ELSEVIER

VOLUME TWO

ADVANCED MATHEMATICAL TOOLS FOR AUTOMATIC CONTROL ENGINEERS

Stochastic Techniques

ALEXANDER S. POZNYAK

Advanced Mathematical Tools for Automatic Control Engineers

To the memory of my teacher
Prof. Yakov Z. Tsypkin (1919–1997):
an outstanding scientist and great person.

Advanced Mathematical Tools for Automatic Control Engineers

Volume 2: Stochastic Techniques

Alexander S. Poznyak

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, the Netherlands

First edition 2009

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978 0 08 044673 8

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in Great Britain

09 10 10 9 8 7 6 5 4 3 2 1

Contents

This page intentionally left blank

Preface

This book contains four parts:

- **Basics of Probability**
- **Discrete Time Processes**
- **Continuous Time Processes**
- **Applications**.

The first part concerns the basics of *Probability Theory* which, in fact, is the *probability space*. The key idea behind probability space is the stabilization of the relative frequencies when one performs 'independent' repetition of a random experiment and records whether each time 'event', say A, occurs or not. Define the *characteristic function of event* A during trial $t = 1, 2, \ldots$ by $\chi(A_t)$, namely,

$$\boxed{\chi(A_t) := \begin{cases} 1 & \text{if } A \text{ occurs, i.e., } A_t = A \\ 0 & \text{if not} \end{cases}} \tag{1}$$

Denoted by

$$\boxed{r_n(A) := \frac{1}{n} \sum_{t=1}^{n} \chi(A_t)} \tag{2}$$

the relative frequency of event A after the first n trials, because of the dawn of history one can observe the stabilization of the relative frequencies; that is, it seems natural that as $n \to \infty$

$r_n(A)$ *converges to some real number called the probability of* A.

Although games of chance have been performed for thousands of years, probability theory, as a *science*, originated in the middle of the 17th century with Pascal (1623–1662), Fermat (1601–1655) and Huygens (1629–1695). The real history of probability theory began with the works of James Bernoulli (1654–1705) and De Moivre (1667–1754). Bernoulli was probably the first to realize the importance of consideration of infinite sequences of random trials and made a clear distinction between the probability of an event and the frequency of its realization. In 1812 there appeared Laplace's (1749–1827) great treatise containing the analytical theory of probability with application to the analysis of observation errors. Then limit theorems were studied by Poisson (1781–1840) and Gauss (1777–1855).

The next important period in the development of probability theory is associated with the names of P.L. Chebyshev (1857–1894), A.A. Markov (1856–1922) and A.M. Lyapunov

(1857–1918), who developed effective methods for proving limit theorems for sums of independent but arbitrarily distributed random variables. Before Chebyshev the main interest had been in the calculation of the probabilities of random events. He, probably, was the first to understand clearly and exploit the full strength of the concepts of random variables. The number of Chebyshev's publications on probability theory is not large – four in all – but it would be hard to overestimate their role and in the development of the *classical Russian school* of that subject.

The *modern period* in the development of probability theory began with its axiomatization due to the publications of S.N. Bernstein (1880–1968), R. von Mises (1883–1953) and E. Borel (1871–1956). But the first mathematically rigorous treatment of probability theory came only in the 1930s by the Russian mathematician A.N. Kolmogorov (1903–1987) in his seminal paper (Kolmogorov, 1933). His first observation was that a number of rules that hold for relative frequencies $r_n(A)$ should also hold for probabilities. This immediately raises the question: *which is the minimal set of such rules?* According to Kolmogorov, the answer is based on several axiomatic concepts. These fundamental concepts are:

(1) *σ-algebra*

(2) *probability measure*

(3) *probability space*

(4) *distribution function*.

In this volume we will discuss each of these principal notions in details.

The second part deals with discrete time processes, or, more exactly, random sequences where the main role is played by *Martingale Theory*, which takes a central place in Discrete-Time Stochastic Process Theory because of the asymptotic properties of martingales providing a key prototype of probabilistic behavior which is of wide applicability. The first appearance of a martingale as a mathematical term was due to J. Ville (1939). The major breakthrough was associated with the classic book *Stochastic Processes* by J. Doob (1953). Other recent books are J. Neveu (1975), R. Liptser and A. Shiryaev (1989) and D. Williams (1991). The martingale is a sequence $\{\xi_n, \mathcal{F}_n\}_{n\geq 1}$ of random variables ξ_n associated with a corresponding prehistory (*σ-algebra*) $\mathcal{F}_{n-1}$ such that the conditional mathematical expectation of ξ_n under fixed $\mathcal{F}_{n-1}$ is equal to ξ_{n-1} with probability 1, that is,

$$\mathsf{E}\{\xi_n/\mathcal{F}_{n-1}\} \overset{a.s.}{=} \xi_{n-1}$$

Martingales are probably the most inventive and generalized of sums of independent random variables with zero-mean. Indeed, any random variable ξ_n (maybe, dependent) can be expressed as a sum of 'martingale-differences'

$$\Delta_n := \mathsf{E}\{\xi_n/\mathcal{F}_k\} - \mathsf{E}\{\xi_n/\mathcal{F}_{k-1}\}, \quad \mathsf{E}\{\xi_n/\mathcal{F}_0\} = 0, k = 1, \ldots, n$$

because of the representation

$$\xi_n = (\xi_n - \mathsf{E}\{\xi_n/\mathcal{F}_{n-1}\}) + (\mathsf{E}\{\xi_n/\mathcal{F}_{n-1}\} - \mathsf{E}\{\xi_n/\mathcal{F}_{n-2}\}) + \ldots + (\mathsf{E}\{\xi_n/\mathcal{F}_1\} - \mathsf{E}\{\xi_n/\mathcal{F}_0\}) = \sum_{k=1}^{n} \Delta_k$$

In some sense martingales occupy the intermediate place between independent and dependent sequences. The independence assumption has proved inadequate for handling contemporary developments in many fields.

Then, based on such considerations, there are presented and discussed in detail the three most important probabilistic laws: the *Weak Law of Large Numbers* (**LLN**) and its strong version known as the *Strong Law of Large Numbers* (**SLLN**), the *Central Limit Theorem* (**CLT**), and, finally, the *Law of the Iterated Logarithm* (**LIL**). All of them may be interpreted as *invariant principles* or *invariant laws* because of the independence of the formulated results of the distribution of random variables forming considered random sequences.

The third part discusses *Continuous-Time Processes* basically governed by stochastic differential equations. The notion of the mean-square continuity property is introduced along with its relation with some properties of the corresponding auto-covariance matrix function. Then processes with orthogonal and independent Increments are introduced, and, as a particular case, the Wiener process or Brownian motion is considered. A detailed analysis and discussion of an invariance principle and LIL for Brownian motion are presented.

The so-called *Markov Processes* are then introduced. A stochastic dynamic system satisfies the Markov property if the probable (future) state of the system is independent of the (past) behavior of the system. The relation of such systems to diffusion processes is deeply analyzed. The ergodicity property of such systems is also discussed.

Next, the most important constructions of stochastic integrals are studied: namely

• *a time-integral* of a sample path of a second order (s.o.) stochastic process;

• the so-called *Wiener integral* of a deterministic function with respect to a stationary orthogonal increment random process such that this integral is associated with the Lebesgue integral, it is usually referred to as a stochastic integral with respect to an *'orthogonal random measure'* ;

• the so-called *Itô integral* of a random function with respect to a stationary orthogonal increment random process;

• and, finally, the so-called *Stratonovich integral* of a random function with respect to an s.o. stationary orthogonal increment random process where the 'summation' on the right-hand side is taken in a special sense.

All of these different types of stochastic integral are required for the mathematically rigorous definition of a solution of a stochastic differential equation. We discuss the class of the so-called *Stochastic Differential Equation*, introduced by K. Itô, whose basic theory was developed independently by Itô and I. Gihman during 1940s. There the Itô-type integral calculus is applied. The principal motivation for choosing the *Itô approach* (as opposed to the Stratonovich calculus as another very popular interpretation of stochastic integration) is that the Itô method extends to a broader class of equations and transformation the probability law of the Wiener process in a more natural way. This approach implements the so-called *diffusion approximation* which arises from random difference equation models and has wide application to control problems in engineering sciences, motivated by the need for more sophisticated models which spurred further work on these types of equation in the 1950s and 1960s.

The fourth part is dedicated to applications. The vitality and principal source of inspiration of Probability Theory comes from its applications. The mathematical modeling of physical reality and the inherent ‘nondeterminism’ of many systems provide an expanding domain of rich pickings in which, in particular, martingale limit results are demonstrably of great usefulness. The applications presented here reflect the author’s interest, although it is hoped that they are diverse enough to establish beyond any doubt the usefulness of the methodology.

First, we apply the stochastic technique presented to *Identification of Dynamics Models* containing stochastic perturbations. Here, identification is associated with on-line parameters estimating of some practically important models such as NARMAX and other autoregression widely applied in mathematical economics over last two decades.

Then the problems of *Filtering*, *Prediction* and *Smoothing* are considered. We concentrate on the problem of estimation for discrete-time and continuous-time processes based on some available observations of these processes which are obviously statistically dependent on the processes considered. The physical device generating any one of these estimates from the observed data is called *a filter*. In detail, the so-called *Kalman filter* is analyzed and discussed.

In certain statistical applications (such as bioassay, sensitivity testing, or fatigue trials) some problems arising can be conveniently attacked using the so-called *Stochastic Approximation Method* (SAM), which requires minimum distributional assumptions. SAM is closely related to recursive *Least Squares* and to the estimation of parameters of a nonlinear regression. The control engineering literature also contains many applications of SAM, basically related to identification problems (see, for example, Ya.Z. Tsypkin, 1971 and G. Saridis, 1977). Quite a large number of stochastic approximation schemes have been discussed in the literature, but they essentially amount to modifications of two basic schemes: the *Robbins–Monro procedure* dealing with a nonlinear regression problem when only measurements of a regression function corrupted by noise are available, and the *Kiefer–Wolfowitz procedure* dealing with an optimization problem when only measurements of a function to be optimized corrupted by noise are available in any predetermined point. We present some extensions of these methods related to the stochastic gradient algorithm, its robustification and the conditions when these procedures work under correlated (dependent) noises.

Finally, a version of the *Robust Stochastic Maximum Principle* (*RSMP*) is designed, being applied to the min-max Mayer Problem formulated for stochastic differential equations with the control-dependent diffusion term. The parametric families of first and second order adjoint stochastic processes are introduced to construct the corresponding Hamiltonian formalism. The Hamiltonian function used for the construction of the robust optimal control is shown to be equal to the Lebesgue integral over a parametric set (which may be a compact or a finite set) of the standard stochastic Hamiltonians corresponding to a fixed value of the uncertain parameter. Robust stochastic LQ control designing is discussed in detail.

Acknowledgements

I feel deep gratitude to my colleagues and students over the years, without whose evident collaboration, comments, enjoyment and expressed appreciation the current work would not have been undertaken. I wish also to acknowledge the editors of Elsevier Limited for being so cooperative during the production process of both volumes of the book.

Alexander S. Poznyak
Avandaro, Mexico

This page intentionally left blank

Notations and Symbols

$A \cup B := \{x : x \in A \text{ or } x \in B\}$ — the union of sets.

$A \cap B := \{x : x \in A \text{ and } x \in B\}$ — the intersection of sets.

$A^c = \bar{A} := \{x : x \notin A\} := \{z \in \mathbb{C} \mid \operatorname{Re} z < 0\}$ — the complement to a set A.

$A \setminus B := A \cap B^c$ — the difference of sets.

$A - B := (A \setminus B) \cup (B \setminus A)$, $A \triangle B := (A - B) \cup (B - A)$ — the symmetric difference of sets.

$\varnothing$ — the empty set.

$\mathfrak{P}(\Omega) := \{A : A \subset \Omega\}$ — the power set.

$\lim_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} A_n$ for $A_1 \subset A_2 \subset \cdots$.

$\lim_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} A_n$ for $A_1 \supset A_2 \supset \cdots$.

$A_* = \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$.

$A^* = \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$.

$(\Omega, \mathcal{F})$ — a measurable space.

$\mu = \mu(A) \in [0, \infty)$ — a finite additive measure.

$(\Omega, \mathcal{F}, \mathsf{P})$ — a probability space.

P — a probability measure.

$\mathcal{B} := \mathcal{B}(\mathbb{R})$ — Borel σ-algebra.

$F = F(x)$ — a distribution function.

$\mathsf{P}\{(a, b]\} := F(b) - F(a)$.

$f(x) = \frac{d}{dx} F(x) \geq 0$ — the density function.

$\mathsf{P}\{A/B\} := \mathsf{P}\{B \cap A\} / \mathsf{P}\{B\}$ ($\mathsf{P}\{B\} > 0$) the conditional probability.

$\xi = \xi(\omega)$, defined on $(\Omega, \mathcal{F})$ $(\omega : \xi(\omega) \in B) \in \mathcal{F}$ — $\mathcal{F}$-measurable (or, Borel measurable) function or random variable.

$I_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$ — the indicator-function of a set A.

$\xi^+ := \max\{\xi; 0\}$.

$\xi^- := -\min\{\xi; 0\}$.

$\mathsf{E}\{\xi\} := \int\limits_{\omega\in\Omega} \xi(\omega)\mathsf{P}(d\omega) \overset{\text{or}}{=} \int\limits_{\Omega} \xi d\mathsf{P}$ — mathematical expectation.

$m := \mathsf{E}\{\xi\}$.

$\operatorname{var} \xi := \mathsf{E}\left\{(\xi - \mathsf{E}\{\xi\})^2\right\}$.

$\operatorname{med}(\xi) : \mathsf{P}\{\xi \le \operatorname{med}(\xi)\} = \mathsf{P}\{\xi > \operatorname{med}(\xi)\} = \frac{1}{2}$ — the median.

$\lambda_\alpha(\xi) : \mathsf{P}\{\xi \ge \lambda_\alpha(\xi)\} \ge \alpha$ — the α-quantile.

$\operatorname{cov}(\xi, \eta) := \mathsf{E}\{(\xi - \mathsf{E}\{\xi\})(\eta - \mathsf{E}\{\eta\})\}$ — the covariance of ξ and η.

$\rho_{\xi,\eta} := \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\operatorname{var}\xi}\sqrt{\operatorname{var}\eta}}$ — the correlation coefficient of ξ and η.

$g_\cup : \mathbb{R} \to \mathbb{R}$ — a convex downward (or, simply, convex) function.

$g_\cap : \mathbb{R} \to \mathbb{R}$ — a convex upward (or, simply, concave) function.

$\varphi(t) := \int\limits_{\mathbb{R}^n} e^{i(t,x)} dF(x) = \varphi(t) := \mathsf{E}\left\{e^{t\xi}\right\}$ — the characteristic function of n-dimensional distribution function $F = F(x_1, x_2 \ldots, x_n)$ given on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, or equivalently, a random vector ξ having this distribution.

$\xi_k \overset{a.s.}{\to} 0$ — the convergence with probability 1 (or, almost sure).

$\xi_n \uparrow \xi$ — monotonically (non-decreasing) converges to ξ.

$\xi_n \downarrow \xi$ — monotonically (non-increasing) converges to ξ.

$\xi_n \underset{n\to\infty}{\overset{d}{\to}} \xi$ — the convergence in distribution.

$\xi_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} \xi$ — the convergence in probability.

$\xi_n \underset{n\to\infty}{\overset{L^p}{\to}} \xi$ — the convergence in mean of the order p.

$\underset{n\to\infty}{\text{l.i.m.}}\, \xi_n = \xi$ — the mean-square convergence.

$\xi_n \overset{\forall \omega \in \Omega}{\underset{n\to\infty}{\to}} \xi$ — the pointly convergence.

$\xi_n \overset{\text{c.c.}}{\underset{n\to\infty}{\to}} \xi$ — the complete convergence.

$d(F, G) := \sup_{A\in\mathbb{R}} |F(A) - G(A)|$ — the variational distance between the distributions F and G.

$d(\xi, \eta) := \sup_{A\in\mathbb{R}} |\mathsf{P}\{\xi \in A\} - \mathsf{P}\{\eta \in A\}|$ — the distributional distance between two random variables ξ and η.

$\eta(\omega) = \mathsf{E}\{\xi/\mathcal{F}_0\}$ — the conditional mathematical expectation of $\xi(\omega)$ with respect to a sigma-algebra $\mathcal{F}_0$.

$\mathcal{F}_n := \sigma(x_1, x_2, \ldots, x_n)$ — a sigma-algebra constructed from all sets $\{\omega : x_i(\omega) \le c_i, i = 1, 2, \ldots, n\}$ where $x_i = x_i(\omega)$ $(i = 1, 2, \ldots, n)$ are random variables, defined on $(\Omega, \mathcal{F}, \mathsf{P})$, and $c_i \in \mathbb{R}$ are any constants.

