual expected damage cost can be written as

$$E_2(D|q_c^*) = \int_{q_c^*}^{\infty} \left[\int_{q_c^*}^{\infty} D(q_T \,|q_c^*) h_{q_T}(q_T \,|q)\, dq_T \right] f_q(q)\, dq \tag{8.36}$$

Incorporation of hydrologic inherent and parameter and hydraulic uncertainties. To include hydrologic inherent and parameter uncertainties along with the hydraulic uncertainty associated with the flow-carrying capacity, the annual expected damage cost can be written as

$$E_4(D) = \int_0^{\infty} \left\{ \int_{q_c}^{\infty} \left[\int_{q_c}^{\infty} D(q_T \,, q_c) h_{q_T}(q_T)\, dq_T \right] f_q(q)\, dq \right\} g_{q_c}(q_c)\, dq_c$$

$$= \int_0^{\infty} E_2(D) g_{q_c}(q_c)\, dq_c \tag{8.37}$$

Summary. Based on the preceding formulations for computing annual expected damage in risk-based design of hydraulic structures, one realizes that the mathematical complexity increases as more uncertainties are considered. However, to obtain an accurate estimation of annual expected damage associated with the structural failure would require the consideration of all uncertainties, if such can be practically accomplished. Otherwise, the annual expected damage would, in most cases, be underestimated, leading to inaccurate optimal design. In an application to flood levee design (Tung, 1987), numerical investigations indicate that without providing a full account of uncertainties in the analysis, the resulting annual expected damage is significantly underestimated, even with a 75-year-long flood record.

8.3.5 Risk-based design without flood damage information

Conventional risk-based design and analysis of hydrosystems requires information with regard to various flood-related damages. Such information requires an extensive survey of the type and value of various properties, economic and

social activities, and other demographic-related information in the regions that are affected by floods. For areas where flood-related damage data are unavailable, conventional risk-based analysis cannot be implemented, realizing that in any design or analysis of a hydrosystem one normally has to conduct hydraulic simulation to delineate the flood-affected zone and other related flow characteristics, such as water depth and flow velocity. The hydraulic characteristics, combined with property survey data, would allow estimation of flood damage for a specified flood event under consideration. In the situation where flood-related damage data are unavailable, the risk-based analysis of relative economic merit of different flood defense systems still can be made by replacing the flood-related damage functions with relevant physical performance characteristics of the hydrosystems that are either required inputs for hydraulic modeling or can be extracted easily from model outputs. For example, useful physical performance characteristics in urban drainage system design and analysis could be pipe length (or street area) subject to surcharge, volume of surcharged water, and maximum (or average) depth and velocity of overland flow. Although these performance characteristics may not completely reflect what the flood damages are, they nevertheless provide a good indication about the potential seriousness of the flooding situation.

For a given design, the corresponding annual installation cost can be estimated. Also, the system responses under the different hydrologic loadings can be obtained by a proper hydraulic simulation model. Based on the annual project installation cost of the system and the expected hydraulic response of the system, a tradeoff analysis can be performed by examining the marginal improvement in hydraulic responses owing to a one-unit increase in capital investment. Referring to Fig. 8.12 for a study to upgrade the level of protection for an urban drainage system in Hong Kong (Tung and So, 2003), it is observed that the annual expected surcharge volume decreases as the annual capital cost of the system increases owing to increasing level of protection.
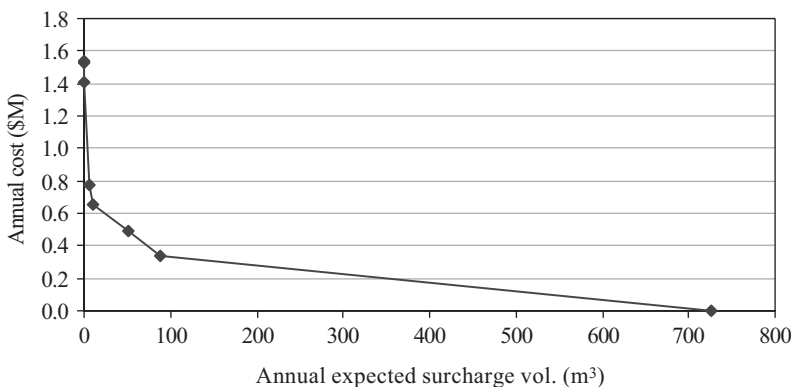
Figure 8.12   Annual project cost versus annual expected surcharge volume. (*After Tung and So, 2003.*)

The *marginal cost MC* corresponding to one reduction in surcharge volume can be written as $MC = -\partial C/\partial S_v$, with C being the capital cost and S_v being the surcharge volume. As can be seen, the value of MC starts very low for the existing system and increases to an annual capital cost around HK\$0.6M (which corresponding to a 10-year protection), beyond which the rate of increase in capital investment per unit reduction in surcharge volume becomes very high. From the trend of marginal cost, a decision maker would be able to choose a sensible level of protection for project implementation.

8.3.6 Risk-based design considering intangible factors

Besides the economic factors that can be quantified in monetary terms in the design of hydrosystems, there are other intangible factors that are noncommensurable and cannot be quantified. Some of the intangible factors might work against the principle of economic efficiency. Examples of intangible factors that are considered in the design and planning of hydrosystems may be potential loss of human lives, stability of water course, impacts on local society and environment, health hazards after floods, litigation potential, maintenance frequency of the systems, and others. The conventional optimal risk-based design yields the most economically efficient system, which may not be acceptable or feasible when other intangible factors are considered.

As more intangible factors are considered in risk-based design, it becomes a multiobjective or *multicriteria decision-making* (MCDM) problem in which economic efficiency is one of many factors to be considered simultaneously. Use of a multiple-criteria approach enhances more realistic decision making, and the design frequency so determined will be more acceptable in practice and defensible during litigation or negotiation with others. Tung et al. (1993) adopted the MCDM framework to incorporate intangible factors in risk-based design of highway drainage structures through which a more defensible extended risk-based design frequency can be determined from integrated consideration of tangible and intangible factors.

In a risk-based design, in addition to quantitative measure of failure probability and risk cost, consideration of intangible factors and societally acceptable risk issues should be included if possible. In the United States, the societally acceptable frequency of flood damage was formally set to once on average in 100 years (the so-called 100-year flood) in the Flood Disaster and Protection Act of 1973; however, the 100-year flood had been used in engineering design for many years before 1973. In this act, the U.S. Congress specified the 100-year flood as the limit of the flood plain for insurance purposes, and this has become widely accepted as the standard of hazard (Linsley and Franzini, 1979, p. 634). This acceptable hazard frequency was to be applied uniformly throughout the United States, without regard to the vulnerability of the surrounding land. The selection was not based on a benefit-cost analysis or an evaluation of probable loss of life. Linsley (1986) indicated that the logic for this fixed level of flood hazard (implicit vulnerability) was that everyone should have the same level of protection. Linsley further pointed out that many hydrologists readily accept

the implicit vulnerability assumption because a relatively uncommon flood is used for the hazard level, and thus

> The probability that anyone will ever point a finger and say "you were wrong" is equally remote. If the flood is exceeded, it is obvious that the new flood is larger than the 10-year or 100-year flood, as the case may be. If the estimate is not exceeded, there is no reason to think about it.

Mitigation of natural hazards requires a more rigorous consideration of the risk resulting from the hazard and society's willingness to accept that risk.

In other cases of disaster, societally acceptable hazard levels also have been selected without formal evaluation of benefits and costs. For example, in the United States, dam-failure hazards are mitigated by designing dams where failure may result in the loss of life to pass the probable maximum flood. Also, in The Netherlands, coastal-protection works normally are designed by application of a semideterministic worst-case approach wherein the maximum storm-surge level (10,000-year storm surge) is assumed to coincide with the minimum interior water level.

In the design of the Eastern Schedlt Storm-Surge Barrier, the Delta Committee in The Netherlands applied a simple risk-cost (in terms of lives) evaluation to set the design safety level. The Delta Committee set the total design load on the storm-surge barrier at the load with an exceedance probability 2.5×10^{-4} per year (i.e., the 4000-year water level) determined by integration of the joint probability distribution among storm-surge levels, basin levels, and the wave-energy spectrum. A single-failure criterion then was developed for the functioning of all major components of the storm-surge barrier (concrete piers, steel gates, foundation, sill, etc.) under the selected design load. The failure criterion was tentatively established at 10^{-7} per year on the basis of the following reasoning. Fatality statistics for The Netherlands indicate that the average probability of death resulting from an accident is 10^{-4} per year. Previous experience has shown that the failure of a sea-defence system may result in 10^3 casualties. Thus a normal safety level can be guaranteed only if the probability of failure of the system is less than or equal to 10^{-7} per year. Comparison of the worst-case approach with the probabilistic-load approach resulted in a 40 percent reduction in the design load when the actual, societally acceptable protection failure hazard was considered (Vrijling, 1993). This illustrates that when a comprehensive risk assessment is performed, societally acceptable safety can be maintained (and in some cases improved) while at the same time effectively using scarce financial resources. Some work on societally acceptable risk and intangible factors can be found elsewhere (Jonkman et al., 2003; Vrijling et al., 1995).

8.4 Applications of Risk-Based Hydrosystem Design

In this section, two examples are described to illustrate the applications of risk-based design of hydrosystems. One is pipe culverts for highway drainage, and the other is flood-damage-reduction projects implemented by the

U.S. Army Corps of Engineers. The first example involves optimal risk-based design considering only hydrologic inherent uncertainty, whereas the second example considers uncertainties from hydraulic and economic aspects.

8.4.1 Optimal risk-based pipe culvert for roadway drainage

The basic functions of highway drainage structures are (1) as hydraulic facilities to safely convey floods under highways during all but the most severe flooding conditions and (2) as portions of the highway to move highway traffic freely over stream channels. There are three general types of drainage structures: bridges, box culverts, and pipe culverts. Conventionally, *bridges* refer to structures measuring more than 20 ft along the roadway centerline (AASHTO, 1979). *Box culverts* are usually built of concrete with rectangular openings. *Pipe culverts* can be in various geometric forms, such as circular, arch, etc., and can be made of several different materials, such as steel, cast iron, concrete, or plastic.

The design of highway drainage structures involves both hydraulic design and structural design. The discharge associated with the critical flood that starts to cause hazards to life, property, and stream stability is termed as the *hydraulic design discharge*. The process to select the design discharge and to perform the necessary hydraulic computations for a proposed highway structure is called *hydraulic design*. In practice, the design discharge is one-to-one related to the design frequency through frequency analysis. Therefore, the design event also can be characterized by the design frequency. In this example, the *design frequency* refers to an annual exceedance probability or its reciprocal, the design return period.

The example problem under consideration is to design a circular culvert under a two-lane highway. The culvert is 100 ft long. The equivalent average daily traffic is 3000 vehicles per day. The discount rate used is 7.125 percent, and the useful service life of the culvert structure is estimated to be 35 years. Detailed descriptions of this example are given by Tung and Bao (1990).

In this example, only the inherent hydrologic and parameter uncertainties are considered. The primary objectives are (1) to search for the optimal design parameters associated with the minimum total annual expected cost for the culvert and (2) to investigate the sensitivity of the optimal design parameters with respect to (a) the hydrologic parameter uncertainty, (b) the length of streamflow records, (c) the distribution model of flood flow, and (d) the maximum flood-damage cost. More specifically, the optimal design parameters considered in this example are the optimal design return period T and the associated least total annual expected cost (LTAEC).

The estimated sample mean and sample standard deviation for the flood flow are 47.9 and 71.9 ft^3/s, respectively. The skewness coefficient of streamflow for the original scale and log-transformed scale are assumed to be 0.5 and 0.2, respectively.

In the sensitivity analysis, the optimal total annual expected cost was calculated for various record lengths n of 10, 20, 40, 60, and 100 years; for maximum

flood damage cost D_{max} of \$928, \$1500, \$3500, and \$4500; and for flood flow distribution models of normal, lognormal, Pearson type 3, and log-Pearson type 3 probability distributions.

The cost function representing the annual installation (first) cost of the culvert is derived on the basis of data from Corry et al. (1980) using regression analysis:

$$\text{AFC} = 1.0215 - 2.62 \times 10^{-7} q_c^2 \tag{8.38}$$

where AFC is the annual first cost (\$), and q_c is the design discharge (ft^3/s). The R^2 of this regression equation is 0.976.

The *damage function* $D(q)$, approximating the original discrete form in Corry et al. (1980) by a continuous function, can be expressed as

$$D(q|q_c) = \begin{cases} D_{max} & q \geq q_{max} \\ D_{max}\left(\dfrac{q - q_c}{q_{max} - q_c}\right) & q_c \leq q \leq q_{max} \\ 0 & q \leq q_c \end{cases} \tag{8.39}$$

where D_{max} is the maximum flood damage cost, q_{max} is the flood magnitude corresponding to D_{max}, and q_c is the design discharge. It is understood that, in general, q_{max} will be increased as a result of raising the design discharge q_c. The rate of increase in q_{max} will slow down as q_c increases. The damage function used, for illustration, is shown in Fig. 8.13, in which q_{max} is determined from

$$q_{max} = 210 + q_c - q_c^{0.94} \tag{8.40}$$
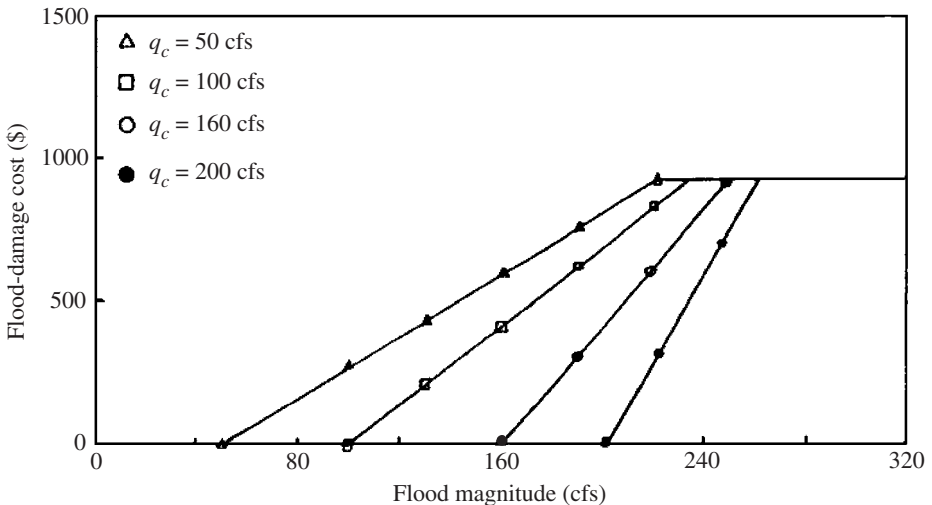
Figure 8.13 Flood-damage function in risk-based culvert design example.

Because of the complexity of the functional form of the objective function, it is difficult to solve Eqs. (8.31) and (8.33) analytically. Therefore, optimum search techniques are useful to solve the problem. However, gradient search techniques are inappropriate for use in this case because the gradient of the objective function is not easily computable. Among the search techniques that do not require knowledge of gradient of the objective function, Fibonacci search is an efficient technique to be used for this single-decision variable-optimization problem (Sivazlian and Stanfel, 1974).

Fibonacci search applies the sequential search strategy that successively reduces the feasible decision variable interval to $1/F_N$ its original size with just N function evaluations. The final decision-variable interval can be made as close to the optimal solution as the desired accuracy. F_N is called the Nth *Fibonacci number* in the Fibonacci sequence F_N, for $i = 0, 1, 2, 3, \ldots$, whose value is given by the recurrence relation

$$F_0 = F_1 = 1$$
$$F_{i+1} = F_i + F_{i-1} \qquad i \geq 1$$

(8.41)

The computational procedures for determination of the optimal return period corresponding to the optimal capacity in the risk-based design of a pipe culvert considering the hydrologic inherent and parameter uncertainties is illustrated in Fig. 8.14.

The optimal design frequency T^* and the associated LTAEC under different record lengths n and streamflow probability distributions with maximum flood damage cost ($D_{\max} = \$928$) are listed in Table 8.3. The values in the columns for $n = 10$–100 are calculated by considering hydrologic parameter uncertainty, whereas the values in the column with $n = \infty$ were calculated without considering hydrologic parameter uncertainty.

Comparing the two design methods, the value of the LTAEC without considering parameter uncertainty is always smaller than the value considering parameter uncertainty regardless of the probability distributions or values of $D_{\max}$. This observation shows that neglect of the hydrologic parameter uncertainty could lead to an underestimation of the total expected cost.

The value of LTAEC decreases as the record length increases. This is expected because the effect of hydrologic parameter uncertainty involved in estimating the second cost diminishes as the record length for streamflow gets longer. The difference in LTAEC values calculated by the two methods, for $D_{\max} = \$928$ and $n > 20$, is only about 3 percent for any of the four probability distributions considered. However, the higher the value of $D_{\max}$, the more dominant the second cost becomes in the objective function evaluation. Therefore, the difference in LTAEC values by the two methods at the same record length will be larger as $D_{\max}$ is increased.