$\langle M \rangle_n = \sum_{j=1}^{n} \mathsf{E}\left\{(\Delta M_j)^2 \mid \mathcal{F}_{j-1}\right\} \left(\Delta M_j := M_j - M_{j-1}\right)$ — the quadratic variation of a square integrable martingale.

$\tau \wedge n := \min\{\tau, n\}$.

$[z]^+ := \begin{cases} z & \text{if } z \ge 0 \\ 0 & \text{if } z < 0 \end{cases}.$

$x^+ := \max\{0; x\}$.

$x^- := -\min\{0; x\}$.

$\mathcal{N}(0, 1)$ — the standard Gaussian distribution with zero-mean and variance equal to one.

$\alpha(\mathcal{H}, \mathcal{G}) := \sup_{A\in\mathcal{H},\, B\in\mathcal{G}} |\mathsf{P}\{A \cap B\} - \mathsf{P}\{A\}\mathsf{P}\{B\}|$ — the coefficient of strong mixing.

$\phi(\mathcal{H}, \mathcal{G}) := \sup_{A\in\mathcal{H},\, B\in\mathcal{G}} |\mathsf{P}\{A/B\} - \mathsf{P}\{A\}|$ — the coefficient of uniform strong mixing .

$\psi(\mathcal{H}, \mathcal{G}) := \sup_{A\in\mathcal{H},\, B\in\mathcal{G},\, \mathsf{P}\{A\}>0,\, \mathsf{P}\{B\}>0} \left| \frac{\mathsf{P}\{A \cap B\}}{\mathsf{P}\{A\}\mathsf{P}\{B\}} - 1 \right|$ — the coefficient of relative uniform strong mixing.

$\rho(\mathcal{H}, \mathcal{G}) := \sup_{x\in\mathcal{H},\, y\in\mathcal{G}} \left|\rho_{x,y}\right|$ — the coefficient of correlative mixing.

$\bigvee_{i=m}^{n} \mathcal{F}_i$ — the sigma-algebra generated by the union of the sigma-algebras $\mathcal{F}_i$.

$\alpha_n := \sup_{k\ge1} \alpha(\bigvee_{i=1}^{k} \mathcal{F}_i, \bigvee_{i=k+n}^{\infty} \mathcal{F}_i)$ — strong mixing coefficient.

$\phi_n := \sup_{k\geq 1} \phi(\bigvee_{i=1}^{k} \mathcal{F}_i, \bigvee_{i=k+n}^{\infty} \mathcal{F}_i)$ — the uniform strong mixing coefficient.

$\psi_n := \sup_{k\geq 1} \psi(\bigvee_{i=1}^{k} \mathcal{F}_i, \bigvee_{i=k+n}^{\infty} \mathcal{F}_i)$ — the relative uniform strong mixing coefficient.

$\rho_n := \sup_{k\geq 1} \rho(\bigvee_{i=1}^{k} \mathcal{F}_i, \bigvee_{i=k+n}^{\infty} \mathcal{F}_i)$ — the correlative mixing coefficient.

$q_{n,m} := \sqrt{\mathsf{E}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}|^2\right\}}$ — the quadratic norm of a conditional mathematical expectation.

$\mathrm{col} X := \left(x_{1,1}, \ldots, x_{1,N}, \ldots, x_{M,1}, \ldots, x_{M,N}\right)^{\mathsf{T}}$ — the spreading operator.

$\otimes$ is the Kronecker product.

$\mathring{N}_t := N_t - \mathsf{E}\{N_t\}$ — a centered random variable.

$\{W_t(\omega)\}_{t\geq 0}$, $W_0(\omega) \overset{a.s.}{=} 0$ — a Wiener process or Brownian motion (BM) is a zero-mean s.o. scalar process with stationary normal independent increments.

$\mathcal{F}_{[t_1,t_2]} := \sigma\{x(t,\omega), t_1 \leq t \leq t_2\}$ — a minimal sigma-algebra generated by the 'intervals' (rectangles, etc.).

$\mathsf{P}\{s, x, t, A\}$ the transition probability or transition function of a stochastic process $\{x(t,\omega)\}_{t\in[t_0,T]}$ where $t_0 \leq s \leq t \leq T$, $x \in \mathbb{R}^n$ and $A \in \mathcal{B}^n$.

$\pi_{i,j}(s,t) := \mathsf{P}\{x(t,\omega) = j \mid x(s,\omega) = i\}$ — the transition probabilities of a given Markov chain defining the conditional probability for a process $\{x(t,\omega)\}_{t\in\mathcal{T}}$ to be in the state j at time t under the condition that it was in the state i at time $s < t$.

$I_{[a,b]}(x) = \int_{\tau=a}^{b} x(\tau,\omega)\, d\tau$ — the time-integral of a sample path of a second order (s.o.) stochastic process $\{x(t,\omega)\}_{t\geq 0}$.

$I^W_{[a,b]}(f) = \int_{\tau=a}^{b} f(\tau)\, dW_\tau(\omega)$ — the Wiener integral of a deterministic function $f(t)$ with respect to an s.o. stationary orthogonal increment random process $W_\tau(\omega)$.

$I_{[a,b]}(g) = \int_{\tau=a}^{b} g(\tau,\omega)\, dW_\tau(\omega)$ — the Itô integral of a random function $g(t,\omega)$ with respect to an s.o. stationary orthogonal increment random process $W_t(\omega)$.

$I^S_{[a,b]}(g) = \int_{\tau=a}^{b} g(\tau,\omega) \overset{\lambda=1/2}{\circ} dW_\tau(\omega)$ — the Stratonovich integral of a random function $g(t,\omega)$ with respect to an s.o. stationary orthogonal increment random process $W_t(\omega)$, where the 'summation' in the right-hand side is taken in a special sense.

$\mathcal{W}_{[0,t]} := \sigma \{x(0,\omega), W_s(\omega), 0 \leq s \leq t\}$ — is a minimal sigma-algebra (sub-sigma-algebra of $\mathcal{F}$) generated by $x(0,\omega)$ and $W_s(\omega)$ $(0 \leq s \leq t)$.

$o_\omega(1) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ — a random sequence tending to zero with probability one (a.s.).

$q : qz_n = z_{n-1}$ — the delay operator.

$\mathbb{I}_F(c,n) := \mathsf{E}\left\{\nabla_c \ln p_{y^n}(y^n \mid c) \nabla_c^{\mathsf{T}} \ln p_{y^n}(y^n \mid c)\right\}$ — the Fisher information matrix.

$R_{\mathring{y},\mathring{y}} := \mathsf{E}\{\mathring{y}\mathring{y}^{\mathsf{T}}\}$ — an auto-covariation matrix.

This page intentionally left blank

List of Figures

This page intentionally left blank

List of Tables

PART I

Basics of Probability

This page intentionally left blank

1 Probability Space

Contents

1.1 Set operations, algebras and sigma-algebras

It will be convenient to start with some useful definitions in algebra of sets. This will serve as a refresher and also as a way of collecting a few important facts that we will often use throughout.

1.1.1 Set operations, set limits and collections of sets

Let $A, A_1, A_2, \ldots$ and $B, B_1, B_2, \ldots$ be sets.

Definition 1.1. *There are defined the following operations over sets:*

1. *the* ***union*** *of sets:*

$$\boxed{A \cup B := \{x : x \in A \text{ or } x \in B\}} \tag{1.1}$$

2. *the* ***intersection*** *of sets:*

$$\boxed{A \cap B := \{x : x \in A \text{ and } x \in B\}} \tag{1.2}$$

3. *the* ***complement*** *to a set:*

$$\boxed{A^c = \bar{A} := \{x : x \notin A\}} \tag{1.3}$$

4. *the* ***difference*** *of sets:*

$$\boxed{A \setminus B := A \cap B^c} \tag{1.4}$$

5. *the* ***symmetric difference*** *of sets:*

$$\boxed{A - B := (A \setminus B) \cup (B \setminus A)} \tag{1.5}$$

Also some additional terminology (definitions) will be used.

Definition 1.2.

1. *The* ***empty set*** $\varnothing$ *is a set which does not contain elements.*
2. A *is a* ***subset*** *of* B*, denoted by* $A \subset B$*, if* $x \in A$ *then obligatory* $x \in B$*.*
3. *The* ***power set*** $\mathfrak{P}(\Omega)$*, generated by a set* Ω*, is a collection of all subsets of* Ω*, that is,*

$$\boxed{\mathfrak{P}(\Omega) := \{A : A \subset \Omega\}} \tag{1.6}$$

4. *A sequence* $\{A_n, n \geq 1\}$ *of sets is non-decreasing, that denoted by* $\{A_n \nearrow\}$*, if*

$$\boxed{A_1 \subset A_2 \subset \cdots} \tag{1.7}$$

5. *A sequence* $\{A_n, n \geq 1\}$ *of sets is non-increasing, that denoted by* $\{A_n \searrow\}$*, if*

$$\boxed{A_1 \supset A_2 \supset \cdots} \tag{1.8}$$

Lemma 1.1. (the de Morgan formulas) *The following relations hold:*

$$\boxed{\left(\bigcup_{k=1}^{n} A_k\right)^c = \bigcap_{k=1}^{n} A_k^c, \quad \left(\bigcap_{k=1}^{n} A_k\right)^c = \bigcup_{k=1}^{n} A_k} \tag{1.9}$$

Proof. The first formula reflects the fact that an element that does not belongs to any A_k whatsoever belongs to all complements, and therefore to their intersection. The second formula means that an element that does not belong to every A_k belongs to at least one of the complements. □

Corollary 1.1. *The formula (1.9) for the case of two sets is as follows:*

$$\boxed{\overline{A_1 \cup A_2} = \bar{A}_1 \cap \bar{A}_2; \quad \overline{A_1 \cap A_2} = \bar{A}_1 \cup \bar{A}_2} \tag{1.10}$$

and may be easily illustrated by the so-called Venn diagram (see Fig. 1.1).

Define below the **limits of sets**. Not every sequence of sets has a limit. One possibility when a limit exists is the 'monotone' sequences of sets (see (1.7) and (1.8)).

Definition 1.3. *Let* $\{A_n, n \geq 1\}$ *be a sequence of subsets from* Ω*.*

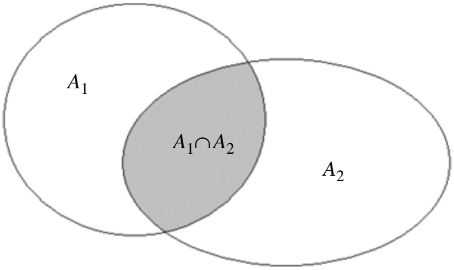

Fig. 1.1. Venn diagram.

1. *If it is non-decreasing, i.e.,* $A_1 \subset A_2 \subset \cdots$, *then define a limit as*

$$\boxed{\lim_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} A_n} \tag{1.11}$$

2. *If it is non-increasing, i.e.,* $A_1 \supset A_2 \supset \cdots$, *then define a limit as*

$$\boxed{\lim_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} A_n} \tag{1.12}$$

This definition looks correct since both $\bigcup_{n=1}^{\infty} A_n$ and $\bigcap_{n=1}^{\infty} A_n$ always exist.

Definition 1.4. *Let* $\{A_n, n \geq 1\}$ *be a sequence of subsets from* Ω. *Then by the definition*

$$\boxed{\begin{aligned} A_* &= \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \\ A^* &= \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \end{aligned}} \tag{1.13}$$

Remark 1.1. *The following inclusion holds:*

$$\boxed{A_* \subseteq A^*} \tag{1.14}$$

Indeed, since $\bigcap_{m=n}^{\infty} A_m \subset \bigcup_{m=n}^{\infty} A_m$, *one has*

$$\left(\bigcap_{m=2}^{\infty} A_m\right) \cup \left(\bigcap_{m=1}^{\infty} A_m\right) = \bigcap_{m=2}^{\infty} A_m \subset \bigcup_{m=2}^{\infty} A_m = \left(\bigcup_{m=2}^{\infty} A_m\right) \cap \left(\bigcup_{m=1}^{\infty} A_m\right)$$

and, hence,

$$\left(\bigcap_{m=n}^{\infty} A_m\right) \cup \cdots \cup \left(\bigcap_{m=1}^{\infty} A_m\right) \equiv \bigcap_{m=n}^{\infty} A_m \subset$$
$$\bigcup_{m=n}^{\infty} A_m = \left(\bigcup_{m=n}^{\infty} A_m\right) \cap \cdots \cap \left(\bigcup_{m=1}^{\infty} A_m\right)$$

Taking limits in both sides of the last inclusion implies (1.14).

1.1.2 Algebras and sigma-algebras

Let $\Omega \triangleq \{\omega\}$ be a set of points ω and be named as *a sample space* or *space of elementary events*.

The next definition concern the notion of an algebra and a σ-algebra of all possible events generated by a given set Ω of elementary events ω.

Definition 1.5. *A system $\mathcal{A}$ of subsets of Ω is called **an algebra** if*

1.

$$\boxed{\Omega \subset \mathcal{A}}$$

2. *for any finite $n < \infty$ and for any subsets $A_i \in \mathcal{A}$ $(i = 1, \ldots, n)$*

$$\boxed{\bigcup_{i=1}^{n} A_i \in \mathcal{A}, \quad \bigcap_{i=1}^{n} A_i \in \mathcal{A}}$$

3. *for all $A \in \mathcal{A}$ its complement $\bar{A}$ is also from $\mathcal{A}$, that is,*

$$\boxed{\bar{A} \triangleq \{\omega \in \Omega | \omega \notin A\} = \Omega \setminus A \in \mathcal{A}}$$

Corollary 1.2. *$\mathcal{A}$ contains the **empty set** $\varnothing$, that is,*

$$\boxed{\varnothing \subset \mathcal{A}}$$

Proof. Indeed, if $\mathcal{A}$ contains some set A, then by (3), it contains also $\bar{A} := \Omega \setminus A$. But by (2), $\mathcal{A}$ contains their union $A \cup \Omega \setminus A = \Omega$ and its complement $\bar{\Omega} := \Omega \setminus \Omega = \varnothing$. □

Definition 1.6. *The collection $\mathcal{F}$ of subsets from Ω is called **an σ-algebra** (a power set) or an **event space** if*

1. *it is not empty, i.e.,*

$$\boxed{\mathcal{F} \neq \varnothing}$$

2. *it is algebra;*

3. *for any sequences of subsets* $\{A_i\}$, $A_i \in \mathcal{F}$ *it follows*

$$\boxed{\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}, \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}}$$

If, for example, Ω is a set whose points correspond to the possible outcomes of a random experiment, certain subsets of Ω will be called *'events'*. Intuitively, we can consider A as an event if the question *'Does any elementary realization* ω *really belong to a set* A*?'* has a definite yes or no answer after the experiment is performed and *'the output'* corresponds to the point $\omega \in \Omega$. So, if we can answer the question *'Is really* $\omega \in A$*?'* we can certainly answer the question *'Is really* $\omega \in \bar{A}$*?'*, and if for each $i = 1, \dots, n$ we can decide whether or not ω belongs to A_i, then we can definitely determine whether or not ω belongs to $\bigcup_{i=1}^{n} A_i$ or $\bigcap_{i=1}^{n} A_i$.

The next example illustrates this point of view and concentrates on the physical sense of the considered class of sets.