Examining the T^* values in Table 8.3, the difference in T^* between the two methods is less than 20 percent in most cases. Also, for fixed distribution and sample size, the optimal T^* increases as $D_{\max}$ increases (see Table 8.4).

```
┌─────────────────────────────────┐
│ Input flood distribution fq(q), │
│ record length n, and cost        │
│ functions.                       │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│ Select a design return period   │◄──────┐
│ T° by Fibonacci sequence.        │       │
└─────────────────────────────────┘       │
                │                          │
                ▼                          │
┌─────────────────────────────────┐       │
│ Calculate design discharge      │       │
│ qc = f(T°) and first cost FC(T°) │       │
└─────────────────────────────────┘       │
                │                          │
                ▼                          │
┌─────────────────────────────────┐       │
│ Calculate average damage cost   │       │
│ D(T) for different frequencies  │       │
│ T by Eq. (8.35)                  │       │
└─────────────────────────────────┘       │
                │                          │
                ▼                          │
┌─────────────────────────────────┐       │
│ Calculate annual expected       │       │
│ damage cost for T° by Eq. (8.36)│       │
└─────────────────────────────────┘       │
                │                          │
                ▼                          │
┌─────────────────────────────────┐       │
│ Calculate total annual expected │       │
│ cost for T° TAEC(T°)             │       │
└─────────────────────────────────┘       │
                │                          │
                ▼                          │
          ◇─────────────◇      No          │
         < Last design T? >───────────────┘
          ◇─────────────◇
                │ Yes
                ▼
┌─────────────────────────────────┐
│ Determine optimal T associated  │
│ with the least total annual     │
│ expected cost.                   │
└─────────────────────────────────┘
```

Figure 8.14   Flowchart of optimal risk-based design of a pipe culvert.

This confirms the original intuition. However, there does not exist the same consistent tendency in T^* as with the LTAEC shown earlier. Therefore, when T^* is considered as a criterion in the comparison of the two design methods, it is difficult to conclude which method tends to be more conservative.

Figure 8.15 shows of the total annual expected cost function, annual first-cost function, and annual second-cost function versus the design return period T with record length varying from 10 to 100 years at $D_{max} = \$4500$ for the log-normal probability distribution. Similar behavior was observed for three other

TABLE 8.3   Optimal Design Return Period T^* and LTAEC for Different Distributions and Record Lengths When D_{max} = $928

Flood distribution	Optimal design	Record length (in years)					
		10	20	40	60	100	∞
Normal	T^* (years)	4.82	4.62	4.55	4.55	4.20	4.00
	LTAEC ($)	473.5	461.7	456.2	454.7	453.5	448.1
Lognormal	T^* (years)	5.79	6.00	6.27	6.55	6.20	6.96
	LTAEC ($)	446.7	433.9	428.1	425.9	423.2	420.2
Pearson type 3	T^* (years)	4.52	4.62	4.55	4.55	4.55	4.00
	LTAEC ($)	479.9	468.3	462.6	461.0	459.8	454.1
Log-Pearson type 3	T^* (years)	5.79	6.00	6.20	6.41	6.13	6.68
	LTAEC ($)	450.5	438.0	432.3	430.2	427.4	424.6

SOURCE: After Tung and Bao (1990).

types of distributions. From Fig. 8.15 it is clear that the annual second cost (ASC) and the total annual expected cost (TAEC) decrease as the record length increases. Therefore, the LTAEC will be smaller when the record length gets longer. However, the corresponding T^*, as discussed earlier, may not necessarily

TABLE 8.4   List of Optimal Design Return Period (T^*) Under Different Record Lengths, Flood Distributions, and Maximum Flood Damage

D_{max}	Flood distribution	Record length (in years)					
		10	20	40	60	100	∞
$928	N	4.82	4.62	4.55	4.55	4.20	4.00
	LN	5.79	6.00	6.27	6.55	6.20	6.96
	P3	4.62	4.62	4.55	4.55	4.55	4.00
	LP3	5.79	6.00	6.20	6.41	6.13	6.68
$1500	N	6.62	6.75	6.37	7.03	6.68	6.96
	LN	7.51	7.99	7.99	7.79	7.58	7.03
	P3	6.41	6.62	7.03	6.75	6.48	6.96
	LP3	7.37	6.68	7.86	7.51	7.03	7.03
$2500	N	10.47	8.13	9.51	8.48	8.75	9.17
	LN	10.41	10.61	10.06	9.44	9.30	10.96
	P3	9.17	9.72	8.89	8.13	7.79	7.03
	LP3	9.85	9.92	9.51	9.44	8.75	10.34
$3500	N	13.51	12.75	12.89	13.30	12.54	11.03
	LN	12.27	12.27	11.37	11.65	12.13	11.03
	P3	11.71	11.65	11.16	11.44	11.37	11.03
	LP3	11.51	11.58	11.30	11.44	11.65	11.03
$4500	N	17.98	17.71	16.26	14.75	15.71	12.89
	LN	14.61	13.92	13.85	13.92	13.64	11.03
	P3	14.26	13.78	13.92	13.85	12.68	11.03
	LP3	13.57	13.09	13.37	13.02	12.61	11.03

NOTE: N = normal; LN = lognormal; P3 = Pearson type 3; LP3 = log-Pearson type 3
SOURCE: After Tung and Bao (1990).
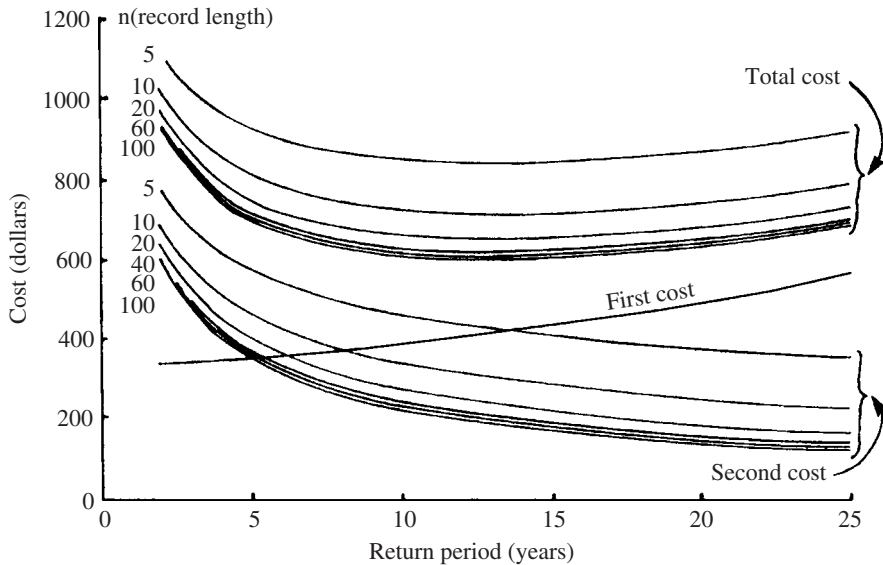
Figure 8.15 Total annual expected costs in optimal risk-based design of pipe culvert with various record lengths under lognormal distribution. (*After Tung and Bao, 1990.*)

become smaller as the record length increases. The inconsistent behavior between T^* and LTAEC in comparing the two design methods is mainly attributed to the nonlinear and nonmonotonic relationship between T^* and LTAEC.

It can be seen from Fig. 8.15 that the total annual expected cost (TAEC) curves are very flat in a range of design frequencies from 5 to 20 years for this example. Therefore, from a practical point of view, a pipe culvert could be overdesigned about 5 to 10 years above the optimal design frequency to give more confidence in the safety protection of the structure with only a small fraction of extra annual capital investment.

8.4.2 Risk-based analysis for flood-damage-reduction projects

A flood-damage-reduction plan includes measures that decrease damage by reducing discharge, stage, and/or damage susceptibility (U.S. Army Corps of Engineers, 1996). For federal projects in the United States, the objective of the plan is to solve the problem under consideration in a manner that will "...contribute to national economic development (NED) consistent with protecting the Nation's environment, pursuant to national environmental statutes, applicable executive orders, and other Federal planning requirements" (U.S. Water Resources Council, 1983). In the flood-damage-reduction planning traditionally done by the U.S. Army Corps of Engineers (Corps), the level of protection provided by the project was the primary performance indicator (Eiker and Davis, 1996). Only projects that provided a set level of protection (typically from the 100-year flood) would be evaluated to determine their contribution to NED, effect on the environment, and other issues. The level of protection was set

without regard to the vulnerability level of the land to be protected. In order to account for uncertainties in the hydrologic and hydraulic analyses applied in the traditional method, safety factors, such as freeboard, are applied in project design in addition to achieving the specified level of protection. These safety factors were selected from experience-based rules and not from a detailed analysis of the uncertainties for the project under consideration.

The Corps now requires risk-based analysis in the formulation of flood-damage-reduction projects (Eiker and Davis, 1996). In this risk-based analysis, each of the alternative solutions for the flooding problem is evaluated to determine the expected net economic benefit (benefit minus cost), expected level of protection on an annual basis and over the project life, and other decision criteria. These expected values are computed with explicit consideration of the uncertainties in the hydrologic, hydraulic, and economic analyses used in plan formulation. The risk-based analysis is used to formulate the type and size of the optimal plan that will meet the study objectives. The Corps policy requires that this plan be identified in every flood-damage-reduction study. This plan may or may not be the recommended plan based on "additional considerations" (Eiker and Davis, 1996). These additional considerations include environmental impacts, potential for fatalities, and acceptability to the local population.

In the traditional approach to planning flood-damage-reduction projects, a discharge-frequency relation for the project site is obtained through frequency analysis at or near gauge locations or through frequency transposition, regional frequency relations, rainfall-runoff models, or other methods described by the U.S. Army Corps of Engineers (1996) at ungauged stream locations. Hydraulic models are used to develop stage-discharge relations for the project location. Typically, one-dimensional steady flows are analyzed with a standard step-backwater model, but in some cases complex hydraulics are simulated using an unsteady-flow model or a two-dimensional flow model. Stage-damage relations are developed from detailed economic evaluations of primary land uses in the flood plain, as described in U.S. Army Corps of Engineers (1996). Through integration of the discharge-frequency, stage-discharge, and stage-damage relations, a damage-frequency relation is obtained. By integration of the damage-frequency relations for without-project and various with-project conditions, the damages avoided by implementing the various projects on an average annual basis can be computed. These avoided damages constitute the primary benefit of the projects, and by subtracting the project cost (converted to an average annual basis) from the avoided damages, the net economic benefit of the project is obtained.

The traditional approach to planning of flood-damage-reduction projects seeks to maximize net economic benefits subject to the constraint of achieving a specified level of protection. That is, the flood-damage-reduction alternative that maximizes net economic benefits and provides the specified level of protection would be the recommended plan unless it was unacceptable with respect to the additional considerations.

Risk-based analysis offers substantial advantages over traditional methods because it requires that the project resulting in the maximum net economic

benefit be identified without regard to the level of protection provided. There-fore, the vulnerability (from an economic viewpoint) of the flood-plain areas affected by the project is considered directly in the analysis, whereas environmental, social, and other aspects of vulnerability are considered through the additional considerations in the decision-making process. In the example presented in the Corps manual on risk-based analysis (U.S. Army Corps of Engineers, 1996), the project that resulted in the maximum net economic benefit provided a level of protection equivalent to once, on average, in 320 years. However, it is possible that in areas of low vulnerability, the project resulting in the maximum net economic benefit could provide a level of protection less than once, on average, in 100 years. A more correct level of protection is computed in the risk-based analysis by including uncertainties in the probability model of floods and the hydraulic transformation of discharge to stage rather than accepting the expected hydrologic frequency as the level of protection. This more complete computation of the level of protection eliminates the need to apply additional safety factors in the project design and results in a more correct computation of the damages avoided by implementation of a proposed project.

Monte Carlo simulation is applied in the risk-based analysis to integrate the discharge-frequency, stage-discharge, and stage-damage relations and the respective uncertainties. These relations and the respective uncertainties are shown in Fig. 8.16. The uncertainty in the discharge-frequency relation is determined by the methods used to compute confidence limits described by the Interagency Advisory Committee on Water Data (1982), which are reviewed in Sec. 3.8. For gauged locations, the uncertainty is determined directly from the gauge data; for ungauged locations, the probability distribution is fit to the estimated flood quantiles, and an estimated equivalent record length is used to compute uncertainty through the confidence-limits approach. The uncertainty in the stage-discharge relation is determined from gauge data, if available, calibration results if a sufficient number of high-water marks are available, or Monte Carlo simulation considering the uncertainties in the component input variables (Manning's n and cross-sectional geometry) for the hydraulic model (e.g., U.S. Army Corps of Engineers, 1986). The uncertainty in the stage-damage relation is determined by using Monte Carlo simulation to aggregate the uncertainties in components of the economic evaluation. At present, uncertainty distributions for structure elevation, structure value, and contents value are considered in the analysis.

The Monte Carlo simulation procedure for the risk-based analysis of flood-damage-reduction alternatives includes the following steps applied to both without-project and with-project conditions (U.S. Army Corps of Engineers, 1996):

1. A value for the expected exceedance (or nonexceedance) probability is selected randomly from a uniform distribution. This value is converted into a random value of flood discharge by inverting the expected flood-frequency relation.
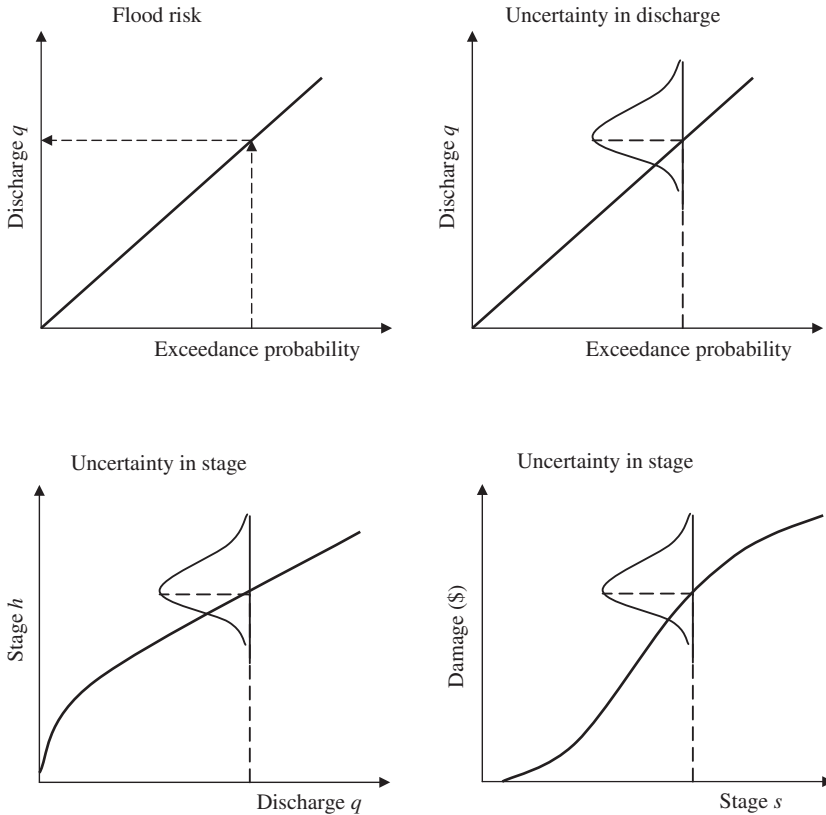
Figure 8.16 Uncertainty in discharge, stage, and damage as considered in the U.S. Army Corps of Engineers risk-based approach to flood-damage reduction studies. (*After Tseng et al., 1993*.)

2. A value of a standard normal variate is selected randomly, and it is used to compute a random value of error associated with the flood discharge obtained in step 1. This random error is added to the flood discharge obtained in step 1 to yield a flood-discharge value that includes the effect of uncertainty in the probability model of floods.

3. The flood discharge obtained in step 2 is converted to the expected flood stage using the expected stage-discharge relation.

4. A value of a standard normal variate is selected randomly, and it is used to compute a random value of error associated with the flood stage computed in step 3. This random error is added to the flood stage computed in step 3 to yield a flood stage that includes the effects of uncertainty in the stage-discharge relation and the probability model of floods. If the performance of a proposed project is being simulated, level of protection may be determined empirically by counting the number of flood stages that are higher than the project capacity and dividing by the number of simulations.

5. The flood stage obtained in step 4 is converted to the expected flood damage using the expected flood-damage relation. If the performance of a proposed project is simulated, the simulation procedure may end here if the simulated flood stage does not result in flood damage.

6. A value of a standard normal variate is selected randomly, and it is used to compute a random value of error associated with the flood damage obtained in step 5. This random error is added to the flood damage obtained in step 5 to yield a flood-damage value that includes the effects of all the uncertainties considered. If the flood-damage value is negative, it should be set equal to zero.

Steps 1 through 6 are repeated as necessary until the values of the relevant performance measures (average flood damage, level of protection, probability of positive net economic benefits) stabilize to consistent values. Typically, 5000 simulations are used in Corps projects.

The risk-based approach, summarized in steps 1 through 6, has many similarities with traditional methods, particularly in that the basic data and discharge-frequency, stage-discharge, and stage-damage relations are the same. The risk-based approach extends traditional methods to consider uncertainties in the basic data and relations. The major new task in the risk-based approach is to estimate the uncertainty in each of the relations. Approaches to estimate these uncertainties are described in detail by the U.S. Army Corps of Engineers (1996) and are not trivial. However, the information needed to estimate uncertainty in the basic components variables is often collected in traditional methods but not used. For example, confidence limits often are computed in flood-frequency analysis, error information is available for calibrated hydraulic models, and economic evaluations typically are done by studying in detail several representative structures for each land-use category, providing a measure of the variability in the economic evaluations. Therefore, an excessive data-analysis burden relative to traditional methods may not be imposed on engineers and planners in risked-based analysis.

Because steps 1 through 6 are applied to each of the alternative flood-damage-reduction projects, decision makers will obtain a clear picture of the tradeoff among level of protection, cost, and benefits. Further, with careful communication of the results, the public can be better informed about what to expect from flood-damage-reduction projects and thus can make better-informed decisions (U.S. Army Corps of Engineers, 1996).

8.5   Optimization of Hydrosystems by Chance-Constrained Methods

In all fields of science and engineering, the decision-making process depends on several parameters describing system behavior and characteristics. More often than not, some of these system parameters cannot be assessed with certainty. In a system-optimization model, if some of the coefficients in the constraints

are uncertain, the compliance with the constraints, under a given set of solutions, cannot be ensured with certainty. Owing to the random nature of the constraint coefficients, a certain likelihood that constraints will be violated always exists. The basic idea of *chance-constrained methods* is to find the solution to an optimization problem such that the constraints will be met with a specified reliability. Chance-constrained formulations have been applied to various types of water resource problems such as groundwater quantity management (Tung, 1986), groundwater quality management (Gorelick, 1982; Wagner and Gorelick, 1987, 1989; Morgan et al., 1993; Ritzel et al., 1994) and monitoring network design (Datta and Dhiman, 1996), reservoir operation (Loucks et al., 1981; Houck, 1979; Datta and Houck, 1984), waste-load allocation (Lohani and Thanh, 1978; Fujiwara et al., 1986, 1987; Ellis, 1987; Tung and Hathhorn, 1990), water distribution systems (Lansey et al., 1989), and freshwater inflow for estuary salinity management (Tung et al., 1990; Mao and Mays, 1994). This section describes the basic properties of chance-constrained models. In the next section an application to waste-load allocation is presented for illustration.

Refer to the general nonlinear optimization problem as stated in Eqs. (8.1a–c). Consider a constraint $g(\boldsymbol{x}) \leq b$, with $\boldsymbol{x}$ being a vector of decision variables. In general, decision variables $\boldsymbol{x}$ in an optimization model are controllable without uncertainty. Suppose that some of the parameters on the left-hand-side (LHS) of the constraint $g(\boldsymbol{x})$ and/or the right-hand-side (RHS) coefficient b are subject to uncertainty. Because of the uncertainty, the compliance with the constraint under a given solution set $\boldsymbol{x}$ cannot be ensured with absolute certainty. In other words, there is a possibility that for any solution $\boldsymbol{x}$, the constraint will be violated. Consequently, the chance-constrained formulation expresses the original constraint in a probabilistic format as

$$P\left[g(\boldsymbol{x}) \leq b\right] \geq \alpha \tag{8.42}$$

where $P[\cdot]$ is the probability and α is the specified reliability for constraint compliance. Since this chance-constrained formulation involves probability, it is not mathematically operational for algebraic solution. For this reason, the *deterministic equivalent* must be derived. There are three cases in which the random elements in Eq. (8.42) could occur: (1) only elements in $g(\boldsymbol{x})$ are random, (2) only the RHS b is random, and (3) both $g(\boldsymbol{x})$ and b are random.

The simplest case is the case 2, where only the RHS coefficient b is random. The derivation of deterministic equivalent of the chance-constraint for this case can be done as follows: The constraint can be rewritten as

$$P\left[g(\boldsymbol{x}) \leq B\right] \geq \alpha \tag{8.43}$$

where B is a random RHS coefficient. Since Eq. (8.43) can be written as

$$P\left[g(\boldsymbol{x}) \geq B\right] \leq 1 - \alpha \tag{8.44}$$

then Eq. (8.44) can be expressed as

$$F_B[g(\boldsymbol{x})] \le 1 - \alpha \tag{8.45}$$

in which $F_B[\cdot]$ is the cumulative distribution function (CDF) of random RHS B. The deterministic equivalent of the original chance-constraint Eq. (8.43) can be expressed as

$$g(\boldsymbol{x}) \le b_{1-\alpha} \tag{8.46}$$

where, referring to Fig. 8.17, $b_{1-\alpha}$ is the $(1-\alpha)$th quantile of the random RHS coefficient B satisfying $P(B \le b_{1-\alpha}) = 1 - \alpha$. If the RHS coefficient B is a normal random variable with mean μ_b and standard deviation σ_b, Eq. (8.46) can be written as

$$g(\boldsymbol{x}) \le \mu_b + z_{1-\alpha}\sigma_b \tag{8.47}$$

with $z_{1-\alpha}$ being the $(1-\alpha)$th standard normal quantile.

In the case that only the elements in $g(\boldsymbol{x})$ are random and the distribution functions are known, the chance-constraint can be expressed as

$$P[G(\boldsymbol{x}) \le b] \ge \alpha \tag{8.48}$$

For a general nonlinear function $G(\boldsymbol{x})$, the difficulty lies in the derivations of exact probability distribution and statistical moments of $G(\boldsymbol{x})$ as functions of unknown decision variables. In this circumstance, statistical moments of $G(\boldsymbol{x})$ can be estimated by uncertainty-analysis methods such as those described in Tung and Yen (2005). The assessment of the distribution for $G(\boldsymbol{x})$ is, at best, to be made subjectively. For the sake of discussion, assume that the distribution function of $G(\boldsymbol{x})$ is known. The deterministic equivalent of the chance-constraint Eq. (8.48) can be expressed as

$$F_{G(\boldsymbol{x})}^{-1}(\alpha) \ge b \tag{8.49}$$
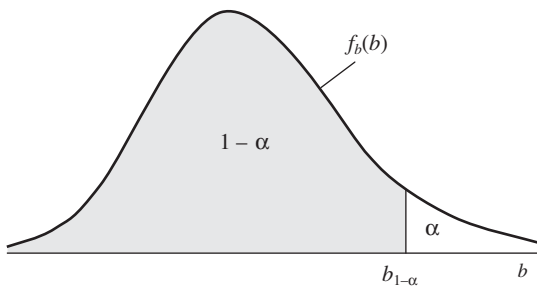
Figure 8.17 Probability density function of the random right-hand-side coefficient B.

where $F_{G(\boldsymbol{x})}^{-1}(\alpha)$ is the αth quantile of the random $G(\boldsymbol{x})$, which is the function of unknown decision variables $\boldsymbol{x}$. In general, $F_{G(\boldsymbol{x})}^{-1}(\alpha)$ in Eq. (8.48) is a nonlinear equation of $\boldsymbol{x}$ even if $G(\boldsymbol{x})$ is a linear function of $\boldsymbol{x}$, as will be shown later.

The third case is when both $G(\boldsymbol{x})$ and the RHS coefficient B are random. The chance-constraint then can be expressed as

$$P\left[G(\boldsymbol{x}) - B \le 0\right] \ge \alpha \tag{8.50}$$

The deterministic equivalent of Eq. (8.49) can be derived as

$$F_{G(\boldsymbol{x})-B}^{-1}(\alpha) \ge 0 \tag{8.51}$$

where $F_{G(\boldsymbol{x})-B}^{-1}(\alpha)$ is the inverse of the CDF of random $G(\boldsymbol{x}) - B$ taken on the value of α.

As a special case, consider an LP formulation as stated in Eq. (8.3) in which technological coefficients $\boldsymbol{A}$ and/or RHS coefficients $\boldsymbol{b}$ are subject to uncertainty. By imposing a reliability restriction $\boldsymbol{\alpha}$ on the system constraints, the LP model can be transformed into the following chance-constrained formulation:

$$\text{Maximize} \quad \boldsymbol{c}^t \boldsymbol{x} \tag{8.52a}$$

$$\text{subject to} \quad P(\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}) \ge \boldsymbol{\alpha} \tag{8.52b}$$

In a chance-constrained LP model, the elements in $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ can be random variables. When the objective function coefficient c_j's are random variables, it is common to replace them by their expected values. Consider the following three cases: (1) elements of the technological coefficient matrix (A_{ij}'s) are random variables, (2) elements of the RHS vector B_i's are random variables, and (3) elements A_{ij} and B_i are simultaneously random variables. In the following derivations, it is assumed that random technological coefficients and random RHS coefficient are correlated within a constraint and that these coefficients are uncorrelated between constraints.

Consider that the RHS of the ith constraint B_i is subject to uncertainty. Furthermore, assume that its distribution and statistical moments are known. In this case, the deterministic equivalent of the chance-constraint can be obtained easily from Eq. (8.46) as

$$\sum_{j=1}^{n} a_{ij} x_j \le b_{i, 1-\alpha_i} \qquad \text{for } i = 1, 2, \ldots, m \tag{8.53}$$

and the constraint form remains linear.

Consider the case that the technological coefficients a_{ij}'s of the ith constraint are random. The deterministic equivalent of the chance-constraint

$$P\left(\sum_{j=1}^{n} A_{ij} x_j \le b_i\right) \ge \alpha_i$$

can be derived as (Kolbin, 1977; Vajda, 1972)

$$\sum_{j=1}^{n} E(A_{ij})x_j + F_{Z_i}^{-1}(\alpha_i)\sqrt{\boldsymbol{x}^t \boldsymbol{C}_i \boldsymbol{x}} \leq b_i \tag{8.54}$$

where $E(A_{ij})$ is an expectation of the technological coefficient A_{ij}, $\boldsymbol{C}_i$ is an $n \times n$ covariance matrix of n random technological coefficients $(A_{i1}, A_{i2}, \ldots, A_{in})$ in the ith constraint, and $F_{Z_i}^{-1}(\alpha_i)$ is the appropriate quantile for the α_i percentage given by the CDF of standardized left-hand-side (LHS) terms. That is,

$$Z_i = \frac{\text{LHS}_i - E(\text{LHS}_i)}{\sqrt{\text{Var}(\text{LHS}_i)}} = \frac{\sum_{j=1}^{n} A_{ij}x_j - \sum_{j=1}^{n} E(A_{ij})x_j}{\sqrt{\boldsymbol{x}^t \boldsymbol{C}_i \boldsymbol{x}}} \tag{8.55}$$

If all A_{ij}'s are independent random variables, that is, $\rho(A_{ij}, A_{ij'}) = 0$, for $j \neq j'$, matrix $\boldsymbol{C}_i$ is a diagonal matrix of variances of A_{ij}. To quantify $F_{Z_i}^{-1}(\alpha_i)$, the distribution of LHS must be known or assumed. Note that the LHSs in an LP model are the sum of several random variables. By the central limit theorem (see Sec. 2.6.1), the random LHS can be approximated as a normal random variable. Therefore, Eq. (8.54) can be written as

$$\sum_{j=1}^{n} E(A_{ij})x_j + \Phi^{-1}(\alpha_i)\sqrt{\boldsymbol{x}^t \boldsymbol{C}_i \boldsymbol{x}} \leq b_i \tag{8.56}$$

with $\Phi(\cdot)$ being the standard normal CDF. From Eq. (8.55) one realizes that when A_{ij}'s are random, the resulting deterministic equivalents of the chance constraints are no longer linear functions of the decision variables. The chance-constrained model has to be solved by nonlinear programming algorithms. In the next subsection of application, a sequential LP algorithm is used to linearize Eq. (8.56).

Finally, when both the technological coefficients and the RHS coefficient of the ith constraint are random, the chance-constraint format, referring to Eq. (8.50), can be written as

$$P\left(\sum_{j=1}^{n} A_{ij}x_j - B_i \leq 0\right) \geq \alpha_i \tag{8.57}$$

Following the same procedure as described earlier, the deterministic equivalent of Eq. (8.57) can be derived, assuming normal distribution for the random terms in Eq. (8.57), as

$$\sum_{j=1}^{n} E(A_{ij})x_j + \Phi^{-1}(\alpha_i)\sqrt{\boldsymbol{x}^t \boldsymbol{C}_i \boldsymbol{x} + 2\sum_{j=1}^{n} x_j \text{Cov}(A_{ij}, B_i) + \sigma_{B_i}^2} \leq 0 \tag{8.58}$$

in which $\text{Cov}(A_{ij}, B_i)$ is the covariance between the random technological coefficient A_{ij} and the random RHS coefficient B_i for the ith constraint.

8.6 Chance-Constrained Water-Quality Management

Water-quality management is the practice of protecting the physical, chemical, and biologic characteristics of various water resources. Historically, such efforts have been guided toward the goal of assessing and controlling the impacts of human activities on the quality of water. To implement water-quality management measures in a conscientious manner, one must acknowledge both the activities of the society and the inherently random nature of the stream environment itself (Ward and Loftis, 1983). In particular, the environments in which decisions are to be made concerning in-stream water-quality management are inherently subject to many uncertainties. The stream system itself, through nature, is an environment abundant with ever-changing and complex processes, both physically and biologically.

Public Law 92-500 (PL 92-500) in the United States provided impetus for three essential tasks, one of which is to regulate waste-water discharge from point sources from industrial plants, municipal sewage treatment facilities, and livestock feedlots. It also requires treatment levels based on the best available technology. However, if a stream segment is *water-quality-limited*, in which the waste assimilative capacity is below the total waste discharge authorized by PL 92-500, more stringent controls may be required.

For streams under water-quality-limited conditions or where effluent standards are not implemented, the *waste-load-allocation* (WLA) problem is concerned with how to effectively allocate the existing assimilative capacity of the receiving water body among several waste dischargers without detrimental effects to the aquatic environment. As an integral part of water-quality management, WLA is an important but complex decision-making task. The results of WLA have profound implications on regional environmental protection. A successful WLA decision requires sound understanding of the physical, biologic, and chemical processes of the aquatic environment and good appreciation for legal, social, economical, and environmental impacts of such a decision.

Much of the research in developing predictive water-quality models has been based on a deterministic evaluation of the stream environment. Attempts to manage such an environment deterministically imply that the compliance with water-quality standards at all control points in the stream system can be ensured with absolute certainty. This, of course, is unrealistic. The existence of the uncertainties associated with stream environments should not be ignored. Thus it is more appropriate in such an environment to examine the performance of the constraints of a mathematical programming model in a probabilistic context. The random nature of the stream environment has been recognized in the WLA process. Representative WLA using a chance-constrained formulation can be found elsewhere (Lohani and Thanh, 1979; Yaron, 1979; Burn and McBean, 1985; Fujiwara et al., 1986, 1987; Ellis, 1987; Tung and Hathhorn, 1990).

In the context of stochastic management, the left-hand-side (LHS) coefficients of the water-quality constraints in a WLA model are functions of various random water-quality parameters. As a result, these LHS coefficients are random

variables as well. Furthermore, correlation exists among these LHS coefficients because (1) they are functions of the same water-quality parameters and (2) some water-quality parameters are correlated with each other. Moreover, the water-quality parameters along a stream are spatially correlated. Therefore, to reflect the reality of a stream system, a stochastic WLA model should account for the randomness of the water-quality parameters, including spatial and cross-correlations of each parameter.

The main objective of this section is to present methodologies to solve a stochastic WLA problem in a chance-constrained framework. The randomness of the water-quality parameters and their spatial and cross-correlations also are taken into account. A six-reach example is used to demonstrate these methodologies. Factors affecting the model solution to be examined are (1) the distribution of the LHS coefficients in water-quality constraints and (2) the spatial correlation of water-quality parameters.

8.6.1 Optimal stochastic waste-load allocation

Deterministic waste-load allocation model. Although any number of pollutants may be considered in the overall quality management of a river system, in this example, application a biochemical oxygen demand–dissolved oxygen (BOD-DO) water-quality model is considered.

In LP format, the *deterministic WLA model* considered herein can be written as

$$\text{Maximize} \quad \sum_{j=1}^{N}(B_j + D_j) \tag{8.59a}$$

subject to

1. Constraints on water quality:

$$a_{0i} + \sum_{j=1}^{n_i} \Theta_{ij} B_j + \sum_{j=1}^{n_i} \Omega_{ij} D_j \le \text{DO}_i^{\text{sat}} - \text{DO}_i^{\text{std}} \qquad \text{for } i = 1, 2, \ldots, M \tag{8.59b}$$

2. Constraints on treatment equity:

$$\left| \frac{B_j}{I_j} - \frac{B_{j'}}{I_{j'}} \right| \le E_a \qquad \text{for } j \ne j' \tag{8.59c}$$

3. Constraints on treatment efficiency:

$$\underline{e}_j \le 1 - \frac{B_j}{I_j} \le \bar{e}_j \qquad \text{for } j = 1, 2, \ldots, N \tag{8.59d}$$

where B_j, D_j, and I_j are the effluent waste concentrations (mg/L BOD), effluent DO deficit concentration (mg/L), and raw waste influent concentration (mg/L BOD) at discharge location j, respectively, and N is the total number of waste dischargers. The LHS coefficients a_{oi}, Θ_{ij}, and Ω_{ij} in Eq. (8.59b) are

the technological transfer coefficients relating impact on DO concentrations at downstream locations i resulting from the background waste and waste input at an upstream location j. These technological transfer coefficients are functions of water-quality parameters such as reaeration and deoxygenation rates, flow velocity, etc. DO_i^{std} and DO_i^{sat} represent the required DO standard and saturated DO concentration at control point i, respectively. Finally, E_a is the allowable difference (i.e., equity) in treatment efficiency between two waste dischargers, and $\underline{e}_j$ and $\bar{e}_j$ are the lower and upper bounds of waste-removal efficiency for the jth discharger, respectively. The importance of incorporating the treatment equity in the WLA problems is discussed by many researchers (Gross, 1965; Loucks et al., 1967; Miller and Gill, 1976; Brill et al., 1976; Chadderton et al., 1981).