Example 1.1. (A toss of two coins) *Two fair coins are continuously flipped. The elementary events corresponding to the described situation can be presented as follows:*

$$\Omega = \{\omega\} = \{\underbrace{HH}_{\omega_1}, \underbrace{HT}_{\omega_2}, \underbrace{TH}_{\omega_3}, \underbrace{TT}_{\omega_4}\}$$

where the elementary event, for example, HT*, means that the first coin flipped with the pattern* H *and the second one flipped with the pattern* T*. Consider now a fixed number* n *of two coins tossed and define an event* A_1 *by the following way: we will say that the event* A_1 *has been realizable during the experiment (toss)* i*, if there appeared at least one* H*, i.e.,*

$$A_1^{(i)} \stackrel{\Delta}{=} \{\omega^{(i)} \in \Omega : \omega^{(i)} = \omega_1 = HH, \omega^{(i)} = \omega_2 = HT, \omega^{(i)} = \omega_3 = TH\}$$

The next events have the following interpretation:

-
$$\bar{A}_1^{(i)} \stackrel{\Delta}{=} \{\omega^{(i)} \in \Omega : \omega^{(i)} = TT = \omega_4\}$$

-
$$\bigcup_{i=1}^{n} A_1^{(i)} = \sum_{i=1}^{n} \left[\chi(\omega^{(i)} = \omega_1) + \chi(\omega^{(i)} = \omega_2) + \chi(\omega^{(i)} = \omega_3)\right]$$

  where the characteristic function (an indicator) $\chi(\cdot)$ *of an event is defined by (1) and in this case is as follows:*

$$\chi(\omega^{(i)} = \omega_k) \stackrel{\Delta}{=} \begin{cases} 1 & \text{if } \omega^{(i)} = \omega_k \\ 0 & \text{if } \omega^{(i)} \neq \omega_k \end{cases}$$

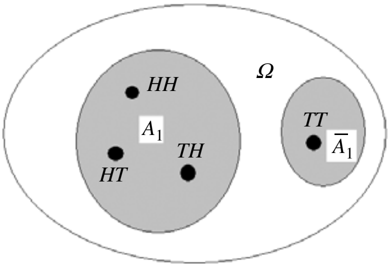

Fig. 1.2. The decomposition of a sample space Ω into subsets $A_1^{(i)}$ and $\bar{A}_1^{(i)}$, i.e., $\Omega = A_1^{(i)} \cup \bar{A}_1^{(i)}$.

i.e., the event $\bigcup_{i=1}^{n} A_1^{(i)}$ *consists of the fact that after n experiments (tosses) at least one H has appeared in its realization:*

- $\bigcap_{i=1}^{n} A_1^{(i)} = \prod_{i=1}^{n} \left[\chi(\omega^{(i)} = \omega_1) + \chi(\omega^{(i)} = \omega_2) + \chi(\omega^{(i)} = \omega_3)\right]$ *i.e., at each experiment there appears at least one H.*
- *It is easy to conclude that for any* $i = 1, \ldots, n$

$$\Omega = \{A_1^{(i)}, \bar{A}_1^{(i)}\}$$

(see Fig. 1.2). This example shows that σ-algebra $\mathcal{F}$ in this case presents the combination of the following sets:

$$\mathcal{F} \triangleq \left\{ A_1^{(1)}, \bar{A}_1^{(1)}, A_1^{(2)}, \bar{A}_1^{(2)}, \ldots, A_1^{(n)}, \bar{A}_1^{(n)}, \right.$$
$$A_1^{(1)} \cup A_1^{(2)}, A_1^{(1)} \cup \bar{A}_1^{(2)}, \bar{A}_1^{(1)} \cup A_1^{(2)}, \bar{A}_1^{(1)} \cup \bar{A}_1^{(2)},$$
$$A_1^{(1)} \cap A_1^{(2)}, A_1^{(1)} \cap \bar{A}_1^{(2)}, \bar{A}_1^{(1)} \cap A_1^{(2)}, \bar{A}_1^{(1)} \cap \bar{A}_1^{(2)},$$
$$\overline{A_1^{(1)} \cup A_1^{(2)}}, \overline{A_1^{(1)} \cup \bar{A}_1^{(2)}}, \overline{\bar{A}_1^{(1)} \cap A_1^{(2)}}, \overline{\bar{A}_1^{(1)} \cap \bar{A}_1^{(2)}},$$
$$\cdots$$
$$\bigcup_{i=1}^{n} A_1^{(i)}, \overline{\bigcup_{i=1}^{n} A_1^{(i)}}, \bigcap_{i=1}^{n} A_1^{(i)}, \overline{\bigcap_{i=1}^{n} A_1^{(i)}}, \ldots,$$
$$\bigcup_{i=1}^{n} \bar{A}_1^{(i)}, \overline{\bigcup_{i=11}^{n} \bar{A}_1^{(i)}}, \bigcap_{i=1}^{n} \bar{A}_1^{(i)}, \overline{\bigcap_{i=1}^{n} \bar{A}_1^{(i)}}, \ldots,$$
$$\cdots$$
$$\left. \bigcup_{i=1}^{\infty} A_1^{(i)}, \overline{\bigcup_{i=1}^{\infty} A_1^{(i)}}, \bigcap_{i=1}^{\infty} A_1^{(i)}, \overline{\bigcap_{i=1}^{\infty} A_1^{(i)}} \right\}$$

i.e. $\mathcal{F}$ (σ-algebra of events initiated by an elementary set Ω) consists of all possible results of infinite numbers of experiments $i = 1, 2, \ldots$.

Remark 1.2. *Let an $\mathcal{A}$ be a collection of subsets of Ω. Evidently the power set $\mathfrak{P}(\Omega)$ (1.6) is a σ-algebra. But then, there exists at least one σ-algebra containing $\mathcal{A}$. Since, moreover,* *the intersection of any numbers of σ-algebras is, again, a σ-algebra, there exists a* ***smallest σ-algebra*** *containing $\mathcal{A}$. In fact, the smallest σ-algebra containing $\mathcal{A}$ equals*

$$\boxed{\mathcal{F}^* = \bigcap_{\mathcal{G} \in \{\sigma\text{-algebra} \supset \mathcal{A}\}} \mathcal{G}} \tag{1.15}$$

Below, we will always assume that we deal with a smallest σ-algebra $\mathcal{F}^*$ when we mention a σ-algebra.

1.2 Measurable and probability spaces

Length, area and volume, as well as probability, are an instance of the measure concept that we are going to discuss in this section. In general, the *measure* is a set function which defines an assignment of a number $\mu(A)$ to each set A of events in a certain class. Some structure must be imposed on the class of sets on which μ is defined and probability considerations provide a good motivation for the type of structure required.

1.2.1 Measurable spaces and finite additive measures

Definition 1.7. *The pair $(\Omega, \mathcal{F})$ is called* ***a measurable space****.*

The definition given above presents only the notion commonly used in mathematical literature, not more. But the next one establishes the central definitions of this book that play a key role in Probability and Stochastic Processes theories.

Definition 1.8. *Any nonnegative function of set $A \in \mathcal{F}$*

$$\boxed{\mu = \mu(A) \in [0, \infty)} \tag{1.16}$$

is called ***a finite additive measure*** *if for any finite collection $\{A_i, i = 1, \ldots, n\}$ of all pairwise disjoint subsets*

$$\boxed{A_1, A_2, \ldots, A_n \qquad (A_i \cap A_j = \varnothing)} \tag{1.17}$$

the following properties hold:

- $$\boxed{\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i)} \tag{1.18}$$

- $$\boxed{\mu(\Omega) < \infty} \tag{1.19}$$

If, keeping (1.19), for any collection $\{A_n, n \geq 1\}$ of pair-wise disjoint subsets instead of (1.18) the following property holds:

$$\boxed{\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)} \tag{1.20}$$

then such measure is called ***a countable additive measure****.*

1.2.2 The Kolmogorov axioms and the probability space

Now we have sufficient mathematical notions at our disposal to introduce a formal definition of a probability space which is the central one in Modern Probability Theory.

Definition 1.9. *An ordered triple $(\Omega, \mathcal{F}, \mathsf{P})$ is called a* ***probability space*** *if*

- *Ω is a sample space;*
- *$\mathcal{F}$ is a σ-algebra of measurable subsets (events) of Ω;*
- *P is a* ***probability measure*** *on $\mathcal{F}$, that is, P satisfies the following* ***Kolmogorov axioms*** *(Kolmogorov, 1933):*

1. *for any $A \in \mathcal{F}$ there exists a number*

$$\boxed{\mathsf{P}\{A\} \geq 0} \tag{1.21}$$

   called the ***probability of the event*** *A;*
2. *the probability measure is* ***normalized****, i.e.,*

$$\boxed{\mathsf{P}\{\Omega\} = 1} \tag{1.22}$$

3. *P is a* ***countable additive measure*** *satisfying (1.20), namely,*

$$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \sum_{i=1}^{\infty} \mathsf{P}\{A_i\}} \tag{1.23}$$

Departing from the Kolmogorov axioms only, one can derive the following properties of the probability measure.

Lemma 1.2. *Let $A, A_1, A_2, \ldots$ be measurable sets from $\mathcal{F}$. Then*

1.

$$\boxed{\mathsf{P}\{\varnothing\} = 0} \tag{1.24}$$

2.

$$\boxed{\mathsf{P}\{A_1 \cup A_2\} = \mathsf{P}\{A_1\} + \mathsf{P}\{A_2\} - \mathsf{P}\{A_1 \cap A_2\}} \tag{1.25}$$

3. *if* $A \subseteq B \in \mathcal{F}$ *then*

$$\boxed{\mathsf{P}\{A\} \le \mathsf{P}\{B\}} \tag{1.26}$$

4.

$$\boxed{\mathsf{P}\{\bar{A}\} = 1 - \mathsf{P}\{A\}} \tag{1.27}$$

5.

$$\boxed{\mathsf{P}\{A_1 \cup A_2\} \le \mathsf{P}\{A_1\} + \mathsf{P}\{A_2\}} \tag{1.28}$$

6.

$$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} \le \sum_{i=1}^{\infty} \mathsf{P}\{A_i\}} \tag{1.29}$$

(here A_i, A_j *are not obligatory disjoint).*

Proof. To clarify the proof of this lemma let us give *'a physical interpretation'* of $\mathsf{P}\{A\}$ as *'a square'* (in two dimensional case) and as *'a volume'* (in general case) of a corresponding set A. Then the properties 1–5 becomes evident. To prove 6, define new sets B_i by

$$B_1 \triangleq \bar{\varnothing} \cap A_1, \quad B_2 \triangleq \bar{A}_1 \cap A_2, \quad \ldots,$$
$$B_n \triangleq \bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_{n-1} \cap A_n \quad (n \ge 2)$$

Then the following properties seem to be evident:

-
  $$B_i \cap B_j = \varnothing \qquad i \ne j$$

  Indeed (for example, for $i < j$)

$$\begin{aligned} B_i \cap B_j &= \bigcap_{s=1}^{i-1} \bar{A}_s \cap A_i \bigcap_{l=1}^{j-1} \bar{A}_l \cap A_j \\ &= \bigcap_{s=1}^{i-1} \bar{A}_s \cap \left(\underbrace{A_i \cap \bar{A}_i}_{\varnothing} \bigcap_{l \ne i}^{j-1} \bar{A}_l \right) \cap A_j = \varnothing \end{aligned}$$

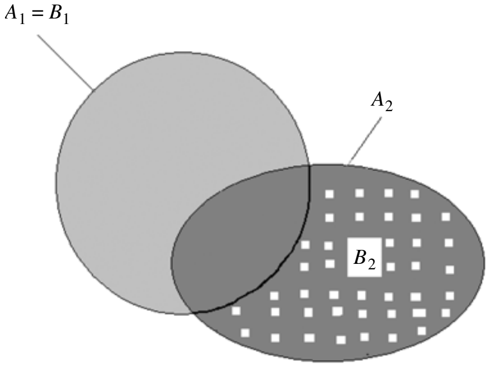

Fig. 1.3. Illustration of the fact: $A_1 \cup A_2 = B_1 \cup B_2$.

- $$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$$

  since

  $$A_1 \cup A_2 = A_1 \cup \left(\bar{A}_1 \cap A_2\right) = B_1 \cup B_2$$

  (see Fig. 1.3).
  Then, by the induction, we derive

  $$\bigcup_{i=1}^{n+1} B_i = \left(\bigcup_{i=1}^{n} B_i\right) \cup B_{n+1} = \bigcup_{i=1}^{n} A_i \cup \left[\bigcap_{i=1}^{n} \bar{A}_i \cap A_{n+1}\right]$$
  $$= \bigcup_{i=1}^{n} A_i \cup \left[\overline{\bigcup_{i=1}^{n} \bar{A}_i} \cap A_{n+1}\right] = \bigcup_{i=1}^{n} A_i \cup A_{n+1} = \bigcup_{i=1}^{n+1} A_i$$

  $$B_i \subseteq A_i$$

Taking into account these properties we derive

$$\mathsf{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \mathsf{P}\left\{\bigcup_{i=1}^{\infty} B_i\right\} = \sum_{i=1}^{\infty} \mathsf{P}\{B_i\} \leq \sum_{i=1}^{\infty} \mathsf{P}\{A_i\}$$

The lemma is proven. □

Based on Lemma 1.2 it is not difficult to prove its following extension.

Proposition 1.1. (the Benferroni inequalities)

- $$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} \leq \sum_{i=1}^{n} \mathsf{P}\{A_i\}} \tag{1.30}$$

- $$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} \geq \sum_{i=1}^{n} \mathsf{P}\{A_i\} - \sum_{1\le i<j\le n} \mathsf{P}\{A_i \cap A_i\}} \tag{1.31}$$

- $$\boxed{\begin{aligned}\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} \leq \sum_{i=1}^{n} \mathsf{P}\{A_i\} - \sum_{1\le i<j\le n} \mathsf{P}\{A_i \cap A_j\} \\ + \sum_{1\le i<j<k\le n} \mathsf{P}\{A_i \cap A_j \cap A_k\}\end{aligned}} \tag{1.32}$$

The last two inequalities (1.31) and (1.32) lead to the following relation.

Proposition 1.2. (the inclusion–exclusion formula)

$$\boxed{\begin{aligned}\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} = \sum_{i=1}^{n} \mathsf{P}\{A_i\} - \sum_{1\le i<j\le n} \mathsf{P}\{A_i \cap A_j\} \\ + \sum_{1\le i<j<k\le n} \mathsf{P}\{A_i \cap A_j \cap A_k\} + \ldots + \\ (-1)^{n+1}\,\mathsf{P}\{A_1 \cap A_1 \cap \cdots \cap\ A_n\}\end{aligned}} \tag{1.33}$$

Proof. Let us apply the induction. For $n = 2$ the relation is true in view of (1.25). Supposing that (1.33) holds for $(n-1)$, denoting $B := \bigcup_{i=1}^{n-1} A_i$ and using (1.25) we get

$$\begin{aligned}
\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} &= \mathsf{P}\{B \cup A_n\} = \mathsf{P}\{B\} + \mathsf{P}\{A_n\} - \mathsf{P}\{B\}\mathsf{P}\{A_n\} = \mathsf{P}\left\{\bigcup_{i=1}^{n-1} A_i\right\} \\
&+ \mathsf{P}\{A_n\} - \mathsf{P}\left\{\bigcup_{i=1}^{n-1} A_i\right\}\mathsf{P}\{A_n\} = \mathsf{P}\left\{\bigcup_{i=1}^{n-1} A_i\right\}(1 - \mathsf{P}\{A_n\}) + \mathsf{P}\{A_n\} \\
&\times \Bigg(\sum_{i=1}^{n-1} \mathsf{P}\{A_i\} - \sum_{1\le i<j\le n-1} \mathsf{P}\{A_i \cap A_j\} + \sum_{1\le i<j<k\le n-1} \mathsf{P}\{A_i \cap A_j \cap A_k\} \\
&+ \ldots + (-1)^n\,\mathsf{P}\{A_1 \cap A_1 \cap \cdots \cap\ A_{n-1}\}\Big)(1 - \mathsf{P}\{A_n\}) + \mathsf{P}\{A_n\}
\end{aligned}$$

which after the arrangements of the corresponding terms leads to (1.33). Lemma proven.

□

Theorem 1.1. *Let A and $\{A_n, n \geq 1\}$ be subsets from Ω such that $A_n \nearrow A$ or $A_n \searrow A$ as $n \to \infty$, and sets A_* and A^* are defined by (1.13). Then*

1. $\mathsf{P}\{A_n\}$ *monotonically converges to* $\mathsf{P}\{A\}$*, namely,*
- $A_n \nearrow A$ implies

$$\boxed{\mathsf{P}\{A_n\} \nearrow \mathsf{P}\{A\}} \tag{1.34}$$

- $A_n \searrow A$ implies

$$\boxed{\mathsf{P}\{A_n\} \searrow \mathsf{P}\{A\}} \tag{1.35}$$

2.

$$\boxed{\mathsf{P}\{A_*\} \leq \liminf_{n\to\infty} \mathsf{P}\{A_n\} \leq \limsup_{n\to\infty} \mathsf{P}\{A_n\} \leq \mathsf{P}\{A^*\}} \tag{1.36}$$

3. $A_n \to \varnothing$ as $n \to \infty$ implies

$$\boxed{\mathsf{P}\{A_n\} \to 0} \tag{1.37}$$

Proof.
1. Define $B_1 := A_1$ and $B_n := A_n \cap \bar{A}_{n-1}$ for $n \geq 2$. Then by the construction $\{B_n, n \geq 1\}$ are disjoint sets, and

$$A_n = \bigcup_{k=1}^{n} B_k, \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} B_k$$

So, by Lemma 1.2 it follows

$$\begin{aligned} \mathsf{P}\{A_n\} &= \sum_{k=1}^{n} \mathsf{P}\{A_k\} \nearrow \sum_{k=1}^{\infty} \mathsf{P}\{A_k\} \\ &= \mathsf{P}\left\{\bigcup_{k=1}^{\infty} B_k\right\} = \mathsf{P}\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \mathsf{P}\{A\} \end{aligned}$$

The case $A_n \searrow A$ follows similarly by considering complements $\bar{A}_n \nearrow \bar{A}$.
2. It follows directly from the statement 1 of this theorem if we take into account that

$$A_* \nwarrow \bigcap_{m=n}^{\infty} A_m \subset A_n \subset \bigcup_{m=n}^{\infty} A_m \searrow A^*$$

3. It results from 1–2 and (1.24). □

1.2.3 Sets of a null measure and completeness

Here we will introduce the notion of a *null set* or a set of a null measure which we will use below.

Definition 1.10. *A set A is set to be a* ***null set*** *if there exists $B \in \mathcal{F}$ such that*

$$\boxed{A \subset B \text{ and } \mathsf{P}\{B\} = 0} \tag{1.38}$$

that is, a set is a null set if it is contained in a measurable set of zero-probability.

In fact, the set A in this definition is not obligatory measurable, namely, it is not obligatory that $A \in \mathcal{F}$. The concept of *completeness* takes care of that problem.

Definition 1.11. *A probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is said to be* ***complete*** *if every null set is measurable, that is, the property $A \subset B \in \mathcal{F}$ with $\mathsf{P}\{B\} = 0$ implies $A \in \mathcal{F}$ (sure with $\mathsf{P}\{A\} = 0$).*

Remark 1.3. *It seems to be evident that it is always possible to extend a given σ-algebra adding within null sets making, thus, the corresponding probability space complete. To avoid any distraction from the main line, we will assume from now on (without further explicit mentioning) that all probability spaces are complete. This property turns out to be very important for the correct definition of stochastic integration in the theory of stochastic processes.*

1.3 Borel algebra and probability measures

Now we consider some examples of measurable space $(\Omega, \mathcal{F})$ which are extremely important in *Probability* and *Stochastic Processes* theories.

1.3.1 The measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

1.3.1.1 Borel-measurable space defined on the real line

Let $\mathbb{R} = (-\infty, \infty)$ be the real line and

$$(a, b] \stackrel{\Delta}{=} \{x \in \mathbb{R} : a < x \leq b\}$$

be the interval *'open from left'* and *'closed from right'*.

Definition 1.12. *If $\Omega = \mathbb{R}$ and contains also single tons $\{a\}$ $(a \in \mathbb{R})$ then the smallest σ-algebra $\mathcal{F}$ of subsets from $\mathbb{R}$ consisting of all intervals*

$$\boxed{\begin{array}{c}(a, b),\ [a, b],\ [a, b),\ (a, b],\ (-\infty, b],\ (-\infty, b),\\ [a, \infty),\ (a, \infty),\ \{a, b : -\infty < a < b < \infty\}\end{array}} \tag{1.39}$$

is called ***Borel σ-algebra*** *and denoted by*

$$\boxed{\mathcal{B} := \mathcal{B}(\mathbb{R})} \tag{1.40}$$

Remark 1.4. *A σ-algebra that contains all intervals of a given type contains also all intervals of any other type (as in (1.39)). Indeed, define $\mathcal{B}(\mathbb{R})$ as a smallest σ-field containing all open sets (a, b) of $\mathbb{R}$. Since a set is open if and only if its complement is closed, $\mathcal{B}(\mathbb{R})$ is the smallest σ-field containing also all closed sets of $\mathbb{R}$. Similarly, if $\mathcal{B}_0$ is the field of finite disjoint units of right semi-closed intervals $(a, b]$, then $\mathcal{B}(\mathbb{R})$ is the smallest σ-field containing the sets from $\mathcal{B}_0$. So, $\mathcal{B}(\mathbb{R})$ may be considered as the smallest σ-field that contains the class of all intervals of $\mathbb{R}$.*

Now let's introduce the probability measure on Borel subsets of a Borel σ-algebra. Consider the event

$$A := (-\infty, x], \quad x \in \mathbb{R}$$

and define the function $F(x)$ in the following way:

$$\boxed{F(x) := \mathsf{P}\{A\} = \mathsf{P}\{(-\infty, x]\}} \tag{1.41}$$

where $\mathsf{P}\{A\}$ is a *probability measure*, defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Lemma 1.3. *The function $F(x)$ (1.41) possesses the following properties:*

1. *$F(x)$ is a* ***non-decreasing*** *function;*
2.

$$\boxed{\begin{array}{c} F(-\infty) = 0, \quad F(+\infty) = 1 \\ F(-\infty) := \lim\limits_{x \downarrow -\infty} F(x), \quad F(+\infty) := \lim\limits_{x \uparrow +\infty} F(x) \end{array}} \tag{1.42}$$

3. *$F(x)$ is* ***continuous on the right and has a limit on the left*** *at each point $x \in \mathbb{R}$.*
4. *$F(x)$ may have at most a* ***countable number of jumps****.*

Proof. The property 1 follows from the property 3 of Lemma 1.2. Since $(-\infty, -\infty] = \varnothing$, then by the property 1 of Lemma 1.2, it follows that

$$F(-\infty) = \mathsf{P}\{(-\infty, -\infty]\} = \mathsf{P}\{\varnothing\} = 0$$

Again, in view of the fact that

$$F(+\infty) = \mathsf{P}\{(-\infty, +\infty]\} = \mathsf{P}\{\Omega\} = 1$$

we prove the property 2. The last statement 3 results directly from the properties of probability measures (Lemma 1.2). To prove 4 for $n \geq 1$ denote

$$\mathbb{J}_F^{(n)} := \left\{ x : F \text{ has a jump at } x \text{ of magnitude form } \left(\frac{1}{n+1}, \frac{1}{n} \right) \right\}$$

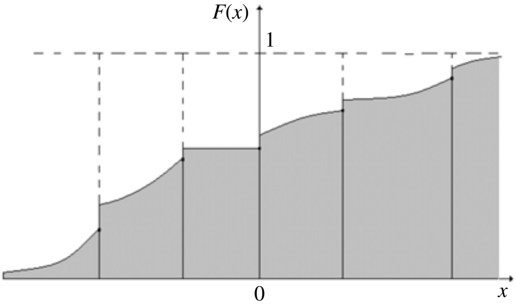

Fig. 1.4. A distribution function $F(x)$.

The total number of points in $\mathbb{J}_F^{(n)}$ is at most equal to $(n+1)$ since F is non-decreasing and does not exceed 1. The conclusion 4 then follows from the fact that the set $\mathbb{J}_F$ of all discontinuities of F is $\mathbb{J}_F = \bigcup_{n=1}^{\infty} \mathbb{J}_F^{(n)}$. Lemma is proven. □

Definition 1.13. *A function $F = F(x)$ defined by (1.41) and, hence, satisfying the conditions 1–4 of Lemma 1.3 is called* ***a distribution function*** *(given on the real line $\mathbb{R}$).*

A typical behavior of a distribution function $F(x)$ is depicted in Fig. 1.4.

The next lemma states a correspondence *('isomorphism')* between a class of distribution functions and a class of probability measures.

Lemma 1.4. *Any probability measure P (a* ***Borel measure****) corresponds (by (1.41)) to some distribution function $F(x)$, and the converse is also true, i.e., for each distribution function $F(x)$ there exists a unique probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

$$\boxed{\mathsf{P}\{(a,b]\} := F(b) - F(a) \qquad \forall a, b : -\infty \leq a < b \leq \infty} \tag{1.43}$$

Proof. The first statement follows from Definition (1.41). It is easy to check that $\mathsf{P}\{(a,b]\}$, defined by (1.43), possesses all properties of a probability measure (see Lemma 1.3). The formulae (1.43) defines it uniquely. □

1.3.1.2 Examples of Borel measures

Consider in detail different probability measures the construction of which is based on (1.41).

Discrete measures are the special Borel measures P for which the corresponding distribution functions $F = F(x)$ are *piecewise constant with 'jumps'* at the points x_i $(i = 1, 2, \ldots)$ (see Fig. 1.5).

Here, as well as for any distribution function, the following properties hold:

$$\boxed{\begin{array}{c} \Delta F(x_i) := F(x_i) - F(x_i - 0) \geq 0 \quad (i = 1, 2, \ldots) \\ \sum_i \Delta F(x_i) = 1 \end{array}} \tag{1.44}$$

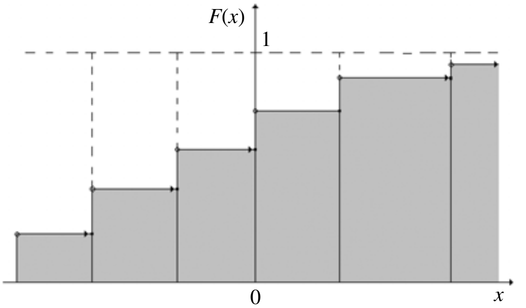

Fig. 1.5. A distribution function $F(x)$ corresponding to a discrete measure.

Table 1.1
Examples of some discrete distributions.

Distribution	Probabilities	Parameters
Discrete uniform	$p_i = 1/n, \ (i = \overline{1, n})$	n
Bernoulli	$p_1 = p, \ p_0 = q$	$p = 1 - q \in [0, 1]$
Binomial	$\mathsf{P}\left\{\begin{array}{l}\text{an event occurs}\\ i\text{-time in } n \text{ trials}\end{array}\right\} = C_n^i p^i q^{n-i} \ (i = 1, \ldots, n)$	$p = 1 - q \in [0, 1]$ $C_n^i := \dfrac{n!}{i!\,(n-i)!}$
Poisson	$\mathsf{P}\{i = \lambda\} = e^{-\lambda} \dfrac{(\lambda)^i}{i!}$	$\lambda > 0$

Definition 1.14. *The set of non-negative numbers* $(p_1, p_2, \ldots)$ *satisfying*

$$\boxed{\begin{array}{c} p_i := \mathsf{P}\{(x_i - 0, x_i)\} = \Delta F(x_i) \geq 0 \\ \sum_i p_i = 1 \end{array}} \tag{1.45}$$

is said to be ***a discrete probability distribution*** *and the corresponding distribution function* $F = F(x)$ *is called* ***a discrete distribution function****.*

Table 1.1 represents some types of discrete probability distributions commonly used in Probability Theory.

The following asymptotic relation between the binomial and Poisson distributions takes place.

Lemma 1.5. (Poisson, 1781–1840) *If*

$$\boxed{n \to \infty, \ p \to 0, \ np \to \lambda} \tag{1.46}$$

then

$$\boxed{\frac{n!}{i!\,(n-i)!}p^i\,(1-p)^{n-i}\to e^{-\lambda}\frac{(\lambda)^i}{i!}} \tag{1.47}$$

Proof. For $i=0$ and $np\to\lambda$ one has

$$\begin{aligned} P_n\,(i)\,|_{i=0} &:= \frac{n!}{i!\,(n-i)!}p^i\,(1-p)^{n-i}\,|_{i=0} \\ &= (1-p)^n = \left(1-\frac{\lambda+o\,(1)}{n}\right)^n \to e^{-\lambda} := P_\infty\,(0) \end{aligned}$$

For any fixed $i>0$ and $np\to\lambda$ we have

$$\begin{aligned} \frac{P_n\,(i)}{P_n\,(i-1)} &= \frac{C_n^i p^i q^{n-i}}{C_n^{i-1}p^{i-1}q^{n-i+1}} = \frac{p}{i\,(n-i)\,(1-p)} \\ &= \frac{[\lambda+o\,(1)]\,/n}{i\,(n-i)\,(1-[\lambda+o\,(1)]\,/n)} \to \frac{\lambda}{i} \end{aligned}$$

which implies

$$\begin{aligned} P_\infty\,(i) &= P_\infty\,(i-1)\,\frac{\lambda}{i} \\ &= P_\infty\,(i-2)\,\frac{\lambda^2}{i\,(i-1)} = \ldots = P_\infty\,(0)\,\frac{\lambda^i}{i!} = e^{-\lambda}\frac{\lambda^i}{i!} \end{aligned}$$

The lemma is proven. □

Absolutely continuous measures are those distribution functions which could be presented as an integral of some other positive function, that is, there exists a non-negative Lebesgue integrable function $f(u)$ such that

$$\boxed{F(x)=\int_{-\infty}^{x} f(u)du, \quad f(u)\geq 0 \text{ for all } u\in\mathbb{R}} \tag{1.48}$$

Definition 1.15. *The function $f=f(u)$ in (1.48) is called* ***the density function*** *of the distribution function $F=F(x)$, and in this case $F=F(x)$ is called* ***an absolutely continuous distribution function****.*

Two main properties of absolutely continuous distribution functions are

- an absolutely continuous distribution functions is differentiable for any $x\in\mathbb{R}$ such that

$$\boxed{\frac{d}{dx}F(x)=f(x)\geq 0} \tag{1.49}$$

Table 1.2
Examples of some absolutely continuous distribution functions.

Distribution	Density $f(x)$	Parameters
Uniform on $[a, b]$	$\frac{1}{b-a},\ x \in [a, b]$	$a, b \in \mathbb{R},\ a < b$
Normal (Gaussian)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}},\ x \in \mathbb{R}$	$a \in \mathbb{R},\ \sigma > 0$
Cauchy	$\frac{a}{\pi\left(x^2 + a^2\right)},\ x \in \mathbb{R}$	$a > 0$
Gamma	$\frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}},\ x \geq 0$	$\alpha > 0,\ \beta > 0$

- a density function is normalized by

$$\boxed{\int_{-\infty}^{\infty} f(x)dx = 1} \tag{1.50}$$

The *'normalization'* property (1.50) results from Lemma 1.3 taking into account that $F(+\infty) = 1$.

Table 1.2 represents some typical examples of density functions $f = f(x)$.

Continuous singular measures are measures with *continuous distribution functions* $F(x)$ *which are differentiable and having the derivative* $F'(x) = 0$ *almost everywhere* (except points of increasing of zero Lebesgue measure).

Consider now an example of a singular measure.

The continuous function $F_1(x)$ (see Fig. 1.6) is a partially linear non-decreasing function satisfying

$$F_1(0) = 0, \quad F_1(1/3) = 1/2, \quad F_1(2/3) = 1/2, \quad F_1(1) = 1$$

Next, define the function $F_2(x)$ by

$$F_1(0) = 0, \quad F_1(1/9) = 1/4, \quad F_1(2/9) = 1/4, \quad F_1(1/3) = 1/2$$
$$F_1(2/3) = 1/2, \quad F_1(7/9) = 3/4, \quad F_1(8/9) = 3/4, \quad F_1(1) = 1$$

Continuing this process we construct the sequence $\{F_n(x)\}$ of the continuous functions which converges to a non-decreasing continuous function $F(x)$, known as the *Cantor measure*:

$$F(x) := \lim_{n \to \infty} F_n(x) \qquad \forall x \in \mathbb{R}$$

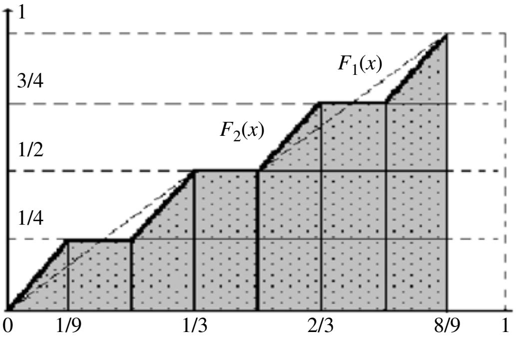

Fig. 1.6. An example of singular distribution function $F(x)$ corresponding to a singular continuous measure.

This function practically always is constant (excluding points of zero measure). Indeed, the total length of the intervals

$$\left(\frac{1}{3}, \frac{2}{3}\right), \left(\frac{1}{9}, \frac{2}{9}\right), \left(\frac{7}{9}, \frac{8}{9}\right), \ldots$$

where this function is constant, is equal:

$$\begin{aligned}\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots &= \frac{1}{3}\left(1 + \frac{2}{3} + \frac{4}{9} + \cdots\right) \\ &= \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n = \frac{1}{3}\left(\frac{1}{1-2/3}\right) = 1\end{aligned}$$

Hence, 'the length of increasing intervals' is equal to 0, or, more exactly, the Lebesgue measure of increasing intervals is equal to zero.

Remark 1.5. *Obviously, discrete distributions (probabilistic measures) are singular but not continuous.*

In fact, these three types of distribution function form any distribution function as their linear combination.

Theorem 1.2. *Any distribution function can be decomposed into a convex combination of three pure types: a* ***discrete*** *one, an* ***absolutely continuous*** *one, and a* ***continuous singular*** *one, that is, if $F(x)$ is a distribution function, then for any $x \in \mathbb{R}$*

$$\boxed{\begin{gathered} F(x) = \lambda_1 F_{discrete}(x) + \lambda_2 F_{abs.cont}(x) + \lambda_3 F_{cont.sing} \\ \lambda_1 + \lambda_2 + \lambda_3 = 1, \quad \lambda_i \geq 0 \quad (i = 1, 2, 3) \end{gathered}} \tag{1.51}$$

where

(a)

$$\boxed{F_{abs.cont}(x) = \int_{-\infty}^{x} f(u)du,\ f(x) = F'_{abs.cont}(x)} \tag{1.52}$$

(b) $F_{discrete}(x)$ *is a pure jump function with at most a countable number of jumps;*
(c) $F_{cont.sing}(x)$ *is continuous and* $F'_{cont.sing}(x) = 0$ *almost everywhere* $(a.e.)$.

Such decomposition is ***unique***.