The water-quality constraint relating the response of DO to the effluent waste can be defined by water-quality models such as the Streeter-Phelps equation (Streeter and Phelps, 1925) or its variations (Dobbins, 1964; Krenkel and Novotny, 1980). To demonstrate the proposed methodologies, the original Streeter-Phelps equation is used herein to derive the water-quality constraints. Expressions for Θ_{ij} and Ω_{ij}, based on the Streeter-Phelps equation, are shown in Appendix 8A. The Streeter-Phelps equation for DO deficit is given as follows:

$$D_x = \frac{K_d L_0}{K_a - K_d}\left(e^{-K_d x/U} - e^{-K_a x/U}\right) + D_0 e^{-K_a x/U} \tag{8.60}$$

where D_x is the DO deficit concentration (mg/L) at stream location x (mi), K_d is the deoxygenation coefficient for BOD (days^{-1}), K_a is the reaeration-rate coefficient (days^{-1}), L_0 is the BOD concentration at the upstream end of the reach (that is, $x = 0$), D_0 is the DO deficit concentration at the upstream end of the reach, and U is the average streamflow velocity in the reach (mi/day).

Chance-constrained waste-load allocation model. The deterministic WLA model presented in Eqs. (8.59a–d) serves as the basic model for deriving the stochastic WLA model. Considering the existence of uncertainty within the stream environment, the water-quality constraints given by Eq. (8.59b) can be expressed probabilistically as

$$P\left\{a_{0i} + \sum_{j=1}^{n_i} \Theta_{ij} B_j + \sum_{j=1}^{n_i} \Omega_{ij} D_j \leq DO_i^{sat} - DO_i^{std}\right\} \geq \alpha_i \quad \text{for } i = 1, 2, \ldots, M \tag{8.61}$$

Based on Eq. (8.53), the deterministic equivalent of Eq. (8.60) can be derived as

$$\sum_{j=1}^{n_i} E(\Theta_{ij}) B_j + \sum_{j=1}^{n_i} E(\Omega_{ij}) D_j + F_Z^{-1}(\alpha_i)\sqrt{(\boldsymbol{B}, \boldsymbol{D})^t \boldsymbol{C}(\boldsymbol{\Theta}_i, \boldsymbol{\Omega}_i)(\boldsymbol{B}, \boldsymbol{D})} \leq R_i' \tag{8.62}$$

in which $R_i' = \mathrm{DO}_i^{\mathrm{sat}} - \mathrm{DO}_i^{\mathrm{std}} - E(a_{0i})$, $(\boldsymbol{B}, \boldsymbol{D})$ is the column vector of BOD and DO deficit concentrations in waste effluent, and $\boldsymbol{C}(\Theta_i, \Omega_i)$ is the covariance matrix associated with the technological transfer coefficients in the ith water-quality constraint, including a_{0i}. The stochastic WLA model to be solved consists of Eqs. (8.58a), (8.62), (8.58c), and (8.58d).

Assessments of statistical properties of random technological coefficients.   To solve the stochastic WLA model, it is necessary to assess the statistical properties of the random LHS in the chance-constraint Eq. (8.62). As shown in Appendix 8A, the technological transfer coefficients Θ_{ij} and Ω_{ij} are nonlinear functions of the stochastic water-quality parameters that are cross-correlated among them within each stream reach and spatially correlated between stream reaches. Furthermore, the complexities of the functional relationships between these transfer coefficients and the water-quality parameters increases rapidly as the control point moves downstream. Hence the analytical derivation of the statistical properties of Θ_{ij} and Ω_{ij} becomes a formidable task given even a small number of reaches. As a practical alternative, simulation procedures may be used to estimate the mean and covariance structure of the random technological coefficients within a given water-quality constraint.

The assumptions made in the Monte Carlo simulation to generate water-quality parameters in all reaches of the stream system are as follows: (1) The representative values for the reaeration coefficient, deoxygenation coefficient, and average flow velocity in each reach are *second-order stationary*; i.e., the spatial covariance functions of water quality-parameters are dependent only on the "space lag" or separation distance; (2) correlation between the reaeration coefficient and average flow velocity exists only within the same stream reach; (3) background DO and BOD concentrations at the upstream end of the entire stream system are independent of each other and of all other water-quality parameters; and (4) all water-quality parameters follow a normal distribution.

In the simulation, variance-covariance matrices representing the spatial correlation of a water-quality parameter can be derived from the *variogram models* (Journel and Huijbregts, 1978) in the field of geostatistics. Three commonly used variogram models are:

1. *Transitive variogram model:*

$$\mathrm{Cov}(|\boldsymbol{h}|) = \sigma^2 \left(1 - \frac{|\boldsymbol{h}|}{h_o} \right) \tag{8.63}$$

2. *Spherical variogram model:*

$$\mathrm{Cov}(|\boldsymbol{h}|) = \sigma^2 \left[1 - \frac{3}{2} \left(\frac{|\boldsymbol{h}|}{h_o} \right) + \frac{1}{2} \left(\frac{|\boldsymbol{h}|}{h_o} \right)^3 \right] \tag{8.64}$$

3. *Gaussian variogram model:*

$$\text{Cov}(|\boldsymbol{h}|) = \sigma^2 \left[\exp \left(\frac{-|\boldsymbol{h}|^2}{2h_o^2} \right) \right] \tag{8.65}$$

in which $\text{Cov}(|\boldsymbol{h}|)$ represents the value of covariance between two measurements of the same water-quality parameter separated by a distance $|\boldsymbol{h}|$ apart, h_o is the length of zone of influence, and σ^2 is the variance of the water-quality parameter within a given reach. The value of correlation coefficient $\rho(|\boldsymbol{h}|)$ can be calculated as $\rho(|\boldsymbol{h}|) = \text{Cov}(|\boldsymbol{h}|)/\sigma^2$. When the distance between reaches exceeds h_o, the value of the covariance function goes to zero, and the corresponding correlation coefficient is zero as well. Graphically, the three variogram models are shown in Fig. 8.18.

To illustrate the concept, consider the water-quality parameters' reaeration coefficient K_a and average flow velocity U. From the variogram models, the correlation matrix for the two parameters can be constructed as follows:

$$\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{U}) = \begin{bmatrix} \boldsymbol{R}_{K_a, K_a} & \boldsymbol{R}_{K_a, U} \\ \boldsymbol{R}_{U, K_a} & \boldsymbol{R}_{U, U} \end{bmatrix} \tag{8.66}$$

in which $\boldsymbol{K}_a = (K_{a,1}, K_{a,2}, \dots, K_{a,N})^t$ and $\boldsymbol{U} = (U_1, U_2, \dots, U_N)^t$ are vectors of the reaeration coefficient and average velocity in each stream reach, respectively (see Fig. 8.19). In Eq. (8.66), $\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{K}_a)$, $\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{U})$, $\boldsymbol{R}(\boldsymbol{U}, \boldsymbol{U})$ are $N \times N$ square symmetric correlation matrices, with N being the number of stream reaches in the WLA model. Submatrices $\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{K}_a)$ and $\boldsymbol{R}(\boldsymbol{U}, \boldsymbol{U})$ define the spatial correlation of K_a and U between the reaches, whereas submatrix $\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{U})$ defines the cross-correlation between K_a and U within the same reach. Under assumption 2 previously mentioned, the submatrix $\boldsymbol{R}(\boldsymbol{K}_a, \boldsymbol{U})$ is a diagonal matrix. For water-quality parameters that are not cross-correlated with other parameters but are spatially correlated, the associated correlation matrix has a form similar to $\boldsymbol{R}(\boldsymbol{U}, \boldsymbol{U})$. For parameters that are spatially independent, their correlation matrices are in the form of an identify matrix. Once the correlation matrix of normal stochastic parameters within a reach and between reaches is established according to the specified variogram model, generation of stochastic water-quality parameters can be obtained easily by the procedures for generating multivariate random variates described in Sec. 7.5.2.

In summary, the simulation for generating spatially and cross-correlated water-quality parameters can be outlined as the follows:

1. Select an appropriate variogram model for a given water-quality parameter, and construct the corresponding covariance matrix $\boldsymbol{C}$ or correlation matrix $\boldsymbol{R}$.

2. Apply procedures described in Sec. 7.5.2 to obtain the values of the water-quality parameter for each reach in the WLA model.

3. Repeat steps 1 and 2 for all water-quality parameters.
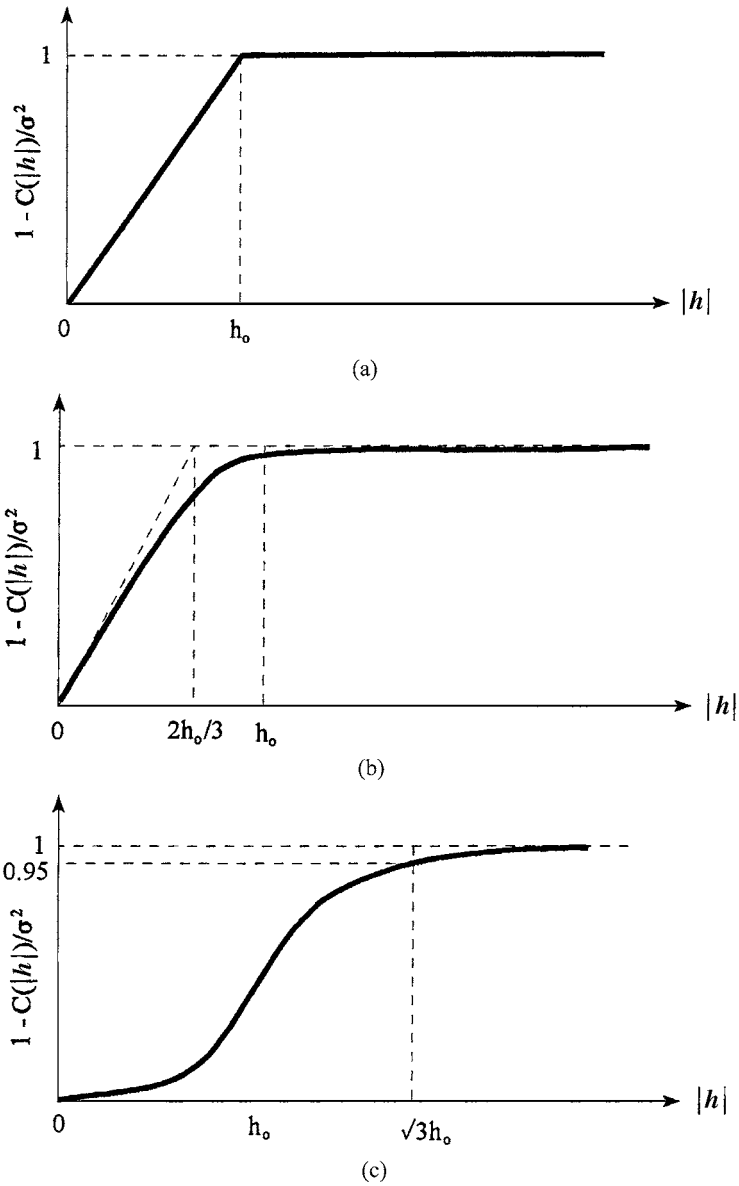
Figure 8.18   Variograms of different types: (a) transitive model; (b) spherical model; (c) Gaussian model.

$$
\begin{bmatrix}
\sigma(K_{a_1},K_{a_1}) & \sigma(K_{a_1},K_{a_2}) & \cdots & \sigma(K_{a_1},K_{a_N}) & \sigma(K_{a_1},U_1) & 0 & \cdots & 0 \\
\sigma(K_{a_2},K_{a_1}) & \sigma(K_{a_2},K_{a_2}) & \cdots & \sigma(K_{a_2},K_{a_N}) & \sigma(K_{a_2},K_{a_1}) & 0 & \cdots & 0 \\
\cdot & & \cdot & & & \cdot & \cdot & \cdot \\
\cdot & & \cdot & & & \cdot & \cdot & \cdot \\
\sigma(K_{a_N},K_{a_1}) & \sigma(K_{a_N},K_{a_2}) & \cdots & \sigma(K_{a_N},K_{a_N}) & 0 & 0 & \cdots & \sigma(K_{a_N},U_N) \\
\sigma(U_1,K_{a_1}) & 0 & \cdots & 0 & \sigma(U_1,U_1) & \sigma(U_1,U_2) & \cdots & \sigma(U_1,U_N) \\
0 & \sigma(U_2,K_{a_2}) & \cdots & 0 & \sigma(U_2,K_1) & \sigma(U_2,U_2) & \cdots & \sigma(U_2,U_N) \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & \cdots & \sigma(U_N,K_{a_N}) & \sigma(U_N,U_1) & \sigma(U_N,U_2) & \cdots & \sigma(U_N,U_N)
\end{bmatrix}
$$

Figure 8.19 Structure of covariance matrix $C(K_a, U)$ for N-reach stream system.

For each set of the water-quality parameters generated by steps 1 through 3, the values of the technological coefficients are computed. Based on the simulated values, the mean and covariance matrices of the random technological coefficients for each water-quality constraint are calculated and used in solving the stochastic WLA problem. The simulation procedure described in this subsection to account for spatial correlation is called *unconditional simulation* in the field of *geostatistics* (Journel and Huijbregts, 1978).

Technique for solving an optimal stochastic WLA model. The deterministic WLA model presented previously follows an LP format and can be solved using the simplex algorithm. However, the deterministic equivalents of the chance-constrained water-quality constraints are nonlinear. Thus the problem is one of nonlinear optimization, which can be solved by various nonlinear programming techniques mentioned in Sec. 8.1.2.

In this example, linearization of the chance-constrained water-quality constraints is done, and the linearized model is solved iteratively using the LP simplex technique. More specifically, the algorithm selects an assumed solution to the stochastic WLA model that is used to calculate the value of the nonlinear terms in Eq. (8.62). The nonlinear terms then become constants and are moved to the RHS of the constraints. The resulting linearized water-quality constraints can be written as

$$
\sum_{j=1}^{n_i} E(\Theta_{ij})B_j + \sum_{j=1}^{n_i} E(\Omega_{ij})D_j \leq R_i' - F_Z^{-1}(\alpha_i)\sqrt{(\hat{B},\hat{D})^t C(\Theta_i,\Omega_i)(\hat{B},\hat{D})}
$$

(8.67)

in which $\hat{B}$ and $\hat{D}$ are assumed solution vectors to the stochastic WLA model.

The linearized stochastic WLA model, replacing Eq. (8.62) by Eq. (8.67), can be solved using LP techniques repeatedly, each time updating the previous

solution values with those obtained from the current iteration, resulting in new values for the RHS. The procedure is repeated until convergence criteria are met between any two successive iterations. A flowchart depicting the procedures is given in Fig. 8.20. Of course, alternative stopping rules could be incorporated in the algorithm to prevent excessive iteration during the computation. Prior to the application of these solution procedures, an assumption for the distribution of the random LHS must be made to determine the appropriate value for the term $F_z(\alpha_i)$ in Eq. (8.67).

Owing to the nonlinear nature of the stochastic WLA model, the global optimum solution, in general, cannot be guaranteed. It is suggested that a few runs of the solution procedure with different initial solutions be carried out to ensure that the model solution converges to the overall optimum. A reasonable initial solution is to select the waste effluent concentration for each discharger associated with the upper bounds of their respective treatment levels. By doing so, the initial solution corresponds to waste discharge at their respective lower
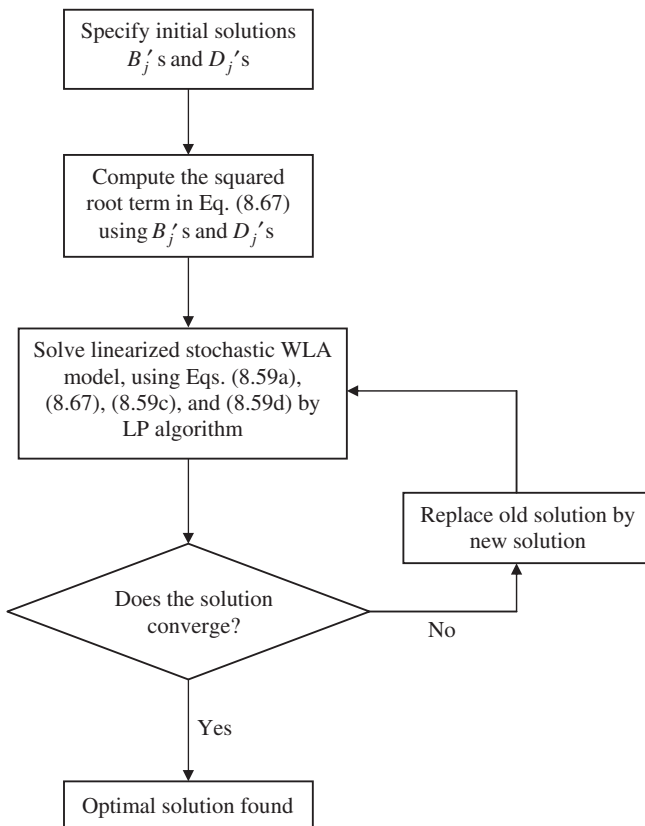
Figure 8.20 Flowchart for solving linearized chance-constrained WLA model.

limits. If the stochastic WLA solution is infeasible during the first iteration, it is likely that the feasible solution to the stochastic WLA problem does not exist. Hence time and computational effort could be saved in searching for an optimal solution that might not exist.