To prove this theorem we need the following two preliminary results: the first one shows that any distribution function can be split into absolutely continuous components and a singular one; the second one provides the decomposition of a distribution function into a discrete component and a continuous one.

Theorem 1.3. (Lebesgue decomposition theorem) *Every distribution function* $F(x)$ *can be decomposed into a convex combination of an absolutely continuous distribution function* $F_{abs.cont}(x)$ *and a singular one* $F_{sing}(x)$ *(including discrete and continuous singular components), that is, for all* $x \in \mathbb{R}$ *there exists* $\alpha \in [0, 1]$ *such that*

$$\boxed{F(x) = \alpha F_{abs.cont}(x) + (1 - \alpha) F_{sing}(x)} \tag{1.53}$$

Proof. If $F(x)$ is a distribution function, and hence, it is a monotonically non-decreasing function, then, by the Lebesgue theorem, it is differentiable almost everywhere and equates to $F^*_{abs.cont}(x) = \int_{-\infty}^{x} F'(s)\,ds$. Evidently that $F^*_{abs.cont}(x)$ is non-decreasing continuous and

$$F^*_{abs.cont}(-\infty) = 0, \quad F^*_{abs.cont}(+\infty) \leq 1$$

Define then $F^*_{sing}(x) := F(x) - F^*_{abs.cont}(x)$ which is also non-decreasing with $F^*_{sing}(-\infty) = F_{cont.sing}(x) = 0$ and $F^*_{sing}(+\infty) \leq 1$. Putting then

$$F_{abs.cont}(x) := \frac{F^*_{abs.cont}(x)}{F^*_{abs.cont}(+\infty)}, \quad F_{sing}(x) := \frac{F^*_{sing}(x)}{F^*_{sing}(+\infty)}$$

we complete the proof. □

Theorem 1.4. *Every distribution function* $F(x)$ *can be represented as a convex combination of discrete distribution function* $F_{discrete}(x)$ *and a continuous one* $F_{cont}(x)$*, that is, for all* $x \in \mathbb{R}$ *there exists* $\beta \in [0, 1]$ *such that*

$$\boxed{F(x) = \beta F_{discrete}(x) + (1 - \beta) F_{cont}(x)} \tag{1.54}$$

Proof. By the property 4 in Lemma 1.3 it follows that $F(x)$ may have at most a countable number of jumps. Let $\{x_j\}$ be those *jumps* (if they exist). Denote the values of the jumps by

$$p_j := F\left(x_j + 0\right) - F\left(x_j - 0\right)$$

and define *the sum of all jumps to the left of* x as

$$F^*_{discrete}(x) := \sum_{x_j \le x} p_j, \quad x \in \mathbb{R}$$

Then define

$$F^*_{cont}(x) := F(x) - F^*_{discrete}(x)$$

By the construction both functions $F^*_{discrete}(x)$ and $F^*_{cont}(x)$ are non-negative, non-decreasing and satisfying

$$\lim_{x\to-\infty} F^*_{discrete}(x) = \lim_{x\to-\infty} F^*_{cont}(x) = 0$$
$$\lim_{x\to\infty} F^*_{discrete}(x) \le 1, \quad \lim_{x\to\infty} F^*_{cont}(x) \le 1$$

Notice also that $F^*_{cont}(x)$ is right-continuous (since $F(x)$ is right-continuous). But $F^*_{cont}(x)$ is also left-continuous. Indeed,

$$F^*_{cont}(x) - F^*_{cont}(x-0) = F(x) - F^*_{discrete}(x)$$
$$-\left[F(x-0) - F^*_{discrete}(x-0)\right] = \begin{cases} p_j - p_j & \text{when } x = x_j \text{ for some } j \\ 0 & \text{otherwise} \end{cases}$$

which proves the left-continuity. Therefore, $F^*_{cont}(x)$ is continuous. Using then the normalization factor, namely, putting

$$F_{discrete}(x) = \frac{F^*_{discrete}(x)}{F^*_{discrete}(+\infty)}, \quad F_{cont}(x) = \frac{F^*_{cont}(x)}{F^*_{cont}(+\infty)}$$

we finish the proof. □

Proof. **(Theorem 1.2)** By Theorems 1.3 and 1.4 we know that

$$F(x) = \alpha F_{abs.cont}(x) + (1-\alpha) F_{sing}(x)$$
$$F_{sing}(x) = \beta F_{discrete}(x) + (1-\beta) F_{cont.sing}(x)$$

that implies

$$F(x) = \alpha F_{abs.cont}(x) + (1-\alpha)\left[\beta F_{discrete}(x) + (1-\beta) F_{cont.sing}(x)\right]$$
$$= \alpha F_{abs.cont}(x) + (1-\alpha)\beta F_{discrete}(x) + (1-\alpha)(1-\beta) F_{cont.sing}(x)$$

Taking $\lambda_1 := \alpha$, $\lambda_2 := (1-\alpha)\beta$ and $\lambda_3 := (1-\alpha)(1-\beta)$ we have

$$\lambda_1 + \lambda_2 + \lambda_3 = \alpha + (1-\alpha)\beta + (1-\alpha)(1-\beta) = 1$$

The uniqueness can be proven by contradiction. The theorem is proven. □

1.3.2 The measurable space $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$

This section deals with a simple generalization of the previous single dimensional case ($N = 1$). In fact, for Borel sets in higher finite dimensions the definitions below extend ones to higher dimensional rectangles.

Theorem 1.5. *An N-**dimensional distribution function** $F : \mathbb{R}^N \to \mathbb{R}$ is a function $F = F(x) := F(x_1, x_2, \ldots, x_N)$ with the following properties:*

1. *for any $a, b \in \mathbb{R}^N \quad (b_i \geq a_i, i = 1, \ldots, N)$*

$$\boxed{\begin{aligned}\Delta_{a,b}F(x) := F(b_1, b_2, \ldots, b_N) - F(a_1, a_2, \ldots, a_N)\\ = F(b) - F(a) \geq 0\end{aligned}}$$

2. *$F(x)$ is a continuous function on right, i.e.,*

$$\boxed{F(x^{(k)}) \downarrow F(x) \quad if \quad x^{(k)} \downarrow x}$$

and has a bounded limit on left;

3.

$$\boxed{F(+\infty, +\infty, \ldots, +\infty) = 1}$$

and

$$\boxed{\lim_{x \downarrow y} F(x) = 0}$$

if at least one of coordinate y_i of a vector $y \in R^N$ is equal to $(-\infty)$.

The complete analog of Lemma 1.4 also takes place.

Lemma 1.6. *For each distribution function $F = F(x), x \in \mathbb{R}^N$ there exists a unique probability measure P on $\left(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N)\right)$ such that for any $a, b : -\infty \leq a_i < b_i \leq \infty, i = 1, \ldots, N$*

$$\boxed{\mathsf{P}\{(a, b]\} = \Delta_{a,b}F(x) = F(b) - F(a)} \tag{1.55}$$

and the converse is also true.

Proof. It completely repeats the proof of Lemma 1.4, and therefore, it is omitted. □

Analogously to the single dimensional case, there exist *discrete, absolutely continuous* and *singular* measures of Borel cylinders in $\mathbb{R}^N$. So, for an absolutely continuous measure there exists a function $f : \mathbb{R}^N \to \mathbb{R}$ such that the following presentation holds:

$$\boxed{F(x) = \int_{u_1=-\infty}^{x_1} \cdots \int_{u_n=-\infty}^{x_N} f(u_1, \ldots, u_N) du_1 \cdots du_N} \tag{1.56}$$

where a corresponding *function of a distribution density* $f(x)$ satisfies the *'normalization conditions'*:

$$\boxed{\begin{array}{c} f(x_1, \ldots, x_N) \geq 0 \\ \int\limits_{u_1=-\infty}^{\infty} \cdots \int\limits_{u_n=-\infty}^{\infty} f(u_1, \ldots, u_N) du_1 \cdots du_N = 1 \end{array}} \tag{1.57}$$

The next example represents one of the distribution functions most commonly used in theoretical considerations.

Example 1.2. (*N*-dimensional Gaussian distribution)

$$\boxed{f(x) = \frac{|\det R|^{-1/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}(x-a)^T R^{-1}(x-a)\right\}} \tag{1.58}$$

where $R = R^T > 0, a \in \mathbb{R}^N$.

1.3.3 The measurable space $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$

This type of measurable space is a generalization of finite dimensional spaces $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ up to infinite dimensional ones.

Let T be a set of indices $t \in T$ and $\mathbb{R}_t$ is a real line in $\mathbb{R}$ associated with the index t. One can correctly define the finite dimensional space $(\mathbb{R}^\tau, \mathcal{B}(\mathbb{R}^\tau))$ where $\tau := [t_1, t_2, \ldots, t_n]$. Let P_τ be a probability measure on $(\mathbb{R}^\tau, \mathcal{B}(\mathbb{R}^\tau))$.

Definition 1.16. *We say that* ***the family*** P_τ ***of probability measures*** *(where* τ *runs through all finite unordered sets)* ***is consistent*** *if for all collections* $\tau = [t_1, t_2, \ldots, t_n]$ *and* $\sigma = [s_1, s_2, \ldots, s_k]$ *such that for* $\sigma \subseteq \tau$ *the following property holds:*

$$\mathsf{P}_\sigma \left\{x_s : (x_{s_1}, \ldots, x_{s_k}) \in B\right\} = \mathsf{P}_\tau \left\{(x_{t_1}, \ldots, x_{t_n}) : (x_{s_1}, \ldots, x_{s_k}) \in B\right\}$$

for each $B \in \mathcal{B}(\mathbb{R}^\sigma)$.

The following result known as Kolmogorov's theorem on the extension of measures on $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T)\right)$ takes place.

Theorem 1.6. (The Kolmogorov theorem) *Let* $\{\mathsf{P}_\tau\}$ *be a consistent family of probability measures on* $(\mathbb{R}^\tau, \mathcal{B}(\mathbb{R}^\tau))$. *Then there exists a* ***unique probability measure*** P *on* $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T)\right)$ *such that*

$$\boxed{\mathsf{P}\{\mathcal{J}_\tau(B)\} = \mathsf{P}_\tau\{B\}} \tag{1.59}$$

for all unordered sets $\tau = [t_1, t_2, \ldots, t_n]$ *of different indices* $t_i \in T$, $B \in \mathcal{B}(R^T)$, *where*

$$\mathcal{J}_\tau(B) \triangleq \left\{x \in \mathbb{R}^T : (x_{t_1}, \ldots, x_{t_n}) \in B\right\}$$

Proof. For details of this proof see Shiryayev (1984) (Theorem 2.3.4). □

Remark. In other words, a measure P on $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T)\right)$ is defined correctly if there are defined any *'finite dimensional'* measure P_τ for all sets $\tau = [t_1, t_2, \ldots, t_n]$.

1.3.4 Wiener measure on $(\mathbb{R}^{[0,\infty]}, \mathcal{B}(\mathbb{R}^{[0,\infty]}))$

This subsection deals with the most commonly used example of '*infinite dimensional measures*' defined in the previous subsection for partial case of the set T. Let $T := [0, \infty)$. Then $\mathbb{R}^T$ is a space of all real function $x = x(t)$, $t \geq 0$. Consider now the following family $\{\phi_t(y|x)\}_{t \geq 0}$ of all Gaussian (normal) densities (as the function of y for any fixed x):

$$\boxed{\phi_t(y|x) := \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{(y-x)^2}{2t}\right\}, \quad y, x \in \mathbb{R}, t > 0} \tag{1.60}$$

For each $\tau = [t_1, t_2, \ldots, t_n]$, $(t_1 < t_2 < \cdots < t_n)$ and each set $B = I_1 \times \cdots \times I_n$, $I_k := (a_k, b_k]$ let us construct the measure $\mathsf{P}_\tau\{B\}$ according to the following definition:

$$\boxed{\begin{array}{c} \mathsf{P}_\tau\{I_1 \times \cdots \times I_n\} \\ := \int\limits_{I_1} \cdots \int\limits_{I_n} \phi_{t_1}(y_1|0)\phi_{t_2-t_1}(y_2|y_1) \cdots \phi_{t_n-t_{n-1}}(y_n|y_{n-1}) dy_1 \cdots dy_n \end{array}} \tag{1.61}$$

Here the integration in (1.61) is in the usual Riemann sense.

If we will interpret $\phi_{t_k - t_{k-1}}(a_k|a_{k-1})$ as the probability that '*a particle*', starting at the point a_{k-1}, at the time interval $t_k - t_{k-1}$ arrives to a neighborhood of the point a_k, then the product of densities describes a certain independence of the increments of the displacement of '*the moving particles*' in time interval $(0, t_1]$, $(t_1, t_2]$, ..., $(t_{n-1}, t_n]$. The family of such measures P_τ (constructed by this way) is easily seen to be a consistent one, and hence, can be extended up to the measurable space $\left(\mathbb{R}^{[0,\infty]}, \mathcal{B}(\mathbb{R}^{[0,\infty]})\right)$ by Theorem 1.6. The measure we obtain is known as the **Wiener measure**.

1.4 Independence and conditional probability

1.4.1 Independence

One of the central concepts of Probability Theory is *independence*. In different applications it means, for example, that

- some successive experiments do not influence each other;
- the future does not depend on the past;
- knowledge of the outcomes so far does not provide any information on future experiments.

Definition 1.17. *The events* $\{A_k, k = 1, \ldots, n\}$ *are said to be* ***independent*** *if and only if*

$$\boxed{\mathsf{P}\left\{\bigcap_k A_{i_k}\right\} = \prod_k \mathsf{P}\left\{A_{i_k}\right\}} \tag{1.62}$$

where intersections and products, respectively, are taken over all subsets of $\{1, \ldots, n\}$*. The* $\{A_n, n \geq 1\}$ *are independent if* $\{A_k, k = 1, \ldots, n\}$ *are independent for all* n*.*

Proposition 1.3. *If* $A \subset \Omega$ *and* $B \subset \Omega$ *are independent, then so are* A *and* B^c*,* A^c *and* B*,* A^c *and* B^c*.*

Proof. One has

$$\begin{aligned}\mathsf{P}\left\{A \cap B^c\right\} &= \mathsf{P}\{A\} - \mathsf{P}\{A \cap B\} \\ &= \mathsf{P}\{A\} - \mathsf{P}\{A\}\mathsf{P}\{B\} = \mathsf{P}\{A\}(1 - \mathsf{P}\{B\}) = \mathsf{P}\{A\}\mathsf{P}\left\{B^c\right\}\end{aligned}$$

Analogously,

$$\begin{aligned}\mathsf{P}\left\{A^c \cap B\right\} &= \mathsf{P}\{B\} - \mathsf{P}\{A \cap B\} \\ &= \mathsf{P}\{B\} - \mathsf{P}\{A\}\mathsf{P}\{B\} = \mathsf{P}\{B\}(1 - \mathsf{P}\{A\}) = \mathsf{P}\{B\}\mathsf{P}\left\{A^c\right\}\end{aligned}$$

and

$$\begin{aligned}\mathsf{P}\left\{A^c \cap B^c\right\} &= \mathsf{P}\left\{A^c\right\} - \mathsf{P}\left\{A^c \cap B\right\} \\ &= \mathsf{P}\left\{A^c\right\} - \mathsf{P}\left\{A^c\right\}\mathsf{P}\{B\} = \mathsf{P}\left\{A^c\right\}(1 - \mathsf{P}\{B\}) = \mathsf{P}\left\{A^c\right\}\mathsf{P}\left\{B^c\right\}\end{aligned}$$

which completes the proof. □

Lemma 1.7. *If* $\{A_k, k = 1, \ldots, n\}$ *is the collection of independent events, then*

•

$$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^n A_i\right\} = 1 - \prod_{i=1}^n (1 - \mathsf{P}\{A_i\})} \tag{1.63}$$

•

$$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^n A_i\right\} \geq 1 - \exp\left\{-\sum_{i=1}^n \mathsf{P}\{A_i\}\right\}} \tag{1.64}$$

Proof. It can be done by induction. For $n = 2$ in view of the relation

$$\mathsf{P}\{A_1\} + \mathsf{P}\{A_2\} = 1$$

It is evident (see (1.25)) that

$$\begin{aligned}\mathsf{P}\{A_1 \cup A_2\} &= \mathsf{P}\{A_1\} + \mathsf{P}\{A_2\} - \mathsf{P}\{A_1 \cap A_2\} \\ &= \mathsf{P}\{A_1\} + \mathsf{P}\{A_2\} - \mathsf{P}\{A_1\}\mathsf{P}\{A_2\} \\ &= 1 - (1 - \mathsf{P}\{A_2\})(1 - \mathsf{P}\{A_1\})\end{aligned}$$

Supposing that (1.63) holds for $(n-1)$ and denoting $B := \bigcup_{i=1}^{n-1} A_i$ we have

$$\bigcup_{i=1}^{n} A_i = B \cup A_n, \quad \mathsf{P}\{B\} + \mathsf{P}\{A_n\} = 1$$

and

$$\begin{aligned}\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} &= \mathsf{P}\{B\} + \mathsf{P}\{A_n\} - \mathsf{P}\{B \cap A_n\} \\ &= 1 - \mathsf{P}\{B\}\mathsf{P}\{A_n\} = 1 - (1 - \mathsf{P}\{A_n\})(1 - \mathsf{P}\{B\}) \\ &= 1 - (1 - \mathsf{P}\{A_n\}) \prod_{i=1}^{n-1} (1 - \mathsf{P}\{A_i\}) = 1 \prod_{i=1}^{n} (1 - \mathsf{P}\{A_i\})\end{aligned}$$

that proves the validity of (1.63) for any integer n. The inequality (1.64) follows from the inequality

$$1 - x \leq \exp\{-x\}$$

valid for any $x \in \mathbb{R}$ being applied to (1.63) since

$$\begin{aligned}\mathsf{P}\left\{\bigcup_{i=1}^{n} A_i\right\} &= 1 - \prod_{i=1}^{n} (1 - \mathsf{P}\{A_i\}) \\ &\geq 1 - \exp\{-\mathsf{P}\{A_n\}\} \prod_{i=1}^{n-1} (1 - \mathsf{P}\{A_i\}) \\ &\geq \cdots \geq 1 - \prod_{i=1}^{n} \exp\{-\mathsf{P}\{A_i\}\} = 1 - \exp\left\{-\sum_{i=1}^{n} \mathsf{P}\{A_i\}\right\}\end{aligned}$$

The lemma is proven. □

Corollary 1.3. *In the setup of Lemma 1.7 (when $\{A_k, k \geq 1\}$ is the collection of independent events) the property*

$$\boxed{\sum_{i=1}^{n} \mathsf{P}\{A_i\} = \infty} \tag{1.65}$$

implies that

$$\boxed{\mathsf{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = 1} \tag{1.66}$$

Proof. It results from (1.64) and the relations

$$1 \geq \mathsf{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} \geq 1 - \exp\left\{-\sum_{i=1}^{n} \mathsf{P}\{A_i\}\right\} = 1 \qquad \square$$

1.4.2 Pair-wise independence

This independence concept is slightly weaker that the independence of a collection (see Definition 1.17).