Numerical example.   The preceding chance-constrained WLA model is applied to a six-reach example shown in Fig. 8.21. The means and standard deviations for the water-quality parameters in each reach are given in Tables 8.5 and 8.6 based on the data reported in the literature (Churchill et al., 1962; Chadderton et al., 1982; Zielinski, 1988).

To assess the mean and correlation matrix of the random technological coefficients in the water-quality constraints, the Monte Carlo simulation procedure described in Sec. 6.5.2 is implemented to generate multivariate normal water-quality parameters. Different numbers of simulation sets are generated to examine the stability of the resulting means and covariance matrix of the technological coefficients. It was found that the statistical properties of Θ_{ij} and Ω_{ij} become stable using 200 sets of simulated parameters. In the example, a positive correlation coefficient of 0.8 between the reaeration coefficient and average flow velocity is used. Both normal and lognormal distributions are assumed for the random LHS of the water-quality constraints

$$a_{0i} + \sum_{j=1}^{n_i} \Theta_{ij} B_j + \sum_{j=1}^{n_i} \Omega_{ij} D_j \tag{8.68}$$

in Eq. (8.67). Various reliability levels α_i ranging from 0.85 to 0.99 for the water-quality constraints are considered.

Background Characteristics:
$L_o = 5.0$ mg/L
$Q_o = 115$ cfs
$D_o = 1.0$ mg/L

Discharger No. 6:
I = 410 mg/L
q = 0.78 cfs
x = 125 miles

Discharger No. 4:
I = 910 mg/L
q = 35.81 cfs
x = 75 miles

Discharger No. 2:
I = 6.0 mg/L
q = 44.0 cfs
x = 25 miles

Discharger No. 1:
I = 1370 mg/L
q = 0.15 cfs
x = 0 miles

Discharger No. 3:
I = 665 mg/L
q = 4.62 cfs
x = 50 miles

Discharger No. 5:
I = 1500 mg/L
q = 3.2 cfs
x = 100 miles
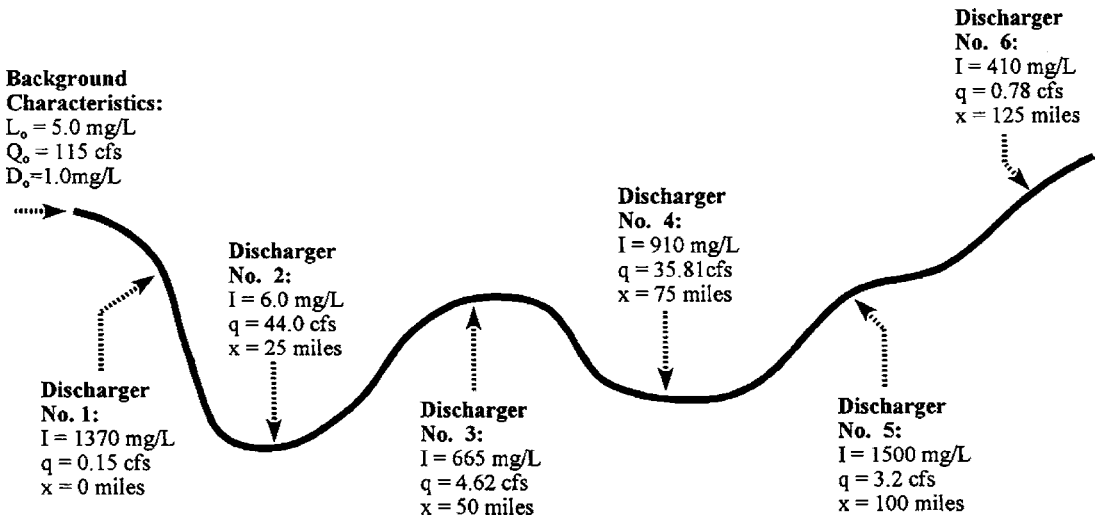
Figure 8.21   Schematic sketch of hypothetical stream in the waste-load allocation (WLA) example. (*After Tung and Hathhorn, 1989.*)

TABLE 8.5   Mean Values of Physical Stream Parameters Used in WLA Example

(a) Mean stream characteristics for each reach

Reach	Deoxygenation coefficient K_d (L/day)	Reaeration coefficient K_a (L/day)	Average stream velocity (mi/day)	Raw waste concentration I (mg/L BOD)	Effluent flow rate (ft^3/s)
1	0.6	1.84	16.4	1370	0.15
2	0.6	2.13	16.4	6.0	44.00
3	0.6	1.98	16.4	665	4.62
4	0.6	1.64	16.4	910	35.81
5	0.6	1.64	16.4	1500	3.20
6	0.6	1.48	16.4	410	0.78

(b) Background characteristics

Upstream waste concentration L_0 (mg/L BOD)	Upstream flow rate Q_0 (ft^3/s)	Upstream DO deficit D_0 (mg/L)
5.0	115.0	1.0

In the example, the length of each reach in the system is 10 mi, and the spatial correlation of representative water-quality parameter values between two reaches is computed based on the separation distance between the centers of the two reaches. To examine the effect of spatial correlation structure on the optimal waste-load allocation, two zones of influence ($h_o = 15$ mi and $h_o = 30$ mi) along with the three variogram models, Eqs. (8.63) through (8.65), are used. A value of $h_o = 15$ mi implies that the water-quality parameters in a given reach are spatially correlated only with the two immediate adjacent reaches. For $h_o = 30$ mi, the spatial correlation extends two reaches upstream and downstream of the reach under consideration. The optimal solutions to the stochastic WLA problem under these various conditions are presented in Tables 8.7 and 8.8.

TABLE 8.6   Standard Deviations Selected for Physical Stream Characteristics

(a) For each reach

Reach (units)	Deoxygenation coefficient K_d (L/day)	Reaeration coefficient K_a (L/day)	Average stream velocity U (ft^3/s)
1–6	0.2	0.4	4.0

(b) Background characteristics

Upstream waste concentration L_0 (mg/L BOD)	Upstream flow rate Q_0 (ft^3/s)	Upstream DO deficit D_0 (mg/L)
10.0	20.0	0.3

TABLE 8.7 Maximum Total BOD Load that Can Be Discharged for Different Reliability Levels and Spatial Correlation Structures under Normal Distribution

		$h_o = 15$ mi			$h_o = 30$ mi		
α	I^*	T	S	G	T	S	G
0.85	671[†]	734	737	679	659	664	694
0.90	633	693	695	639	624	625	656
0.95	588	644	646	593	580	578	610
0.99	521	570	572	524	516	511	541

[*] I = independence; T = transitive model; S = spherical model; G = Gaussian model.
[†] Total BOD load concentration in mg/L.

Examining Tables 8.7 and 8.8, the maximum total BOD discharge, under a given spatial correlation structure, reduces as the reliability of water-quality constraints increases. This behavior is expected because an increase in water-quality compliance reliability is equivalent to imposing stricter standards on water-quality assurance. To meet this increased water-quality compliance reliability, the amount of waste discharge must be reduced to lower the risk of water-quality violation at the various control points. When continuing to increase the required reliability for the water-quality constraints, at some point these restrictions could become too stringent, and feasible solutions to the problem are no longer obtainable.

From Tables 8.7 and 8.8, using a lognormal distribution for the LHS of water-quality constraints yields a higher total BOD discharge than that under a normal distribution when the performance reliability requirement is 0.85. However, the results reverse themselves when reliability requirements are greater than or equal to 0.90. This indicates that the optimal solution to the stochastic WLA model depends on the distribution used for the LHS of the water-quality constraints. From the investigation of Tung and Hathhorn (1989), a lognormal distribution was found to best describe the DO deficit concentration in a single-reach case. In other words, each term of the LHS in the water-quality

TABLE 8.8 Maximum Total BOD Load that Can Be Discharged for Different Reliability Levels and Spatial Correlation Structures under Lognormal Distribution

		$h_o = 15$ mi			$h_o = 30$ mi		
α	I^*	T	S	G	T	S	G
0.85	691[‡]	753	755	699	676	686	712
0.90	633	692	694	640	623	626	655
0.95	560	614	616	565	554	551	582
0.99	424	496	498	425	420	388	471

[*] I = independence; T = transitive model; S = spherical model; G = Gaussian model.
[‡] Total BOD load concentration in mg/L.

constraints could be considered as a lognormal random variable. Therefore, the LHS is the sum of correlated lognormal random variables. For the first two or three reaches from the upstream end of the system, the distribution of the LHS may close to lognormal because the number of terms in the LHS is few. However, considering the control point for farther-downstream reaches, the number of terms in the LHS increases, and the resulting distribution may approach to normal from the argument of the central limit theorem. Since the true distribution for the LHS of water-quality constraints is not known, it is suggested that different distributions be used for model solutions and that the least amount of total BOD load be applied for implementation.

Furthermore, the impacts of the extent of the spatial correlation of the water-quality parameters (represented by the length of h_o) and the structure (represented by the form of the variogram) on the results of the stochastic WLA model also can be observed. When $h_o = 15$ mi, where the spatial correlation of the water-quality parameters extends only one reach, the maximum allowable total BOD load, for all three variogram models, is higher than that of spatially independent case. When the spatial correlation extends over two reaches (that is, $h_o = 30$ mi), use of transitive and spherical variogram models results in lower maximum total BOD loads than that of the spatially independent case, whereas use of a Gaussian variogram yields a higher total BOD load. The model results using a transitive variogram are very similar to those of a spherical model.

As a final comment on the computational aspects of the proposed technique for solving the stochastic nonlinear WLA model formulated in this study, it was observed that the iterative technique proposed takes three to five iterations to converge for all the cases investigated. Therefore, the proposed solution procedure is quite efficient in solving the stochastic WLA model.

8.6.2   Multiobjective stochastic waste-load allocation

The WLA problem, by nature, is a multiobjective problem involving several conflicting objectives. The treatment-equity constraint (Eq. 8.59c) is incorporated for the purpose of fairness. Without it, any attempt to maximize waste discharge (or to minimize treatment cost) could result in allocating large quantities of waste to the upstream users, whereas the downstream discharges could be required to treat their effluent at levels of maximum possible efficiency. This is especially true for slow-moving streams. Several articles have discussed the importance of equity considerations in WLA problems (Gross, 1965; Loucks et al., 1967; Miller and Gill, 1976; Brill et al., 1976).

In general, as the requirement for an equity measure (or fairness) is raised, the total waste discharge to the stream system would be reduced. This will be in direct conflict with the maximization of waste discharge associated with the minimization of treatment cost. Furthermore, from the preserving stream water-quality viewpoint, setting a higher water-quality standard is more desirable. However, such an objective cannot be achieved without increasing waste treatment cost. Therefore, the objectives of preserving water quality and of

enhancing economic efficiency are in conflict each other. Lastly, as the require-
ment of reliability in complying with the water-quality standard is raised, the
total waste load that can be discharged expectedly would have to be reduced.
Therefore, the task of solving WLA problems is multiobjective.

From the preceding discussions, four objective functions can be considered
in WLA modeling: (1) maximization of the total waste load, (2) minimization
of differences in treatment levels among waste dischargers, (3) maximization
of allowable in-stream DO concentration, and (4) maximization of the water-
quality standard compliance reliability. The first objective Z_1 can be formulated
as Eq. (8.59a), which is repeated here as

$$\text{Maximize} \quad Z_1 = \sum_{j=1}^{N} (B_j + D_j)$$

For a stream system involving multiple waste dischargers, the difference in
required treatment levels generally would vary. To collapse different values of
equity measure into a single representative indicator, the worst case associated
with the largest differences can be used. With that, the second objective can be
expressed as

$$\text{Minimize} \quad Z_2 = \delta_{\max} = \max_{j \neq j'} \left\{ \left| \frac{B_j}{I_j} - \frac{B_{j'}}{I_{j'}} \right| \right\} \tag{8.69}$$

where $\delta_{\max}$ is a new decision variable for the equity measure representing the
largest difference in treatment levels among waste dischargers.

The third objective considered is the maximization of the lowest allowable
DO concentration level that should be maintained in the stream environment.
This objective can be expressed as

$$\text{Maximize} \quad Z_3 = \text{DO}_{\min}^{\text{std}} \tag{8.70}$$

in which the new decision variable $\text{DO}_{\min}^{\text{std}}$ is the minimum required DO standard
in the stream.

Similar to the differences in treatment levels, the water-quality compliance
reliability at different control points will not be uniform. To use a single repre-
sentative measure of compliance reliability for the entire system, a conservative
view of looking at the lowest reliability was applied. The objective is to maximize
this lowest compliance reliability $\alpha_{\min}$ as

$$\text{Maximize} \quad Z_4 = \alpha_{\min} = \min[\alpha_1, \alpha_2, \ldots, \alpha_M] \tag{8.71}$$

By the definitions of $\text{DO}_{\min}^{\text{std}}$ and $\alpha_{\min}$, the chance constraints for water-quality
compliance (Eq. 8.59b) can be modified as

$$P \left\{ a_{0i} + \sum_{j=1}^{n_i} \Theta_{ij} B_j + \sum_{j=1}^{n_i} \Omega_{ij} D_j + \text{DO}_{\min}^{\text{std}} \leq \text{DO}_i^{\text{sat}} \right\} \geq \alpha_{\min} \qquad \text{for } i = 1, 2, \ldots, M$$

$$\tag{8.72}$$

The corresponding deterministic equivalent of Eq. (8.72) can be expressed as

$$\sum_{j=1}^{n_i} E(\Theta_{ij})B_j + \sum_{j=1}^{n_i} E(\Omega_{ij})D_j + \text{DO}_{\min}^{\text{std}}$$

$$+ F_z^{-1}(\alpha_{\min})\sqrt{(\boldsymbol{B}, \boldsymbol{D})^t \boldsymbol{C}(\Theta_i, \Omega_i)(\boldsymbol{B}, \boldsymbol{D})} \le R_i^{''} \qquad (8.73)$$

in which $R_i^{''} = \text{DO}_i^{\text{sat}} - E(a_{oi})$.

Note that the original objective given in Eq. (8.71) is to maximize $\alpha_{\min}$. However, under the assumption that the standardized left-hand sides of the water-quality constraints are continuous and unimodal random variables, the decision variable $\alpha_{\min}$ would have a strictly increasing relationship with $F_z^{-1}(\alpha_{\min})$. Therefore, maximization of $\alpha_{\min}$ is then equivalent to maximizing $F_z^{-1}(\alpha_{\min})$. By letting $z_{\min} = F_z^{-1}(\alpha_{\min})$, Eq. (8.71) can be written as

$$\text{Maximize} \quad Z_4 = z_{\min} \qquad (8.74)$$

Note that the decision variable $z_{\min}$ is unrestricted in sign. The objective of maximizing the lowest compliance reliability is equivalent to minimizing the highest water-quality violation risk.