Definition 1.18. *The events $\{A_k, k = 1, \ldots, n\}$ are **pair-wise independent** if and only if*

$$\boxed{\mathsf{P}\left\{A_i \cap A_j\right\} = \mathsf{P}\{A_i\}\,\mathsf{P}\left\{A_j\right\}} \tag{1.67}$$

for all $i \neq j$ $(i, j = 1, \ldots, n)$.

Proposition 1.4. *Pair-wise independence of events does not imply their independence.*

Proof. It is sufficient to consider the following counter-example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with all outcomes ω_i which are equiprobable. It is easy to verify that the events

$$A := \{\omega_1, \omega_2\}, \quad B := \{\omega_1, \omega_3\}, \quad C := \{\omega_1, \omega_4\}$$

are pair-wise independent, whereas

$$\mathsf{P}\{A \cap B \cap C\} = \frac{1}{4} \neq \mathsf{P}\{A\}\,\mathsf{P}\{B\}\,\mathsf{P}\{C\} = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \qquad \square$$

1.4.3 Definition of conditional probability

Definition 1.19. *If $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space and $A, B \in \mathcal{F}$ are two events, then the **conditional probability** of the event $A \in \mathcal{F}$ with respect to the event $B \in \mathcal{F}$ is called the function of sets $\mathsf{P}\{A/B\}$ such that the relation*

$$\boxed{\mathsf{P}\{B \cap A\} = \mathsf{P}\{A/B\}\,\mathsf{P}\{B\}} \tag{1.68}$$

is valid for any sets $A, B \in \mathcal{F}$. Thus the conditional probability $\mathsf{P}\{A/B\}$ is the probability of A given that we know that B has occurred.

Several useful corollaries, directly following from this definition, are summarized below.

1.4.4 Bayes's formula

Corollary 1.4. (Bayes's formula) *For any two events (sets) $A, B \in \mathcal{F}$ the following connections hold:*

$$\boxed{\mathsf{P}\{B/A\}\,\mathsf{P}\{A\} = \mathsf{P}\{A/B\}\,\mathsf{P}\{B\}} \tag{1.69}$$

or, in other form,

$$\boxed{\begin{array}{l} \mathsf{P}\{B/A\} = \dfrac{\mathsf{P}\{A/B\}\,\mathsf{P}\{B\}}{\mathsf{P}\{A\}} \quad \text{if } \mathsf{P}\{A\} \neq 0 \\ \\ \mathsf{P}\{A/B\} = \dfrac{\mathsf{P}\{B/A\}\,\mathsf{P}\{A\}}{\mathsf{P}\{B\}} \quad \text{if } \mathsf{P}\{B\} \neq 0 \end{array}} \tag{1.70}$$

Proof. It follows from (1.68) and the fact that for any two events $A, B \in \mathcal{F}$ one has $\mathsf{P}\{B \cap A\} = \mathsf{P}\{A \cap B\}$. Corollary is proved. □

Corollary 1.5. (Law of total probability) *If $\{B_i\}_{i=1,\dots,N}$ is a partition of the set Ω of elementary events, i.e.,*

$$\boxed{B_i \in \mathcal{F},\ B_i \bigcap_{i \neq j} B_j = \emptyset,\ \bigcup_{i=1}^{N} B_i = \Omega} \tag{1.71}$$

then for any event $A \in \mathcal{F}$

$$\boxed{\mathsf{P}\{A\} = \sum_{i=1}^{N} \mathsf{P}\{A/B_i\}\,\mathsf{P}\{B_i\}} \tag{1.72}$$

Proof. Because $\{B_i\}_{i=1,\dots,N}$ is a partition of Ω we can conclude that $\sum_{i=1}^{N} \chi(\omega \in B_i) = 1$. Then using this identity we derive

$$\begin{aligned} \chi(\omega \in A) &= \chi(\omega \in A) \sum_{i=1}^{N} \chi(\omega \in B_i) \\ &= \sum_{i=1}^{N} \chi(\omega \in A)\,\chi(\omega \in B_i) \\ &= \sum_{i=1}^{N} \chi((\omega \in A) \cap (\omega \in B_i))\,\chi(\omega \in B_i) \\ &= \sum_{i=1}^{N} \chi((\omega \in A) \cap (\omega \in B_i)) \end{aligned}$$

Here we have used

$$\begin{aligned}
&\chi(\omega \in A)\,\chi(\omega \in B_i) \\
&\quad = \chi((\omega \in A) \cap (\omega \in B_i))\,\chi(\omega \in B_i) \\
&\quad = \chi((\omega \in A) \cap (\omega \in B_i))
\end{aligned}$$

for any sets $A, B_i \in \mathcal{F}$. This exactly means that $\mathsf{P}\{A\} = \sum_{i=1}^{N} \mathsf{P}\{A \cap B_i\}$. Applying then the formula (1.68) to the right-hand side of this relation for $B = B_i$ we finally obtain (1.72). The corollary is proven. □

Corollary 1.6. (Generalized Bayes formula) *If* $\{B_i\}_{i=1,\dots,N}$ *is a partition of* Ω *(see (1.71)) and if* $\mathsf{P}\{A\} \neq 0$*, then for any* $i = 1, \dots, N$

$$\boxed{\mathsf{P}\{B_i/A\} = \frac{\mathsf{P}\{A/B_i\}\,\mathsf{P}\{B_i\}}{\mathsf{P}\{A\}} = \frac{\mathsf{P}\{A/B_i\}\,\mathsf{P}\{B_i\}}{\sum_{j=1}^{N} \mathsf{P}\left\{A/B_j\right\}\mathsf{P}\left\{B_j\right\}}} \tag{1.73}$$

Proof. It results from the relations

$$\begin{aligned}
&\mathsf{P}\{A\} = \sum_{j=1}^{N} \mathsf{P}\left\{A/B_j\right\}\mathsf{P}\left\{B_j\right\} \\
&\mathsf{P}\left\{B_j/A\right\}\mathsf{P}\{A\} = \mathsf{P}\left\{A/B_j\right\}\mathsf{P}\left\{B_j\right\}
\end{aligned}$$

□

This page intentionally left blank

2 Random Variables

Contents

In this chapter a connection between measure theory and the basic notion of probability theory – a *random variable* – is established. In fact, random variables are the functions from the probability space to some other measurable space. The definition of a random variable as a measurable function is presented. Several simple examples of random variables are considered. The transformation of distributions for the class of functionally connected random variables is also analyzed.

2.1 Measurable functions and random variables

If any real-valued function describes a connection between points of reals and corresponding points of real line, a *random variable* states connection between any arbitrary set of possible outcomes of experiments and extended reals. Here we are going to develop a much more general definition of a measurable function, which realizes this more general notion, provided that certain *'measurability'* conditions are satisfied. Probability considerations may be used to motivate the concept of measurability.

2.1.1 Measurable functions

Let now $(\Omega, \mathcal{F})$ be a measurable space, and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a real line with the system $\mathcal{B}(\mathbb{R})$ of Borel sets. The following definition is the central one in this section.

Definition 2.1. *A real function $\xi = \xi(\omega)$ defined on $(\Omega, \mathcal{F})$ is said to be* ***an $\mathcal{F}$-measurable*** *(or* ***Borel measurable) function*** *or* ***random variable*** *if the following inclusion holds:*

$$\boxed{(\omega : \xi(\omega) \in B) \in \mathcal{F}} \tag{2.1}$$

for each set $B \in \mathcal{B}(\mathbb{R})$ or, equivalently, if the inverse image is a measurable set in Ω, i.e.

$$\boxed{\xi^{-1}(B) := (\omega : \xi(\omega) \in B) \in \mathcal{F}} \tag{2.2}$$

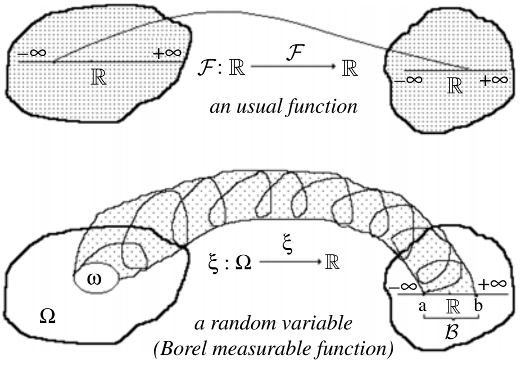

Fig. 2.1. Usual and $\mathcal{F}$-measurable functions.

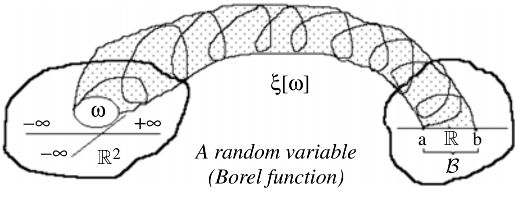

Fig. 2.2. A Borel function.

Remark 2.1. *Fig. 2.1 illustrates the main properties of usual functions, which state correspondence between each point in* $\mathbb{R}$ *(argument) and some point of* $\mathbb{R}$ *(value function), and an* $\mathcal{F}$*-measurable function, which state correspondence between each set B of possible values of function in* $\mathbb{R}$ *and some set B of corresponding realizations* ω *('a random factor') from* $\otimes$.

2.1.2 Borel functions and multidimensional random variables

Analogous definitions could be done in the case of a multidimensional 'random factor' when we consider situations involving more that one random variable associated with the same experiment.

Definition 2.2. *When*

$$\boxed{(\Omega, \mathcal{F}) = \left(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N)\right)} \tag{2.3}$$

then $\mathcal{B}(\mathbb{R}^N)$*-measurable functions are called* ***Borel functions***.

In fact, this is a simple generalization of the previous definition to the N-dimensional random factor. It is illustrated by Fig. 2.2. So, if a random factor ω can be measured by some physical device in R-scale, then the following 'termini' are synonyms:

$$\boxed{\underbrace{\text{An } \mathcal{F}\text{-measurable function} = \text{A Borel function}}_{\text{the mathematical terminus}} = \underbrace{\text{A random variable}}_{\text{the engineering terminus}}}$$

So far we have described the map from Ω to $(\mathbb{R}^N)$. To complete the picture we have to define a third component in the triplet $(\Omega, \mathcal{F}, \mathsf{P})$, namely, the appropriate probability measure.

Definition 2.3. *To each random variable $\xi = \xi(\omega)$ we associate an* ***induced probability measure*** P *through the relation*

$$\boxed{\mathsf{P}\{B\} := \mathsf{P}\{\omega : \xi(\omega) \in B\} = \mathsf{P}\left\{\xi^{-1}(B)\right\}} \tag{2.4}$$

for all $B \in \mathcal{B}$.

The next definition is the central one in this chapter.

Definition 2.4. *We say that* ***the random variable*** $\xi = \xi(\omega)$ ***is given on the probability space*** $(\Omega, \mathcal{F}, \mathsf{P})$ *if*

- *$\xi(\omega)$ is $\mathcal{F}$-measurable on $(\Omega, \mathcal{F})$;*
- *its distribution function $F_\xi(x)$ is defined by*

$$\boxed{F_\xi(x) := \mathsf{P}\{\omega : \xi(\omega) \le x\} \quad \forall x \in \mathbb{R}^N} \tag{2.5}$$

where the event $\{\omega : \xi(\omega) \le x\}$ is to be interpreted component-wise, that is,

$$\boxed{\{\omega : \xi(\omega) \le x\} := \bigcap_{k=1}^{N} \{\omega : \xi_k(\omega) \le x_k\}} \tag{2.6}$$

Intuitively, a *random variable* is a quantity that is measured in connection with a random experiment: if $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space and the outcome of the experiment corresponds to the point $\omega \in \Omega$, a measuring process is carried out to obtain a number $\xi(\omega)$. Thus, $\xi = \xi(\omega)$ is a function from the sample space Ω to the reals (or extended reals including $\pm\infty$) $\mathbb{R}^N$.

Remark 2.2. *If we are interested in a random variable ξ defined on a given probability space, we generally want to know the probability of all events involving ξ. The numbers $F_\xi(x)$ completely characterize the random variable ξ in the sense that they provide the probabilities of all events using the information on ξ. It is useful to understand that this information may be captured by a single function $F_\xi(x)$ from $\mathbb{R}^N$ to $\mathbb{R}$.*

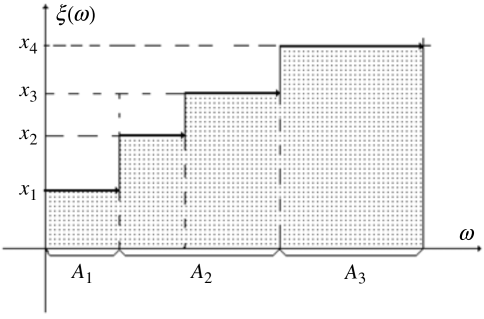

Fig. 2.3. An example of a simple random variable.

2.1.3 Indicators and discrete variables

Consider now the simplest example of a random variable, which we will often use hereafter.

Let $A \in \mathcal{F}$ be some event from a given σ-algebra $\mathcal{F}$ of all possible events. Define the function

$$\boxed{I_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}} \tag{2.7}$$

which we will call **the indicator-function** of a set A.

Definition 2.5. *A random variable $\xi = \xi(\omega)$ with values in $\mathbb{R}$ is called* ***a discrete random variable*** *if it has the representation*

$$\boxed{\xi(\omega) = \sum_{i=1}^{\infty} x_i I_{A_i}(\omega)} \tag{2.8}$$

where

- *x_i is a fixed point from $\mathbb{R}$,*
- $$\boxed{A_i \in \mathcal{F}: \quad \bigcup_{i=1}^{\infty} A_i = \Omega} \tag{2.9}$$

If a sum in (2.8) is finite then $\xi(\omega)$ is called a ***simple random variable****.*

Example 2.1. *Let $x_i = i$ $(i = 1, \ldots, n)$ and $A_i := \{\omega : \omega \in [i-1, i)\}$. Then the random variable $\xi(\omega)$ is a simple random variable (see Fig. 2.3).*

2.2 Transformation of distributions

This section describes some helpful rules for the calculation of distribution functions of random variables in the case of their functional transformations.

2.2.1 Functionally connected random variables

The following lemma gives a rigorous mathematical proof of the trivial *'physically clear'* fact:

'Any deterministic function of a random variable is also a random variable.'

Lemma 2.1. *Let* $\xi = \xi(\omega)$ *be a random variable, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*, and* $\phi = \phi(x)$ *be a Borel function. Then the function* $\eta(\omega) = \phi(\xi(\omega))$ *is a random function defined on the same probability space* $(\Omega, \mathcal{F}, \mathsf{P})$.