The preceding multiobjective WLA problem can be solved by various techniques described in the references cited in Sec. 8.1.2. In the following, the constraint method is used by which the preceding multiobjective WLA problem is expressed as

$$\text{Maximize} \quad z_{\min} \qquad (8.75a)$$

subject to

$$\sum_{j=1}^{n_i} E(\Theta_{ij})B_j + \sum_{j=1}^{n_i} E(\Omega_{ij})D_j + \text{DO}_{\min}^{\text{std}}$$

$$+ z_{\min}\sqrt{(\boldsymbol{B}, \boldsymbol{D})^t \boldsymbol{C}(\Theta_i, \Omega_i)(\boldsymbol{B}, \boldsymbol{D})} \le R_i^{''} \qquad \text{for } i = 1, 2, \ldots, M \quad (8.75b)$$

$$0.35 \le \frac{B_j}{I_j} \le 0.90 \qquad \text{for } j = 1, 2, \ldots, N \qquad (8.75c)$$

$$\left| \frac{B_j}{I_j} - \frac{B_{j'}}{I_{j'}} \right| - \delta_{\max} \le 0 \qquad \text{for all } j \ne j' \qquad (8.75d)$$

$$\sum_{j=1}^{N}(B_j + D_j) \ge Z_1^o \qquad (8.75e)$$

$$\delta_{\max} \le Z_2^o \qquad (8.75f)$$

$$\text{DO}_{\min}^{\text{std}} \ge Z_3^o \qquad (8.75g)$$

and nonnegativity constraints for the decision variables, except for $z_{\min}$. In Eqs. (8.75e–g), the right-hand sides Z_1^o, Z_2^o, and Z_3^o are the values of the objective functions 1, 2, and 3, respectively, which are to be varied parametrically.
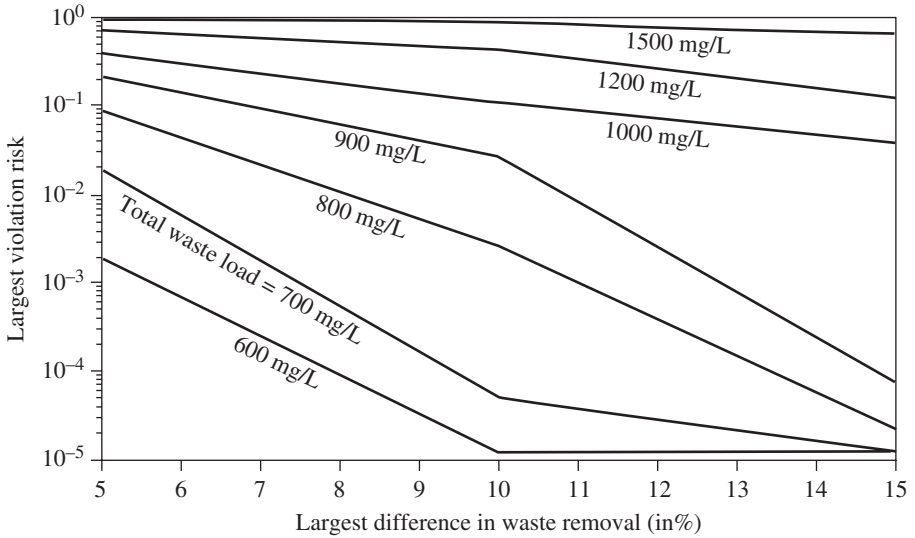
Figure 8.22 Tradeoff curves of various objectives in stochastic WLA problem with 4 mg/L minimum DO standard. (*After Tung and Hathhorn, 1989.*)

Using the same hypothetical stream system as shown in Fig. 8.21 and the corresponding data, the solution to this multiobjective WLA model by the constraint method yields a series of tradeoff curves among the various objectives. Figures 8.22 through 8.24 show the tradeoffs among three objectives for a given
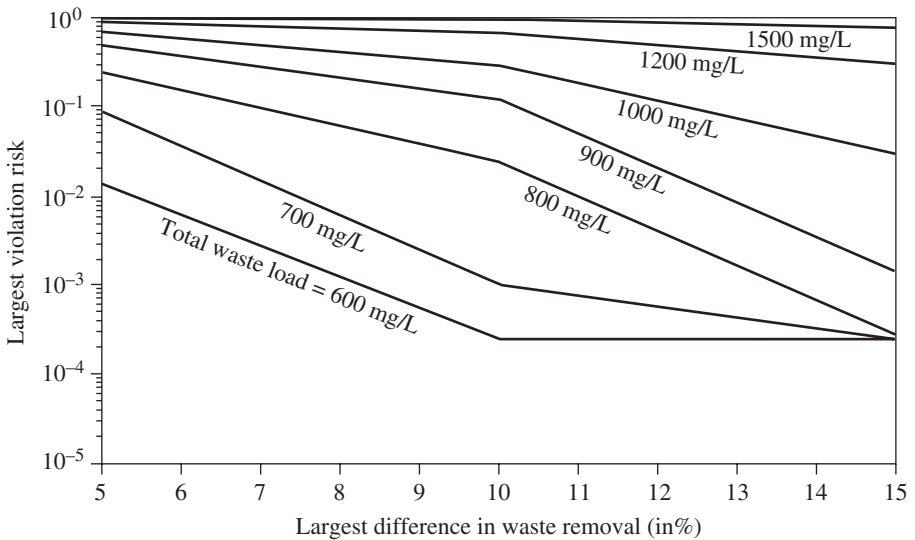
Figure 8.23 Tradeoff curves of various objectives in stochastic WLA problem with 5 mg/L minimum DO standard. (*After Tung and Hathhorn, 1989.*)
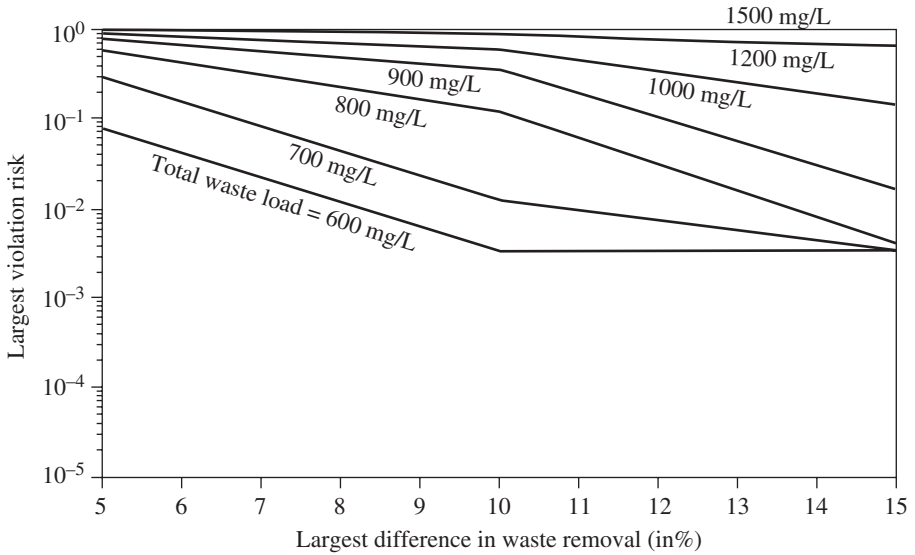
Figure 8.24 Tradeoff curves of various objectives in stochastic WLA problem with 6 mg/L minimum DO standard. (*After Tung and Hathhorn, 1989*.)

minimum DO standard concentration. As can be seen for a specified minimum DO standard and total waste load, the largest water-quality violation risk decreases as the maximum difference in treatment equity increases. An increase in the treatment equity measure by δ_{max} implies a larger tolerance for the
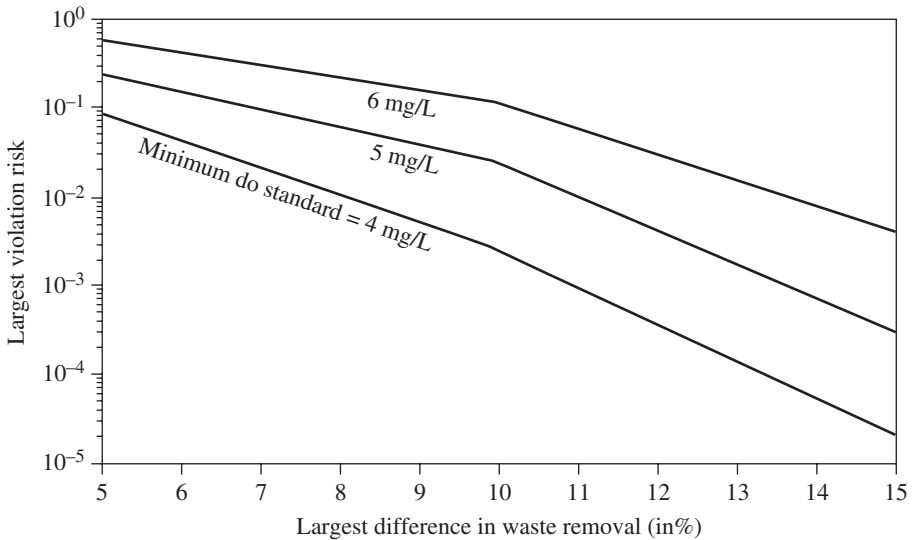
Figure 8.25 Tradeoff curves of the various objectives in stochastic WLA problem with total waste load fixed at 800 mg/L. (*After Tung and Hathhorn, 1989*.)
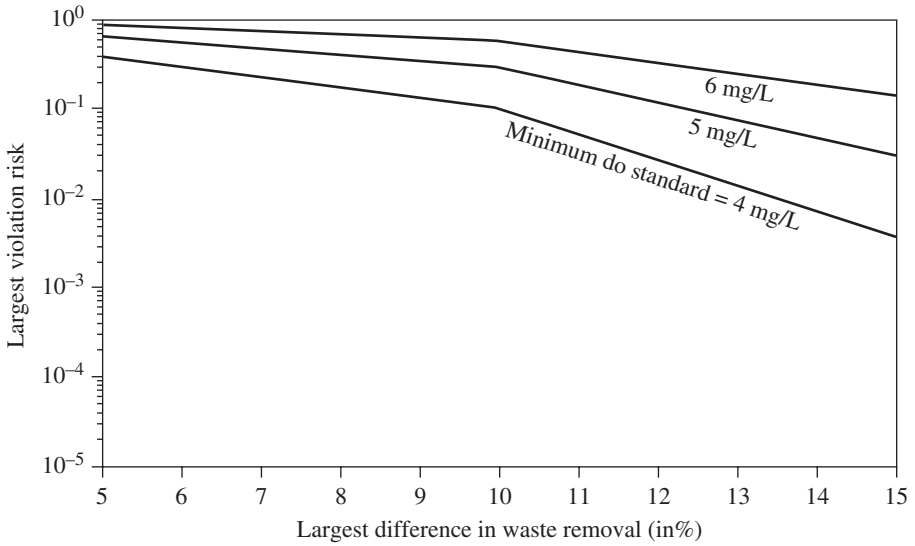
Figure 8.26 Tradeoff curves of the various objectives in stochastic WLA problem with total waste load fixed at 1000 mg/L. (*After Tung and Hathhorn, 1989.*)

unfairness in the treatment requirement among waste dischargers. As the level of the minimum required DO standard is raised, the set of tradeoff curves moves upward. To show the tradeoffs for different minimum DO standard, Figs. 8.25 and 8.26 are plotted for the risk of water-quality standard violation, treatment equity, and water-quality standard while the total waste load to the stream system is fixed at some specified levels.

Appendix 8A: Derivation of Water-Quality Constraints

In a WLA problem, one of the most essential requirements is the assurance of a minimum concentration of dissolved oxygen (DO) throughout the river system in an attempt to maintain desired levels of aquatic biota. The constraint relating the response of DO to the additional of in-stream waste generally is defined by the Streeter-Phelps equation (Eq. 8.60) or its variations (ReVelle et al., 1968; Bathala et al., 1979). To incorporate water-quality constraints into the model formulation, a number of control points are placed within each reach of the river system under investigation. By using the Streeter-Phelps equation, each control point and discharge location becomes a constraint in the LP model, providing a check on water-quality at that location. In a general framework, a typical water quality constraint would be as follows:

$$\sum_{j=1}^{n_i} \theta_{ij} L_j + \sum_{j=1}^{n_i} \psi_{ij} D_j \le R_i \tag{8A.1}$$

where

$$
\theta_{ij} = \frac{q_i}{Q_0 + \sum_{m=1}^{n_i} q_m} \left[\left(\prod_{\ell=j}^{n_i-1} b_{\ell,\ell+1} \right) d_{n_i,i} \right.
$$

$$
\left. + b_{n_i,i}^a \sum_{p=1}^{n_i-1} \left(\prod_{\ell=j}^{n_i-p-1} b_{\ell,\ell+1} \right) d_{n_i-p,n_i-p-1} \left(\prod_{k=n_i-p+1}^{n_i-1} b_{k,k+1}^a \right) \right] \quad \text{(8A.2)}
$$

$$
\psi_{ij} = \frac{q_i}{Q_0 + \sum_{m=1}^{n_i} q_m} \left(\prod_{\ell=j}^{n_i-1} b_{\ell,\ell+1} \right) b_{n_i,i}^a \quad \text{(8A.3)}
$$

$$
R_i = \text{DO}_i^{\text{sat}} - \text{DO}_i^{\text{std}} - \frac{q_i}{Q_0 + \sum_{m=1}^{n_i} q_m} \left[\left(L_0 Q_0 \prod_{\ell=j}^{n_i-1} b_{\ell,\ell+1} \right) \right.
$$

$$
\left. + L_0 Q_0 \sum_{p=2}^{n_i-1} \left(\prod_{\ell=1}^{n_i-p} b_{\ell,\ell+1} \right) d_{n_i-p+1,n_i-p+2} \left(\prod_{k=n_i-p+1}^{n_i-1} b_{k,k+1}^a \right) + \left(D_0 Q_0 \prod_{k=j}^{n_i-1} b_{k,k+1}^a \right) \right]
$$

$$
\text{(8A.4)}
$$

$$
d_{n_i,i} = \frac{K_{n_i}^d}{K_{n_i}^a - K_{n_i}^d} \left(b_{n_i,i} - d_{n_i,i} \right) \quad \text{(8A.5)}
$$

and

$$
b_{n_i,n_i+1} = \exp \frac{-K_{n_i}^d x_{n_i,n_i+1}}{U_{n_i}} \quad \text{(8A.6)}
$$

$$
b_{n_i,n_i+1}^a = \exp \frac{-K_{n_i}^a x_{n_i,n_i+1}}{U_{n_i}} \quad \text{(8A.7)}
$$

in which M is the total number of control points, n_i is the number of discharg-ers upstream of the control point i, $K_{n_i}^a$ and $K_{n_i}^d$ are, respectively, the reaeration and deoxygenation coefficients (days^{-1}) in the reach, L_0, Q_0, and D_0 are the upstream waste concentration (mg/L BOD), flow rate (ft^3/s), and DO deficit (mg/L), respectively, D_{n_i}, L_{n_i}, and q_{n_i} are the DO deficit (mg/L), waste concen-tration (mg/L BOD), and effluent flow rate (ft^3/s) from each discharge location, respectively, $x_{n_i,i}$ is the distance (miles) between discharge location and control point i, and U_{n_i} is the average stream velocity (mi/day) in reach n_i. R_i repre-sents the allowable DO deficit at the control point i, available for utilization of water discharge (mg/L). It should be noted that in addition to each control point i, water quality is also checked at each discharge location n_i.

Problems

8.1   A city in an alluvial valley is subject to flooding. As a matter of good fortune, no serious floods have taken place during the past 50 years, and therefore, no flood-control measure of any significance has been taken. However, last year a serious flood threat developed; people realized the danger they are exposed to, and a flood investigation is under way.

From the hydrologic flood frequency analysis of past streamflow records and hydrometric surveys, the discharge-frequency curve, rating curve, and damage curve under nature condition are derived and shown in the table below and Figs. 8P.1 and 8P.2, respectively. Also, it is known that the flow-carrying capacity of existing channel is 340 m³/s.

T (years)	2	5	10	20	50	100	200	500	1000
Q (m³/s)	255	340	396	453	510	566	623	680	736

Three flood-control alternatives are considered, and they are (1) construction of a dike system throughout the city that will contain a flood peak of 425 m³/s but will fail completely if the river discharge is higher, (2) design of an upstream permanent diversion that would divert up to 85 m³/s if the upstream inflow discharge exceeds existing channel capacity of 340 m³/s, and (3) construction of a detention basin upstream to provide a protection up to a flow of 425 m³/s.

The detention basin will install a conduit with a maximum flow capacity of 340 m³/s. Assume that all flow rates less than 340 m³/s will pass through the conduit without being retarded behind the detention basin. For incoming flow rate between 340 and 425 m³/s, runoff volume will be stored temporally in the detention basin so that outflow discharge does not exceed existing downstream channel capacity. In other words, inflow hydrograph with peak discharge exceeding 425 m³/s could result in spillway overflow, and hence the total outflow discharge would be higher than the channel capacity. The storage-elevation curve at the detention basin site and normalized inflow hydrograph of different return
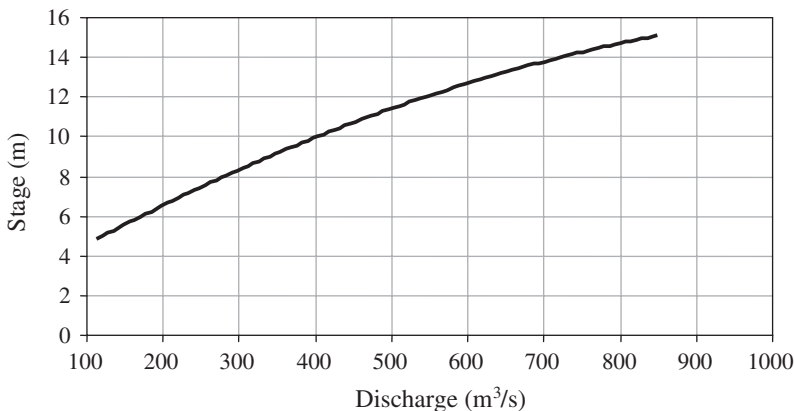
Figure 8P.1   Stage-discharge (rating) curve.
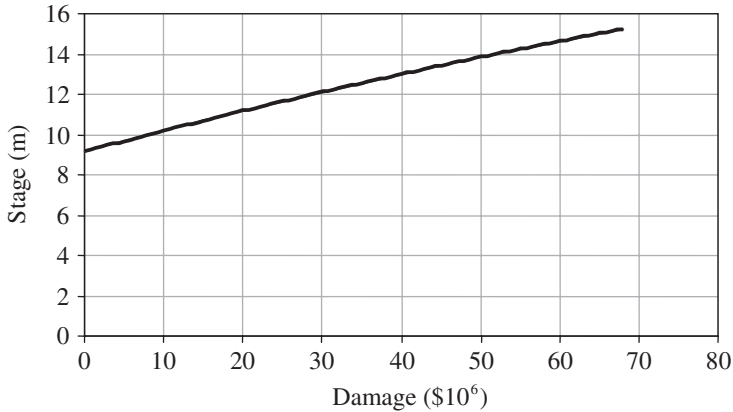
Figure 8P.2 Stage-damage relation.

periods are shown in Figs. 8P.3 and 8P.4, respectively. The flow capacities of the conduit and spillway can be calculated, respectively, by

Conduit: $$Q_c = 159h^{0.5}$$

Spillway: $$Q_s = 67.0(h - h_s)^{1.5}$$

where Q_c and Q_s are conduit and spillway capacity (in m^3/s), respectively, h is water surface elevation in detention basin (in m) above the river bed, and h_s is elevation of spillway crest (in m) above the river bed.

To simplify the algebraic manipulations in the analysis, the basic relations between stage, discharge, storage, and damage are derived to fit the data:
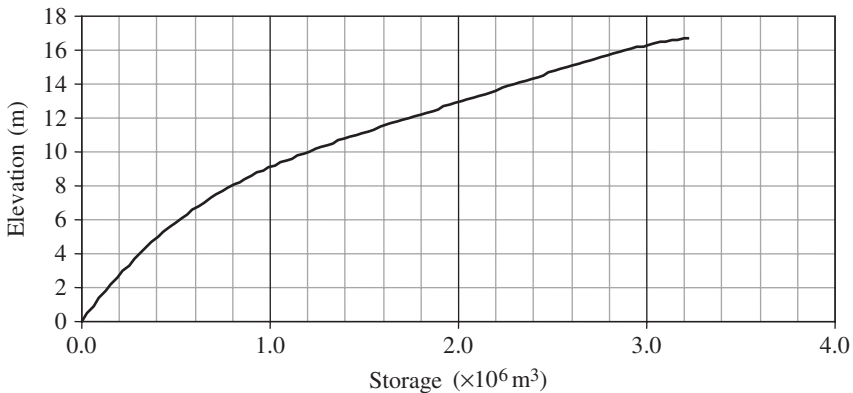
Figure 8P.3 Detention basin storage-elevation relation.
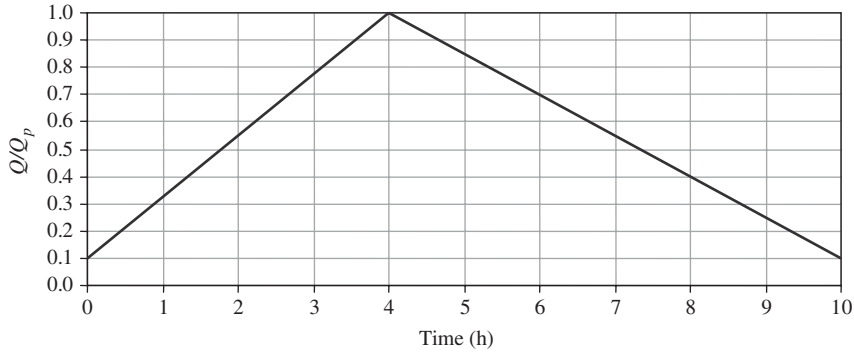
Figure 8P.4   Normalized inflow hydrograph. (*Note*: Qp = peak inflow discharge.)

(i) Stage discharge:	$Q = 8.77 + 7.761H + 3.1267H^2$	
(ii) Stage damage:	$D = \text{Max}(0, -54.443 + 2.8446H + 0.34035H^2)$	
(iii) Storage elevation:	$S = 0.0992 + 0.0021h + 0.011h^2$	
	$h > 0; S = 0$, otherwise	

in which Q is flow rate in channel (m^3/s), H is channel water stage (m), D is flood damage ($\$10^6$), S is detention basin storage (10^6 m^3), and h is water level in detention basin above channel bed (m).

With all the information provided, answer the following questions:

(a) Develop the damage-frequency curve for the natural condition.

(b) What is the height of spillway crest of the detention basin above the river bed?

(c) Develop the damage-frequency curves as the results of each of the three flood control measures.

(d) Rank the alternatives based on their merits on the flood damage reduction.

8.2 Refer to Problem 8.1 and consider the alternative of building a levee system for flood control. It is known that the capital-cost function for constructing the levee system is

$$\text{FC}(Y) = 1.0 + 0.6(Y - 7) + 0.05(Y - 7)^3$$

in which Y is the height of levee, and $\text{FC}(Y)$ is the capital cost (in million dollars). Suppose that the service period of the levee system is to be 50 years and the interest rate is 5 percent. Determine the optimal design return period such that the annual total expected cost is the minimum.

8.3 Consider a confined aquifer with homogeneous soil medium. Use the Thiem equation and the linear superposition principle (see Problem 2.30) to formulate a steady-state optimal groundwater management model for the aquifer system sketched in Fig. 8P.5. The management objective is to determine the maximum total allowable pumpage from the three production wells such that the drawdown of piezometric head at each of the five selected control point would not exceed a specified limit.
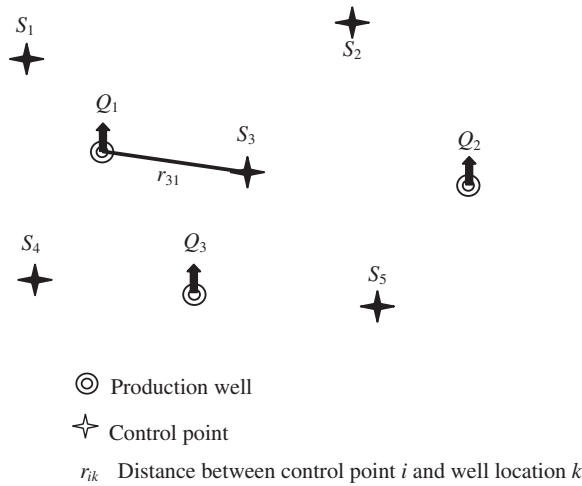
Distance (in ft) between Pumping Wells and Control Points						
Pumping well	Control points					Pumping capacity (gpd)
	1	2	3	4	5	
1	160	380	160	260	430	200,000
2	520	260	300	480	160	200,000
3	450	450	200	200	200	200,000
Maximum allowable drawdown	7 ft	7 ft	15 ft	7 ft	7 ft	

Figure 8P.5 Location of pumping wells and control points for a hypothetical groundwater system (Problems 8.3–8.8). (*After Mays and Tung, 1992.*)

(a) Formulate a linear programming model for the groundwater system as shown in Fig. 8P.5.

(b) Suppose that the radius of influence of all pump wells is 700 ft (213 m) and that the aquifer transmissivity is 5000 gal/day/ft (0.00072 m²/s). Based on the information given in Fig. 8P.5, solve the optimization model formulated in part (a).

8.4 Consider that the soil medium is random and that the transmissivity has a log-normal distribution with mean value of 5000 gal/day/ft and a coefficient of variation of 0.4. Construct a chance-constrained model based on Problem 8.3, and solve the chance-constrained model for a 95 percent compliance reliability of all constraints.

8.5 Modify the formulation in Problem 8.3, and solve the optimization model that maximizes the total allowable pumpage in such a way that the largest drawdown among the five control points does not exceed 10 ft.

8.6 Develop a chance-constrained model based on Problem 8.5, and solve the model for a 95 percent compliance reliability of all constraints.

8.7 Based on the chance-constrained model established in Problem 8.6, explore the tradeoff relationship among the maximum total pumpage, compliance reliability, and the largest drawdown.

8.8 Modify the formulation in Problem 8.6 to develop a chance-constrained management model for the hypothetical groundwater system that maximizes the total allowable pumpage while satisfying the desired lowest compliance reliability for all constraints. Furthermore, solve the model for the hypothetical system shown in Fig. 8P.5 with the lowest compliance reliability of 95 percent.

8.9 In the design of a water supply system, it is general to consider a least-cost system configuration that satisfies the required water demand and pressure head at the demand points. The cost of the system may include the initial investment for the components (e.g., pipes, tanks, valves, and pumps) and the operational costs. The optimal design problem, in general, can be cast into

Minimize Capital cost + energy cost
subject to (1) Hydraulic constraints
(2) Water demands
(3) Pressure requirements

Consider a hypothetical branched water distribution system as shown in Fig. 8P.6. Develop a linear programming model to determine the optimal combination of cast iron pipe length of various commercially available pipe sizes for each branch. The objective is to minimize the total pipe cost of the system, subject to water demand and pressure constraints at all demand points. The new cast iron pipes of all sizes have the Hazen-Williams roughness coefficient of 130. The cost of pumping head is $500/ft, and the pipe costs for available pipe sizes are listed below
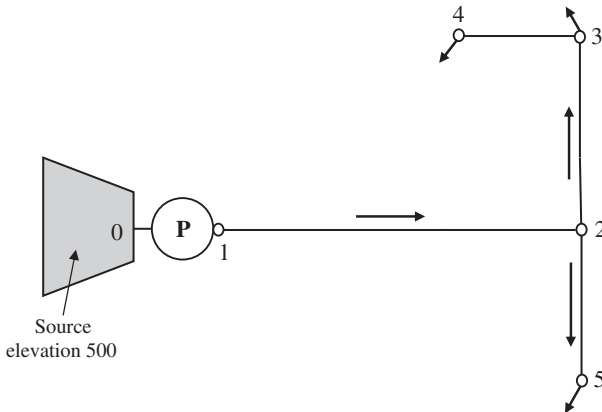
Figure 8P.6 A hypothetical water distribution system.

Pipe diameter (inches)	8	10	12	15	18	21	24	27	30	36	42
Pipe cost ($/ft)	4	5	6	8	9	11	12	14	15	18	21

The length and the plausible pipe sizes for each pipe section are:

	Pipe branches			
	1–2	2–3	3–4	2–5
Pipe length (feet)	6000	4000	3000	3000
Pipe sizes considered (in)	36, 30, 24	24, 21, 18	18, 15, 12	21, 18, 15

To this hypothetical system, the required flow rate and water pressure at each demand node are

Demand node	3	4	5
Required flow rate (ft^3/s)	6	6	10
Minimum pressure (ft)	550	550	550

8.10 Consider that the Hazen-Williams roughness coefficient is a uniform random variable with a mean of 130 and a coefficient of variation of 30. Develop a chance-constrained optimization model for the hypothetical water distribution system shown in Fig. 8P.6. Furthermore, solve the model based on the data given in Problem 8.9.

References

Aarts, E., and Korst, J. (1989). *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, New York.

AASHTO (1979). *Highway Drainage Guidelines*, Vols. I and IV. American Association of State and Highway Transportation Officials, Washington, D.C.

Afshar, A., Barkhordary, A., and Marino, M. A. (1994). Optimzing river diversion under hydarulic and hydrologic uncertainties, *Water Resources Planning and Management*, ASCE, 120(1): 36–47.

Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm, in *Genetic Algorithm and Their Applications*. Proceedings of 2nd International Conference on Genetic Algorithms, pp. 14–21, Lawrence Erlbaum Associates, Hillsdale, NJ.

Bao, Y., Tung, Y. K., and Hasfurther, V. R. (1987). Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures, *Water Resources Research*, AGU, 23(11):2023–2029.

Bathala, C. T., Das, K. C., and Jones, W. D. (1979). Assimilative capacity of small streams, *Journal of Environmental Engineering*, ASCE, 6(12):1049–1060.

Bodo, B., and Unny, T. E. (1976). Model uncertainty in flood frequency analysis and frequency based design, *Water Resources Research* 12(6):1109–1117.

Bradley, J. N. (1973). Hydraulics of bridge waterways, Federal Highway Administration, Hydraulic Design Series No. 1.

Brill, E., Liebman, J., and ReVelle, C. (1976). Equity measures for exploring water quality management alternatives, *Water Resources Research*, AGU, 12(3):845–851.

Brooke, A. D., Kendrick, D., and Meerhaus, A. (1988). *GAMS: A User's Guide*, Scientific Press, Palo Alto, CA.

Buras, N. (1972). *Scientific Allocation of Water Resources*, American Elservier Publishing, New York.

Bureau of Reclamation (1986). Guideline for decision analysis, ACER Technical Memo, No. 7, U.S. Dept. of Interior.

Burn, D. H., and McBean, E. A. (1985). Optimization modeling of water quality in an uncertain environment, *Water Resources Research* 21(7):934–940.

Chadderton, R. A., Miller, A. D., and McDonnell, A. J. (1981). Analysis of waste load allocation procedure, *Water Resources Bulletin* 17(5):760–766.

Chadderton, R. A., Miller, A. C., and McDonnell, A. J. (1982). Uncertainty analysis of dissolved oxygen model, *Journal of Environmental Engineering*, ASCE, 108(5):1003–1012.

Chankong, V., and Haimes, Y. Y. (1983). *Multiobjective Decision Making: Theory and Methodology*, North-Holland Publishing, New York.

Cheng, S. T., Yen, B. C., and Tang, W. H. (1986). Wind-induced overtopping risk of dams, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, pp. 48–58, Water Resources Publications, Littleton, CO.

Chow, V.-T. (1962). Hydrologic determination of waterway areas for the design of drainage structures in small drainage basins, University of Illinois Engineering Experiment Station Bulletin No. 462, Urbana, Ill., 104 p.

Chow, V.-T., Maidment, D. R., and Tauxe, G. W. (1975). Computer time and memory requirements for DP and DDDP in water resources system analysis, *Water Resources Research*, AGU, 11(5):621–628.

Churchill, M. A., Elmore, H. L., and Buckingham, R. A. (1964). The prediction of stream reaeration rates, *Journal of Environmental Engineering*, ASCE, 88(1):1–46.

Cohon, J. L. (1978). *Multiobjective Programming and Planning*, Academic Press, New York.

Cooper, L. L., and Cooper, M. W. (1981). *Introduction to Dynamic Programming*, Pergamon Press, Elmsford, NY.

Corry, M. L., Jones, J. S., and Thompson, D. L. (1980). The design of encroachments of floodplain using risk analysis, Hydraulic Engineering Circular, No. 17, U.S. Dept. of Transportation, Federal Highway Administration, Washington.

Dantzig, G. B. (1963). *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ.

Datta, B., and Dhiman, S. D. (1996). Chance-constrained optimal monitoring network design for pollutants in ground water, *Journal of Water Resources Planning and Management*, ASCE, 122(3):180–188.

Datta, B., and Houck, M. H. (1984). A stochastic optimization model for real-time operation of reservoirs incorporating forecast errors, *Water Resources Research*, AGU, 20(8):1039–1046.

Denardo, E. V. (1982). *Dynamic Programming Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ.

Dobbins, W. E. (1964). BOD and oxygen relationships in stream, *Journal of Environmental Engineering*, ASCE, 90(1):53–78.

Dreyfus, S., and Law, A. (1977). *The Art and Theory of Dynamic Programming*, Academic Press, New York.

Edgar, T. F., and Himmelblau, D. M. (1988). *Optimization of Chemical Processes*, McGraw-Hill, New York.

Eiker, E. E., and Davis, D. W. (1996) Risk-based analysis for Corps flood project studies: A status report, in *Proceedings, Rivertech 96: 1st International Conference on New/Emerging Concepts for Rivers*, ed. by W. H. C. Maxwell, H. C. Preul, and G. E. Stout, pp. 332–339. International Water Resources Association, Albuquerque, NM.

Ellis, J. H. (1987). Stochastic water quality optimization using imbedded chance constraints, *Water Resources Research*, AGU, 23(12):2227–2238.

Fletcher, R. (1980). *Practical Methods of Optimization*, Vol. 1: *Unconstrained Optimization*, John Wiley and Sons, New York.

Fujiwara, O., Gnanendran, S. K., and Ohgaki, S. (1986). River quality management under stochastic streamflow, *Journal of Environmental Engineering*, ASCE, 112(1):185–198.

Fujiwara, O., Gnanendran, S. K., and Ohgaki, S. (1987). Chance-constrained model for river water quality management, *Journal of Environmental Engineering*, ASCE, 113(5):1018–1031.

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical Optimization*, Academic Press, New York.

Goicoechea, A., Hansen, D. R., and Duckstein, L. (1982). *Multiobjective Decision Analysis with Engineering and Business Applications*, John Wiley and Sons, New York.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, New York.

Gorelick, S. M. (1982). A model for managing sources of groundwater pollution, *Water Resources Research*, AGU, 18(4):773–781.

Gross, W. M. (1965). A lawyer looks at stream pollution, *Civil Engineering*, ASCE, 44–45.

Heidari, M., Chow, V. T., Kokotovic, P. V., and Meridith, D. D. (1971). Discrete differential dynamic programming approach to water resources systems optimization, *WaterResources Research*, AGU, 7(2):273–282.