Proof. It follows from the relations:

$$\{\omega : \eta(\omega) \in B\} = \{\omega : \phi(\xi(\omega)) \in B\} = \{\omega : \xi(\omega) \in \phi^{-1}(B)\} \in \mathcal{F}$$

Lemma is proven. □

The following examples represent some random variables, constructed by applying a simple deterministic transformation to a given random variable:

- $$\boxed{\xi^{+} := \max\{\xi; 0\}} \tag{2.10}$$

- $$\boxed{\xi^{-} := -\min\{\xi; 0\}} \tag{2.11}$$

- $$\boxed{|\xi| := \xi^{+} + \xi^{-} \quad (\xi = \xi^{+} - \xi^{-})} \tag{2.12}$$
  $$\boxed{\xi^{n} \quad (n = 1, 2, \dots)} \tag{2.13}$$

All of these functions are also random variables given on the same probability space $(\Omega, \mathcal{F}, \mathsf{P})$ if $\xi = \xi(\omega)$ is a random variable defined on $(\Omega, \mathcal{F}, \mathsf{P})$. The proof of this fact is similar to the one given below.

Proposition 2.1. *Let* $\xi_1, \xi_2 \dots$ *be random variables. Then the following quantities are random variables too:*

1.

$$\boxed{\max\{\xi_1, \xi_2\} \quad \text{and} \quad \min\{\xi_1, \xi_2\}} \tag{2.14}$$

2.

$$\boxed{\sup_n \xi_n \quad \text{and} \quad \inf_n \xi_n} \tag{2.15}$$

3.

$$\boxed{\limsup_n \xi_n \quad \text{and} \quad \liminf_n \xi_n} \tag{2.16}$$

4. If $\{\xi_n(\omega)\}$ converges for all $\omega \in \Omega$, then $\lim_{n\to\infty} \xi_n(\omega)$ is a random variable too.

Proof.
1. For any $x \in \mathbb{R}$ we have

$$\begin{aligned}\{\omega : \max\{\xi_1, \xi_2\}(\omega) \le x\} &= \{\omega : \max\{\xi_1(\omega), \xi_2(\omega)\} \le x\} \\ &= \{\omega : \xi_1(\omega) \le x\} \cap \{\omega : \xi_2(\omega) \le x\} \in \mathcal{F}\end{aligned}$$

and

$$\begin{aligned}\{\omega : \min\{\xi_1, \xi_2\}(\omega) \le x\} &= \{\omega : \min\{\xi_1(\omega), \xi_2(\omega)\} \le x\} \\ &= \{\omega : \xi_1(\omega) \le x\} \cup \{\omega : \xi_2(\omega) \le x\} \in \mathcal{F}\end{aligned}$$

that proves (2.14).
2. Since a countable intersection of measurable sets is measurable, then

$$\left\{\omega : \sup_n \xi_n(\omega) \le x\right\} = \bigcap_n \{\omega : \xi_n(\omega) \le x\} \in \mathcal{F}$$

and, similarly, since a countable union of measurable sets is measurable, it follows

$$\left\{\omega : \inf_n \xi_n(\omega) \le x\right\} = \bigcup_n \{\omega : \xi_n(\omega) \le x\} \in \mathcal{F}$$

that proves (2.15).
3. To prove (2.16) notice that

$$\begin{aligned}\left\{\omega : \limsup_n \xi_n(\omega) \le x\right\} &= \left\{\omega : \lim_{n\to\infty} \sup_{m\ge n} \xi_n(\omega) \le x\right\} \\ = \left\{\omega : \inf_{n\to\infty} \sup_{m\ge n} \xi_n(\omega) \le x\right\} &= \left\{\omega : \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \xi_n(\omega) \le x\right\} \in \mathcal{F}\end{aligned}$$

and

$$\left\{\omega : \liminf_n \xi_n(\omega) \le x\right\} = \left\{\omega : \lim_{n\to\infty} \inf_{m\ge n} \xi_n(\omega) \le x\right\}$$
$$= \left\{\omega : \sup_{n\to\infty} \inf_{m\ge n} \xi_n(\omega) \le x\right\} = \left\{\omega : \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \xi_n(\omega) \le x\right\} \in \mathcal{F}$$

since by 2 $\inf_{m\ge n} \xi_n(\omega)$ is a random variable, and again by 2 $\inf_{n\to\infty} \sup_{m\ge n} \xi_n(\omega) \le x$ is a random variable too. Similarly for $\liminf_n \xi_n(\omega)$.
4. Since in this case $\inf_{n\to\infty} \sup_{m\ge n} \xi_n(\omega) = \lim_{n\to\infty} \inf_{m\ge n} \xi_n(\omega)$ the results is true. □

2.2.2 Transformation of densities

The following theorem gives the formula for calculation of the density $p_\phi(x)$ (1.49) of the distribution for the random variable $\phi(\xi(\omega))$, if it is available a density $p_\xi(x)$ of the distribution for a random variable $\xi(\omega)$.

Theorem 2.1. *Let $\xi = \xi(\omega)$ be a random value defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and its distribution function $F_\xi(x)$ is absolutely continuous on $\mathbb{R}$, i.e., $\xi(\omega)$ has a density of this distribution*

$$\boxed{p_\xi(x) := \frac{d}{dx} F_\xi(x)} \tag{2.17}$$

Let $y = \phi(x)$ be a strictly monotonic (increasing or decreasing) function on $\mathbb{R}$, which is differentiable and its derivative is not equal to zero, that is,

$$\boxed{\phi'(x) : \phi'(x) \ne 0 \quad \forall x \in \mathbb{R}} \tag{2.18}$$

Then $\eta(\omega) = \phi(\xi(\omega))$ is also a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and having the absolutely continuous distribution function $F_\eta(y)$ with the density of distribution $p_\eta(y)$ equal to

$$\boxed{p_\eta(y) = \frac{d}{dx} F_\xi(x) = \left.\frac{p_\xi(x)}{|\phi'(x)|}\right|_{x=\phi^{-1}(y)}, \quad y \in \mathbb{R}} \tag{2.19}$$

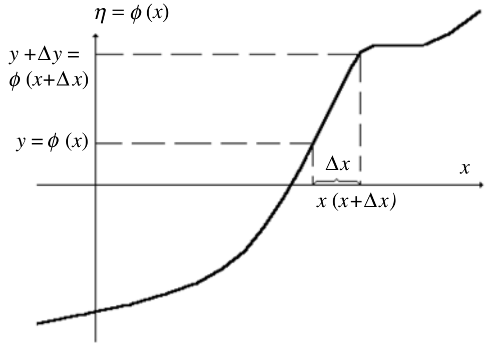

Fig. 2.4. A strictly monotonically increasing function.

Proof. Let $\phi(x)$ be a strictly monotonically increasing function. Then (see Fig. 2.4) by the mean-value theorem one has

$$\begin{aligned}
&\mathsf{P}\{\omega : y = \phi(x) < \eta \le \phi(x + \Delta x) = y + \Delta y\} \\
&\quad = \mathsf{P}\{\omega : x < \xi \le x + \Delta x\} \\
&\quad = \mathsf{P}\{\omega : \phi^{-1}(y) < \xi \le \phi^{-1}(y + \Delta y)\} = \int\limits_{\phi^{-1}(y)}^{\phi^{-1}(y+\Delta y)} p_\xi(x)dx \\
&\quad = p_\xi(\theta)|\phi^{-1}(y + \Delta y) - \phi^{-1}(y)|, \quad \theta \in [\phi^{-1}(y); \phi^{-1}(y + \Delta y)]
\end{aligned}$$

Taking $\Delta y \to 0$ we obtain:

$$\begin{aligned}
&\lim_{\Delta y \to 0} \frac{1}{\Delta y} \mathsf{P}\{\omega : y = \phi(x) < \eta \le \phi(x + \Delta x) = y + \Delta y\} \\
&\quad = p_\eta(y) = p_\xi(\phi^{-1}(y)) \left| \frac{d}{dy} \phi^{-1}(y) \right| = \left. \frac{p_\xi(x)}{|\phi'(x)|} \right|_{x=\phi^{-1}(y)}, \quad y \in \mathbb{R}
\end{aligned}$$

Analogous considerations holds for a monotonically decreasing function. The theorem is proven. □

For the more general class of functions, which are *partially strictly monotonic* (see Fig. 2.5), the following result holds.

Theorem 2.2. *Let, instead of (2.18), a function $y = \phi(x)$ be a partially strictly monotonic at each semi-open interval $X_i := (x_{i-1}, x_i]$ and differentiable on X_i function with a nonzero derivative on* int X_i, *i.e.,*

$$\boxed{\begin{array}{l} \phi'(x) : \phi'(x) \neq 0 \quad \forall x \in X_i \\ X_i \bigcap\limits_{i \neq j} X_j = \varnothing, \quad \bigcup\limits_i X_i = \mathbb{R} \end{array}}$$

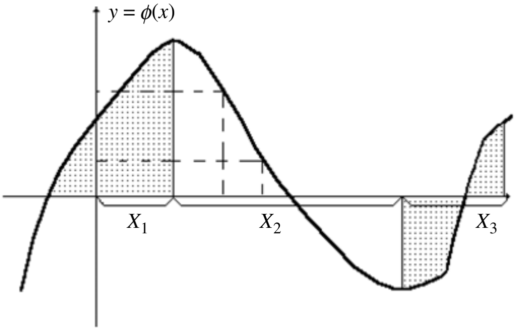

Fig. 2.5. A partially strictly monotonically increasing function.

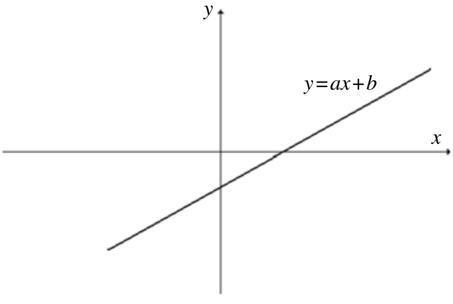

Fig. 2.6. A linear (affine) transformation.

Then $\eta(\omega) = \phi(\xi(\omega))$ *is also a random variable, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$, *and having the absolutely continuous distribution function* $F_\eta(y)$ *with the density of distribution* $p_\eta(y)$ *equal to*

$$\boxed{p_\eta(y) = \sum_i \chi\,(x \in \operatorname{int} X_i)\, \frac{p_\xi(x)}{|\phi'(x)|}\bigg|_{x=\phi^{-1}(y)}} \tag{2.20}$$

for any $y \in \mathbb{R}$ *such that* $x = \phi^{-1}(y) \in \operatorname{int} X_{i_0}$ *for some* $i = i_0$.

Proof. It follows from the equality

$$\mathsf{P}\{\omega : y < \eta \le y + \Delta y\} = \sum_i \mathsf{P}\{\omega : x < \xi \le x + \Delta x, x, (x + \Delta x) \in X_i\}$$

after the application of the formula (2.19) to each term of the last expression and tending Δy to zero. □

Example 2.2. (The class of linear (affine) transformations) *Let (see Fig. 2.6):*

$$\boxed{y = \phi(x) = ax + b, \quad a \neq 0} \tag{2.21}$$

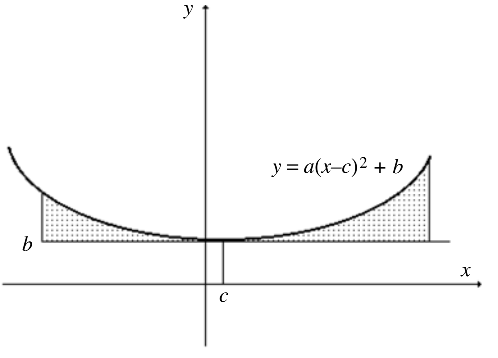

Fig. 2.7. A quadratic transformation.

As in this case we deal with a strictly monotonic functions, we can use the formula (2.19):

$$p_\eta(y) = \left.\frac{p_\xi(x)}{|\phi'(x)|}\right|_{x=\phi^{-1}(y)} = \frac{1}{a} p_\xi\left(\frac{y-b}{a}\right) \tag{2.22}$$

Example 2.3. (The class of quadratic transformations) *Let now (see Fig. 2.7)*

$$\boxed{y = \phi(x) = a(x-c)^2 + b, \quad a > 0} \tag{2.23}$$

The real line in this case could be separated in two regions:

$$X_1 = \{x : -\infty < x \le c\}, \quad X_2 = \{x : c < x < \infty\}$$

Applying to this class of transformation the formula (2.20), for any $y > b$ we obtain:

$$\begin{aligned} p_\eta(y) = {} & \chi\,(x \in X_1) \left.\frac{p_\xi(x)}{|\phi'(x)|}\right|_{x=-\sqrt{\frac{y-b}{a}}+c} + \chi\,(x \in X_2) \left.\frac{p_\xi(x)}{|\phi'(x)|}\right|_{x=\sqrt{\frac{y-b}{a}}+c} \\ = {} & \frac{\chi\left(-\sqrt{\frac{y-b}{a}} + c \in X_1\right) p_\xi\left(-\sqrt{\frac{y-b}{a}} + c\right)}{2\sqrt{a(y-b)}} \\ & + \frac{\chi\left(\sqrt{\frac{y-b}{a}} + c \in X_2\right) p_\xi\left(\sqrt{\frac{y-b}{a}} + c\right)}{2\sqrt{a(y-b)}} \end{aligned} \tag{2.24}$$

2.3 Continuous random variables

This section introduces two useful notions, which help to simplify the terminology in the next texts, and also present a very important theorem on *monotone convergence*, which plays a key role for the construction of *Lebesgue integral* or, in other words, *the operator of mathematical expectation*.

2.3.1 Continuous variables

The following definitions will be useful hereafter.

Definition 2.6. *A random variable* $\xi = \xi(\omega)$ *given on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *is said to be*

- ***continuous***, *if its distribution function* $F_\xi(x)$ *is continuous for all* $x \in \mathbb{R}$*;*
- ***absolutely continuous***, *if there exists a nonnegative function* $f_\xi(x)$*, called its density, such that*

$$\boxed{F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du, \quad x \in \mathbb{R}} \tag{2.25}$$

Surely, the definition of absolutely continuous random variable is closely related to the definition of absolutely continuous measures, given in Chapter 1 (see (1.48)).

2.3.2 Theorem on monotone approximation

This theorem gives a simple approximating presentation for any random variable, using the notion of a simple random variable with a finite number of terms in (2.8).

Theorem 2.3. (on monotone approximation) *For each random variable* $\xi = \xi(\omega)$ *given on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*, there is a sequence of simple random variables* $\xi_1, \xi_2, \ldots$ *having the structure*

$$\boxed{\xi_n(\omega) = \sum_{k=1}^{N(n)} x_{k,n} I_{A_{k,n}}(\omega)} \tag{2.26}$$

(where $N(n)$ *is some given monotonically increasing function of* n*,* $\{x_{k,n}\}$ *are given numbers, and* $I_{A_{k,n}}(\omega)$ *are indicators of the sets* $A_{k,n}$*) such that for all* $\omega \in \Omega$

$$\boxed{|\xi_n(\omega)| \le |\xi(\omega)|} \tag{2.27}$$

and

$$\boxed{|\xi_n(\omega) - \xi(\omega)| \underset{n\to\infty}{\to} 0} \tag{2.28}$$

If $\xi(\omega) \ge 0$ *for all* $\omega \in \Omega$*, then there exists a sequence* $\{\xi_n(\omega)\}_{n\ge1}$ *of simple random variables* $\xi_1, \xi_2, \ldots$ *such that*

$$\boxed{\xi_n(\omega) \underset{n\to\infty}{\uparrow} \xi(\omega)} \tag{2.29}$$

that is, for each $\omega \in \Omega$ *the sequence* $\{\xi_n(\omega)\}_{n\ge1}$ *converges to* $\xi(\omega)$ *monotonically from below.*

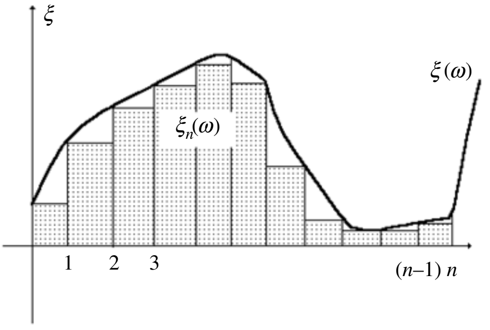

Fig. 2.8. A monotone approximation from down.

Proof. Define $I_{A_{k,n}}$ as an indicator of the set

$$A_{k,n} := \left\{ \omega : \frac{k-1}{2^n} \le \xi(\omega) < \frac{k}{2^n} \right\}$$

and $N(n) := n2^n$. Put then

$$\xi_n(\omega) := \sum_{k=1}^{N(n)} \frac{k-1}{2^n} I_{A_{k,n}}(\omega) + nI\{\omega : \xi(\omega) \ge n\}$$

Then, for a non-negative random variable $\xi(\omega) \ge 0$ and any $\omega \in \Omega$ we have (see Fig. 2.8) $\xi_n(\omega) \le \xi(\omega)$, $\xi_n(\omega) \uparrow \xi(\omega)$.