Hillier, F. S., and Lieberman, G. J. (1990). *Introduction to Operation Research*, 5th ed., McGraw-Hill, New York.

Houck, M. H. (1979). A chance-constrained optimization model for reservoir design and operation, *Water Resources Research*, AGU, 15(12):1011–1016.

Interagency Advisory Committee on Water Data (1982). Guidelines for determining flood flow frequency, Bulletin 17B, U.S. Department of the Interior, U.S. Geological Survey, Office of Water Data Coordination, Reston, VA.

Jonkman, S. N., van Gelder, P. H. A. J. M., and Vrijing, J. K. (2003). An overview of quantitative risk measures for loss of life and economic damage, *Journal of Hazardous Materials*, A99:1–30.

Journel, A. G., and Huijbregts, C. J. (1978). *Mining Geostatistics*, Academic Press, New York.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica* 4(4):373–395.

Khatchian, L. (1979). A polynomial algorithm in linear programming, *Soviet Mathematics*, Doklady 20, pp. 191–194.

Kirkpatrick, S., Gerlatt, C. D. Jr., and Vecchi, M. P. (1983). Optimization by simulated annealing, *Science* 220:671–680.

Kolbin, V. V. (1977). *Stochastic Programming*, Reidal Publishing Co., Boston.

Krenkel, P. A., and Novotny, V. (1980). *Water Quality Management*, Academic Press, New York.

Lansey, K. E., Duan, N., Mays, L. W., and Tung, Y. K. (1989). Water distribution system design under uncertainties, *Journal of Water Resources Planning and Management*, ASCE, 115(5):630–645.

Lasdon, L. S., Waren, A. D., Jain, A., and Ratner, M. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming, ACM *Transactions on Mathematical Software* 4:34–50.

Lasdon, L. S., and Waren, A. D. (1978). Generalized reduced gradient software for linearly and nonlinearly constrained problems, in *Design and Implementation of Optimization Software*, ed. by H. Journal Greenberg, pp.363–397, Sijthoff and Noordhoff, Amsterdam.

Lee, H. L., and Mays, L. W. (1986). Hydraulic uncertainties in flood levee capacity, *Journal of Hydraulic Engineering*, ASCE, 112(10):928–934.

Liebman, J. S., Lasdon, L. S., Schrage, L., and Waren, A. (1986). *Modeling and Optimization with GINO*, Scientific Press, Palo Alto, CA.

Linsley, R. K. (1986). Flood estimates: How good are they? *Water Resources Research* 22(9):159s–164s.

Linsley, R. K., and Franzini, J. B. (1979) *Water Resources Engineering*, McGraw-Hill, New York.

Lohani, B. N., and Thanh, W. R. (1978). Stochastic programming model for water quality management in a river, *Journal of Water Pollution Control Federation* 50:2175–2182.

Loucks, D. P., Revelle, C. S., and Lynn, W. R. (1967). Linear programming models for water pollution control, *Management Science* 14:166–181.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Mao, N., and Mays, L. W. (1994). Goal programming model for determining freshwater inflows to bays and estuaries, *Journal of Water Resources Planning and Management*, ASCE, 120(3):316–329.

Mautaugh, B. A., and Saunders, M. A. (1980). MINOS/AUGMENTED User's Manual, System Optimization Lab. Tech. Report 80-14, Department of Operations Research, Stanford University, Stanford, CA.

Mautaugh, B. A., and Saunders, M. A. (1983). MINOS 5.0 User's Manual, *System Optimization Lab. Tech. Report* 83-20, Department of Operations Research, Stanford University, Stanford, CA.

Mays, L. W. (1979). Optimal design of culverts under uncertainties, *Journal of Hydraulic Engineering*, ASCE, 105(5):443–460.

Mays, L. W., and Tung, Y. K. (1992). *Hydrosystem Engineering and Management*, McGraw-Hill, New York.

Mays, L. W., and Tung, Y. K. (2005). Systems analysis, in *Water Resources Handbook*, 2nd ed., ed. by L. W. Mays, McGraw-Hill Company, New York.

Mays, L. W., and Yen, B. C. (1975). Optimal cost design of branched sewer systems, *Water Resources Research*, 11(1):37–47.

McCann, M. M., Franzini, J. B., Kavazanjiam, E., and Shah, H. C. (1984). Preliminary safety evaluation of existing dams, Vol. 1, Report prepared for Federal Emergency Mgmt. Agency by Dept. of Civ. Eng., Stanford University, Stanford, CA.

McCormick, G. P. (1983). *Nonlinear Programming: Theory, Algorithm, and Applications*, John Wiley and Sons, New York.

Melching, C. S., Wenzel, H. G., and Yen, B. C. (1987). Application of system reliability analysis to flood forecasting, in *Application of Frequency and Risk in Water Resources*, ed. by V. P. Singh, pp. 335–350, D. Reidel Publishing Co., Dordrecht, The Netherlands.

Miller, W. L., and Gill, J. H. (1976). Equity considerations on controlling nonpoint pollution from agricultural sources, *Water Resources Bulletin*, AWRA, 12(2):253–261.

Morgan, D. R., Eheart, J. W., and Valocchi, A. J. (1993). Aquifer redediation design under uncertainty using a new chance constrained programming technique, *Water Resources Research*, AGU, 29(3):551–561.

National Research Council (1983). *Safety of Existing Dams: Evaluation and Improvement*. National Academy Press, Washington.

Parks, G. T. (1990). An intelligent stochastic optimization routine for nuclear fuel cycle design, *Nuclear Technology* 89:233–246.

Pate-Cornell, M. E., and Tagaras, G. (1986). Risk cost for new dams: Economic analysis and effects of monitoring, *Water Resources Research*, AGU, 22(1):5–14.

Pincus, M. (1970). Monte Carlo method for the approximate solution of certain types of constrained optimization problems, *Operations Research* 18:1225–1228.

Plate, E. J. (1992). Stochastic design in hydraulics: Concepts for a broader application, in *Stochastic Hydraulics '92*, Proc. 6th IAHR Int'l Symp., Taipei, ed. by J. T. Kuo and G. F. Lin, pp. 1–15, Water Resources Publications, Wtteton, CO.

Plate, E. J., and Duckstein, L. (1987). Reliability in hydraulic design, in *Engineering Reliability and Risk in Water Resources*, ed. by L. Duckstein and E. J. Plate, pp. 27–60, Martinus Nijhoff, Dordrecht, The Netherlands.

Plate, E. J., and Duckstein, L. (1988). Reliability-based design concepts in hydraulic engineering, *Water Resources Bulletin*, AWRA, 24(2):234–245.

Pritchett, H. D. (1964). Application of the principles of engineering economy to the selection of highway culverts, Stanford University, Report EEP-13.

ReVelle, C. S., Loucks, D. P., and Lynn, W. R. (1968). Linear programming applied to water quality management, *Water Resources Research* 4(1):1–9.

Ritzel, B. J., Eheart, J. W., and Ranjithan, S. (1994). Using genetic algorithm to solve a multiple objective groundwater pollution containment problem, *Water Resources Research*, AGU, 30(5):1589–1603.

Schneider, V. R., and Wilson, K. V. (1980). Hydraulic design of bridges with risk analysis, Report FHWA-TS-80-226, U.S. Dept. of Transporation, Federal Highway Administration, Washington.

Schrage, L. (1986). *LINDO*, Scientific Press, Palo Alto, CA.

Sivazlian, B. D., and Stanfel, L. E. (1974). *Optimization Techniques in Operation Research*. Prentice-Hall, Englewood-Cliffs, NJ.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley and Sons, New York.

Streeter, H. W., and Phelps, E. B. (1925). A study of the pollution and natural purification of the Ohio River, Public Health Bulletin 146, U.S. Public Health Service, Washington, DC.

Taha, A. T. (1987). *Operations Research: An Introduction*, Macmillan, New York.

Tang, W. H., and Yen, B. C. (1972). Hydrologic and hydraulic design under uncertainties, in *Proceedings*, International Symposium on Uncertainties in Hydrologic and Water Resources Systems, Vol. 2, pp. 868–882, Tucson, AZ.

Tang, W. H., May, L. W., and Yen, B. C. (1975). Optimal risk-based design of storm sewer networks, *Journal of Environmental Engineering*, ASCE, 101(3):381–398.

Tang, W. H., Mays, L. W., and Wenzel, H. G. (1976). Discounted flood risks in least cost design of storm sewer networks, in *Proceedings*, Second International IAHR Symposium on Stochastic Hydraulics, Lund, Sweden.

Tang, W. H., and Yen, B. C. (1991). Dam safety inspection scheduling, *Journal of Hydraulic Engineering*, ASCE, 117(2):214–229.

Tang, W. H., and Yen, B. C. (1993). Probabilistic inspection scheduling for dams, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, pp. 107–122, ASCE, New York.

Tseng, M. T., Eiker, E. E., and Davis, D. W. (1993) Risk and uncertainty in flood damage reduction project design, in *Hydraulic Engineering '93*, ed. by H.-W., Shen, S.-T., Su, and F., Wen, pp. 2104–2109, American Society of Civil Engineers, New York.

Tseng, M. T., Knepp, A. J., and Schmalz, R. A. (1975). Evaluation of flood risk factors in the design of highway stream crossings, Vol. IV: Economic risk analysis for design of bridge waterways, Report No. FHWA-RD-75-54, U.S. Department of Transportation, Federal Highway Administration, Offices of Research and Development, Washington.

Tung, Y. K. (1986). Groundwater management by chance-constrained model, *Journal of Water Resources Planning and Management*, ASCE, 112(1):1–19.

Tung, Y. K. (1987). Effects of uncertainties on optimal risk-based design of hydraulic structures, *Journal of Water Resources Planning and Management*, ASCE, 113(5):709–722.

Tung, Y. K. (1994). Probabilistic hydraulic design: a next step to experimental hydraulics, *Journal of Hydraulic Research*, IAHR, 32(3):323–336.

Tung, Y. K., and Bao, Y. (1990). On the optimal risk-based designs of highway drainage structures, *Journal of Stochastic Hydrology and Hydraulics* 4(4):311–324.

Tung, Y. K., and Hathhorn, W. E. (1989). Multiobjective waste load allocation, *International Journal of Water Resources Management* 3:129–140.

Tung, Y. K., and Hathhorn, W. E. (1990). Stochastic waste load allocation, *Journal of Ecological Modelling*, 51(1):29–46.

Tung, Y. K., and Mays, L. W. (1980). Risk analysis for hydraulic structures, *Journal of Hydraulics Engineering*, ASCE, 106(5):893–913.

Tung, Y. K., and Mays, L. W. (1981a). Risk and reliability model for levee design, *Water Resources Research*, AGU, 17(4):833–842.

Tung, Y. K., and Mays, L. W. (1981b). Optimal risk-based design of flood levee systems, *Water Resources Research*, AGU, 17(4):843–852.

Tung, Y. K., and Mays, L. W. (1982). Optimal risk-based hydraulic design of bridges, *Journal of Water Resources Planning and Management*, ASCE, 108(2):191–202.

Tung, Y. K., and So, Y. W. (2003). Risk-based urban drainage system improvement without considering flood damage, in *Proc. of XXX IAHR Congress (Theme-B)*, pp. 533–540, Thessaloniki, Greece.

Tung, Y. K., and Yen, B. C. (2005) *Hydrosystem Engineering Uncertainty Analysis*, McGraw-Hill, New York.

Tung, Y. K., Bao, Y., Mays, L. W., and Ward, G. H. (1990). Optimization of freshwater inflows to estuary, *Journal of Water Resources Planning and Management*, ASCE. 116(4):567–585.

Tung, Y. K., Hasfurther, V., Wacker, A. M., Bao, Y., and Zhao, B. (1993). Least total expected cost (LTEC) analysis for selecting a defensible design flood frequency for highway drainage structures in Wyoming, Technical Report, Wyoming Water Resource Center, University of Wyoming, Laramie.

U.S. Army Corps of Engineers (1986). Accuracy of computed water surface profiles, Research Document 26, Hydrologic Engineering Center, Davis, CA.

U.S. Army Corps of Engineers (1996). Risk-based analysis for flood damage reduction studies, Engineer Manual EM 1110-2-1619, Washington, DC.

U.S. Water Resources Council (1983). *Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies*, Washington, U.S. Government Printing Office.

Vajda, S. (1972). *Probabilistic Programming*, Academic Press, New York.

Vanderbilt, D., and Louie, S. G. (1984). A Monte Carlo simulated annealing approach to optimization over continuous variables, *Journal of Computational Physics* 56:259–271.

Vrijling, J. K. (1993) Development in probabilistic design in flood defenses in The Netherlands, in *Reliability and Uncertainty Analyses in Hydraulic Design*, ed. by B. C. Yen and Y. K. Tung, pp. 133–178, American Society of Civil Engineers, New York.

Vrijling, J. K., van Hengel, W., and Houben, R. J. (1995). A framework for risk evaluation, *Journal of Hazard Materials* 43(3):245–261.

Wagner, B. J., and Gorelick, S. M. (1987). Optimal groundwater quality management under uncertainty, *Water Resources Research*, AGU, 23(7):1162–1172.

Wagner, B. J., and Gorelick, S. M. (1989). Reliable aquifer redediation in the presence of spatially variable hydraulic conductivity: From data to design, *Water Resources Research*, AGU, 25(10):2211–2226.

Ward, R. C., and Loftis, J. C. (1983). Incorporating the stochastic nature of water quality into management, *Journal of Water Pollution Control Federation* 55(4):408–414.

Winston, W. L. (1987). *Operations Research: Applications and Algorithms*, PWS-Kent Publisher, Boston, MA.

Wood, E. F. (1977). An analysis of flood levee reliability, *Water Resources Research*, AGU, 13(3):665–671.

Yaron, D. (1979). A method for the analysis of seasonal aspects of water quality control, *Journal of Environment, Economics and Management* 6(2):140–151.

Yen, B. C., and Ang, A. H. S. (1971). Risk analysis in design of hydraulic projects, in *Stochastic Hydraulics*, 1st International Symposium on Stochastic Hydaulics, pp. 694–709.

Yen, B. C., and Jun, B. H. (1984). Risk consideration in design of storm drains, in *Proceedings*, Third IAHR/IAWPRC International Conference on Urban Storm Drainage, 2, pp. 695–704, Chalmers University of Technology, Goteborg, Sweden.

Yen, B. C., Wenzel, H. G., Jr., Mays, L. W., and Tang, W. H. (1976). Advanced methodologies for design of storm sewer systems, Research Report, No. 112, Water Resources Center, University of Illinois at Urbana-Champaign, IL.

Yen, B. C., Cheng, S. T., and Tang, W. H. (1980). Reliability of hydraulic design of culverts, in *Proceedings*, International Conference on Water Resources Development, IAHR Asian Pacific Division Second Congress, 2, pp. 991–1001, Taipei, Taiwan.

Young, G. K., Taylor, R. S., and Costello, L. S. (1970). Evaluation of the flood risk factor in the design of box culverts, Vols. 1 and 2, U.S. Department of Transportation, Bureau of Public Roads.

Young, G. K., Childrey, M. R., and Trent, R. E. (1974). Optimal design of highway drainage culverts, *Journal of Hydraulics Engineering*, ASCE, 107(7):971–993.

Zielinski, P. A. (1988). Stochastic dissolved oxygen model, *Journal Environmental Engineering*, ASCE, 114(1):74–90.

Index

ABOUT THE AUTHOR

YEOU-KOUNG TUNG, Ph.D., is a Professor of Civil Engineering at Hong Kong University of Science and Technology. The author of numerous technical papers on hydrology and risk analysis, he has won several awards for his research on these topics including the Walter L. Huber Research Prize, ASCE; the Arthur T. Ippen Award, IAHR; and the Collingwood Prize, ASCE. Dr. Tung received his B.S. in Hydraulic Engineering from Tamkang University, Taiwan and his M.S. and Ph.D. in civil engineering from the University of Texas at Austin.

BEN-CHIE YEN, Ph.D., (deceased) was a Professor of Civil and Environmental Engineering at the University of Illinois at Champaign-Urbana. He worked with surface water and urban hydrology problems, risk and reliability analysis, and open channel and river hydraulics for more than 30 years, and was author of over 200 published technical papers and co-author of eight books. He won a number of lifetime achievement awards from various professional societies focusing on hydraulics and civil engineering including the Hunter Rouse Hydraulic Engineering Lecture, ASCE; the Ven Te Chow Memorial Lecture Award, IWRA; and Honorary Membership in IAHR. He held a B.S. in civil engineering from National Taiwan University and M.S. and Ph.D. degrees in civil engineering from the University of Iowa.

CHARLES STEVE MELCHING, Ph.D., P.E., is an Associate Professor of Civil and Environmental Engineering at Marquette University, Milwaukee, Wisconsin. He worked for the U.S. Geological Survey, Water Resources Division, for seven years prior to joining the Marquette faculty in 1999. Much of his research has been centered on the application of reliability and uncertainty analysis to water resources modeling and design. He has been honored for his research with the Walter L. Huber Research Prize, ASCE. He received his B.S. in civil engineering from Arizona State University and his M.S. and Ph.D. in civil engineering from the University of Illinois at Urbana-Champaign.