In general case (when $\xi(\omega)$ is not obligatory non-negative) we have

$$\xi(\omega) = \xi^+(\omega) - \xi^-(\omega)$$

and one can define two sequences such that

$$\xi_n^+(\omega) \uparrow \xi^+(\omega)$$
$$\xi_n^-(\omega) \uparrow \xi^-(\omega)$$

for which the following statements hold: for any $\omega \in \Omega$

$$\xi_n(\omega) := \xi_n^+(\omega) - \xi_n^-(\omega) \underset{n\to\infty}{\to} \xi(\omega)$$
$$|\xi_n(\omega)| = \xi_n^+(\omega) + \xi_n^-(\omega) \le \xi^+(\omega) + \xi^-(\omega) = |\xi(\omega)|$$

Theorem is proven. □

Hereafter we will use only the simple random variable $\xi_n(\omega)$, defined on the $(\Omega, \mathcal{F}, \mathsf{P})$ by the formula

$$\boxed{\begin{aligned} \xi_n(\omega) &:= \xi_n^+(\omega) - \xi_n^-(\omega) \\ \xi_n^+(\omega) &:= \sum_{k=1}^{N(n)} \frac{k-1}{2^n} I_{A_{k,n}^+}(\omega) + nI\{\omega : \xi^+(\omega) \geq n\} \\ A_{k,n}^+ &:= \left\{\omega : \frac{k-1}{2^n} \leq \xi^+(\omega) < \frac{k}{2^n}\right\} \\ \xi_n^-(\omega) &:= \sum_{k=1}^{N(n)} \frac{k-1}{2^n} I_{A_{k,n}^-}(\omega) + nI\{\omega : \xi^-(\omega) \geq n\} \\ A_{k,n}^- &:= \left\{\omega : \frac{k-1}{2^n} \leq \xi^-(\omega) < \frac{k}{2^n}\right\} \end{aligned}} \tag{2.30}$$

This page intentionally left blank

3 Mathematical Expectation

Contents

This chapter introduces the most important notion in probability theory – *mathematical expectation* or (*in mathematical language*) a Lebesgue Integral taken with respect to a probabilistic measure (see, for example, Chapters 15 and 16 in Poznyak (2008)). Physically, this operator presents some sort of '*averaging*', or a probabilistic version of '*the center of gravity of a physical body*'. In fact, it *compresses* the information on a random variable in to a single number.

There exist two possibilities to introduce the operator of the *mathematical expectation:*

- using the so-called **axiomatic approach**, suggested by Whittle (Whittle, 1984), which is based on some evident properties of 'an *averaging operation*', and showing then that it is exactly the Lebesgue–Stieltjes integration;
- introducing directly the mathematical expectation operator as **the Lebesgue–Stieltjes integration with respect to a distribution function or probability measure**.

Here we will present both of these approaches and demonstrate their internal interconnection.

3.1 Definition of mathematical expectation

3.1.1 Whittle axioms

Definition 3.1. (the Whittle axioms) *The operator* $\mathsf{E}\{\cdot\}$ *defining the correspondence between each random value* $x(\omega)$ *(given on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*) and a real variable*

$$\boxed{x(\omega) \overset{\mathsf{E}}{\to} m \in \mathbb{R}} \tag{3.1}$$

is called the ***mathematical expectation of the random variable*** $x(\omega)$*, or, its '****average value****' with respect to all possible realizations* $\omega \in \Omega$*, if it has the following properties:*

A1. 'the average value' of a non-negative random variable is also non-negative, i.e., if $x(\omega) \geq 0$ *for all* $\omega \in \Omega$ *then*

$$\boxed{\mathsf{E}\{x(\omega)\} \geq 0} \tag{3.2}$$

A2. 'the average value' of the sum of two random variables is equal to the corresponding sum of their average values, i.e., for any $c_1, c_2 \in \mathbb{R}$ *and any random variables* $x_1(\omega)$ *and* $x_2(\omega)$ *defined on* $(\Omega, \mathcal{F}, \mathsf{P})$

$$\boxed{\mathsf{E}\{c_1 x_1(\omega) + c_2 x_2(\omega)\} = c_1 \mathsf{E}\{x_1(\omega)\} + c_2 \mathsf{E}\{x_2(\omega)\}} \tag{3.3}$$

A3. 'the average value of a constant' is the same constant (scaling property), i.e.,

$$\boxed{\mathsf{E}\{1\} = 1} \tag{3.4}$$

A4. if any sequence of simple random variables $\xi_n(\omega)$ *monotonically converges to a random variable* $\xi_n(\omega)$ *for all* $\omega \in \Omega$*, i.e.,*

$$\xi_n(\omega) \leq \xi_{n+1}(\omega), \quad \xi_n(\omega) \uparrow \xi(\omega)$$

or

$$\xi_n(\omega) \geq \xi_{n+1}(\omega), \quad \xi_n(\omega) \downarrow \xi(\omega)$$

then the corresponding sequence of their average values also converges monotonically to the average value of the limit random variable, that is,

$$\boxed{\lim_{n\to\infty} \mathsf{E}\{\xi_n(\omega)\} = \mathsf{E}\{\xi(\omega)\}} \tag{3.5}$$

A5. the probability of any event $A \in \Omega$ *is equal to 'the average value' of its characteristic function, i.e.,*

$$\boxed{\mathsf{P}\{A\} := \mathsf{E}\{\chi\,(\omega \in A)\}} \tag{3.6}$$

All axioms presented above have very clear '*physical interpretation*', if we will consider the operator of mathematical expectation as some sort of '**averaging over the ensemble of all possible realizations** ω'. Indeed,

- the axiom A1 illustrates the evident fact that *'the average value of any group of non-negative plants is also non-negative'*;
- the axiom A2 reflects the additivity property: *'the average of the sum of elements is equal to the sum of their average values'*;
- the axiom A3 is also absolutely clear: *'averaging of any constant leads to the same constant'*;
- the axiom A4 is not so evident as any other, but it also doesn't provoke any negative 'emotions': *'if any sequence of random variables monotonically converges to a limit variable, then the corresponding sequence of average values converges to the average value of the limit variable'*;
- the last axiom A5 introduces the natural connection between a given measure P (participating in the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ definition), and the considered 'averaging method': *'the probability of any event* $A \in \Omega$ *is equal to the average value of its characteristic function'*.

3.1.2 Mathematical expectation as the Lebesgue integral

This subsection gives the description of the *mathematical expectation operator* (3.1) based on the *Lebesgue integral definition* (see Chapters 15, 16 in Poznyak (2008)).

A. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $\xi = \xi(\omega)$ be *a simple random variable*, that is,

$$\boxed{\xi(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega)} \tag{3.7}$$

Definition 3.2. *The* ***mathematical expectation*** $\mathsf{E}\{\xi\}$ *of the simple random variable* ξ *(3.7) is defined by*

$$\boxed{\mathsf{E}\{\xi\} := \sum_{i=1}^{n} x_i \mathsf{P}(A_i)} \tag{3.8}$$

This definition is consistence in the sense that $\mathsf{E}\{\xi\}$ does not depend on the particular representation of ξ in the form (3.7).

B. Let now $\xi = \xi(\omega)$ be a non-negative random variable, i.e., $\xi(\omega) \geq 0$. Then by Theorem 2.3 on the monotone approximation it follows that for all $\omega \in \Omega$ there exists a sequence $\{\xi_n(\omega)\}_{n\geq 1}$ of simple random variables $\xi_1, \xi_2, \ldots$ such that

$$\boxed{\xi_n(\omega) \underset{n\to\infty}{\uparrow} \xi(\omega)} \tag{3.9}$$

that is, for each $\omega \in \Omega$ the sequence $\{\xi_n(\omega)\}_{n\geq 1}$ converges to $\xi(\omega)$ monotonically from below. Based on the definition (3.8) we may conclude that

$$\boxed{\mathsf{E}\{\xi_n\} \leq \mathsf{E}\{\xi_{n+1}\}} \tag{3.10}$$

and hence, $\lim\limits_{n\to\infty} \mathsf{E}\{\xi_n\}$ exists (possibly, with the value $+\infty$).

Definition 3.3. *The* ***mathematical expectation*** $\mathsf{E}\{\xi\}$ *of a* ***non-negative random variable*** $\xi(\omega)$ *is defined by*

$$\boxed{\mathsf{E}\{\xi\} = \lim_{n\to\infty} \mathsf{E}\{\xi_n\}} \tag{3.11}$$

To see that this definition is consistent, we need to show that the limit in (3.11) is independent of the approximating sequence $\{\xi_n\}$, or, in other words, it is independent of the partition of Ω.

Lemma 3.1. *Let* $\{A_k : k = 1, \ldots, n\}$ *and* $\left\{B_j : j = 1, \ldots, m\right\}$ *be two partitions of* Ω*, such that*

$$\boxed{X = \sum_{k=1}^{n} x_k I_{A_k}(\omega) \quad \text{and} \quad Y = \sum_{j=1}^{m} y_j I_{B_j}(\omega)} \tag{3.12}$$

Then

$$\boxed{\sum_{i=1}^{n} x_k \mathsf{P}\left(A_k\right) = \sum_{j=1}^{m} y_j \mathsf{P}\left(B_j\right)} \tag{3.13}$$

Proof. One has

$$\mathsf{P}\left(A_k\right) = \sum_{j=1}^{m} \mathsf{P}\left(A_k \cap B_j\right) \quad \text{and} \quad \mathsf{P}\left(B_j\right) = \sum_{k=1}^{m} \mathsf{P}\left(A_k \cap B_j\right)$$

that implies

$$\begin{aligned}
\sum_{k=1}^{n} x_k \mathsf{P}\left(A_k\right) &= \sum_{k=1}^{n} \sum_{j=1}^{m} x_k \mathsf{P}\left(A_k \cap B_j\right) \\
\sum_{j=1}^{m} y_j \mathsf{P}\left(B_j\right) &= \sum_{j=1}^{m} \sum_{k=1}^{n} y_j \mathsf{P}\left(A_k \cap B_j\right)
\end{aligned}$$

Since the sets $\left\{A_k \cap B_j : k = 1, \ldots, n;\ j = 1, \ldots, m\right\}$ also form a partition of Ω it follows that $x_k = y_j$ whenever $A_k \cap B_j \neq \varnothing$ that proves the lemma. □

C. In general case, any random variable $\xi(\omega)$ (not obligatory non-negative) can be represented as

$$\boxed{\begin{gathered}
\xi(\omega) = \xi^{+}(\omega) - \xi^{-}(\omega) \\
\xi^{+}(\omega) := \max\{\xi(\omega); 0\}, \quad \xi^{-}(\omega) := -\min\{\xi(\omega); 0\}
\end{gathered}} \tag{3.14}$$

Definition 3.4. *The* ***mathematical expectation*** $\mathsf{E}\{\xi\}$ ***of a random variable*** $\xi(\omega)$ *given on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*, or, its '****average value****' with respect to all possible realizations* $\omega \in \Omega$*, is defined by*

$$\boxed{\mathsf{E}\{\xi\} := \mathsf{E}\{\xi^{+}\} - \mathsf{E}\{\xi^{-}\}} \tag{3.15}$$

It is also called the ***Lebesgue–Stieltjes*** *(or, simply,* ***Lebesgue****)* ***integral*** *of the* $\mathcal{F}$*-measurable function* $\xi(\omega)$ *with respect to the probability measure* P *for which we shall use the notation*

$$\boxed{\mathsf{E}\{\xi\} := \int_{\omega\in\Omega} \xi(\omega)\mathsf{P}(d\omega) \overset{or}{=} \int_{\Omega} \xi d\mathsf{P}} \tag{3.16}$$

We say that the expectation $\mathsf{E}\{\xi\}$ *is* ***finite*** *if both*

$$\mathsf{E}\{\xi^+\} < \infty \quad \text{and} \quad \mathsf{E}\{\xi^-\} < \infty$$

Remark 3.1. *It is easy to verify that the mathematical expectation* $\mathsf{E}\{\xi\}$ *defined by (3.15) satisfies all Whittle axioms A1–A5, and inverse. So,* ***both definitions*** *(3.1) and (3.15) are* ***equivalent****.*

3.1.3 Moments, mean, variance, median and α-quantile

Definition 3.5. *If* ξ *is a random variable then the*

- ***moments*** *are*

$$\boxed{\mathsf{E}\{\xi^n\}, \quad n = 1, 2, \ldots} \tag{3.17}$$

- ***central moments*** *are*

$$\boxed{\mathsf{E}\left\{(\xi - \mathsf{E}\{\xi\})^n\right\}, \quad n = 1, 2, \ldots} \tag{3.18}$$

- ***absolute moments*** *are*

$$\boxed{\mathsf{E}\{|\xi|^n\}, \quad n = 1, 2, \ldots} \tag{3.19}$$

- ***absolute central moments*** *are*

$$\boxed{\mathsf{E}\left\{|\xi - \mathsf{E}\{\xi\}|^n\right\}, \quad n = 1, 2, \ldots} \tag{3.20}$$

The first moment $\mathsf{E}\{\xi\}$ *is called* ***mean*** *and the second central moment is called* ***variance****. They are denoted by*

$$\boxed{\begin{array}{c} m := \mathsf{E}\{\xi\} \\ \operatorname{var}\xi := \mathsf{E}\left\{(\xi - \mathsf{E}\{\xi\})^2\right\} \end{array}} \tag{3.21}$$

All definitions above are true provided the relevant quantities exist.

Definition 3.6. *The* ***median****, denoted by* $\operatorname{med}(\xi)$*, of a random variable* ξ *is called a real number such that*

$$\boxed{\mathsf{P}\{\xi \le \operatorname{med}(\xi)\} = \mathsf{P}\{\xi > \operatorname{med}(\xi)\} = \frac{1}{2}} \tag{3.22}$$

or, in other words, it is a kind of 'center' of the given distribution P *in the sense that a half of the probability mass lies to the left-hand side of it and a half of it on the right.*

Remark 3.2. *A* ***median always exists*** *in contrast to moments or absolute moments which need not.*

Definition 3.7. *The* α***-quantile*** *of a random variable* ξ *is called a real number* $\lambda_\alpha(\xi)$ *such that*

$$\boxed{\mathsf{P}\{\xi \geq \lambda_\alpha(\xi)\} \geq \alpha} \tag{3.23}$$

Remark 3.3. *The median is thus a 0.5-quantile.*

3.2 Calculation of mathematical expectation

In this section we will present the rules of the mathematical expectation calculation for discrete, continuous and absolutely continuous random variables.

3.2.1 Discrete variables

Lemma 3.2. *If* $\xi_n(\omega)$ *is a simple random variable (a finite discrete variable) (2.26), i.e.,*

$$\boxed{\xi_n(\omega) = \sum_{k=1}^{N(n)} x_{k,n} I_{A_{k,n}}(\omega)} \tag{3.24}$$

then

$$\boxed{\mathsf{E}\{\xi_n(\omega)\} = \sum_{k=1}^{N(n)} x_{k,n} \mathsf{P}\{A_{k,n}(\omega)\}} \tag{3.25}$$

Proof. It results directly from (3.8). □

Lemma 3.3. *If* $\xi(\omega)$ *is* ***a discrete random variable*** *(2.8), i.e.,*

$$\boxed{\xi(\omega) = \sum_{i=1}^{\infty} x_i I_{A_i}(\omega)} \tag{3.26}$$

then

$$\boxed{\mathsf{E}\{\xi(\omega)\} = \sum_{i=1}^{\infty} x_i \mathsf{P}\{A_i(\omega)\}} \tag{3.27}$$

Proof. It follows from Definition 3.3 and Lemma 3.1. □

The following examples illustrate these rules.

Example 3.1. (the toss of a die) *The sample space in this case is*

$$\Omega = \{\omega\} = \{1; 2; 3; 4; 5; 6\}$$

The measure is given by

$$\mathsf{P}\{\omega = i\} = \frac{1}{6} \quad \forall i = 1, \ldots, 6$$

Take $x(\omega) = \omega$. *Applying the previous lemma we obtain:*

$$\begin{aligned} \mathsf{E}\{x(\omega)\} &= \mathsf{E}\left\{\sum_{i=1}^{6} i\chi(\omega = i)\right\} = \sum_{i=1}^{6} i\mathsf{P}\{\omega = i\} \\ &= \sum_{i=1}^{6} \frac{i}{6} = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3.5 \end{aligned}$$

Example 3.2. (Bernoulli's variable) *Bernoulli's variable is a binary variable defined by*

$$x(\omega) = \begin{cases} 1 & \text{with the probability } p \\ 0 & \text{with the probability } q \end{cases}$$

where $q = 1 - p \in [0, 1]$. *Then by the previous lemma*

$$\begin{aligned} \mathsf{E}\{x(\omega)\} &= \mathsf{E}\{1\chi(x(\omega) = 1) + 0\chi(x(\omega) = 0)\} \\ &= \mathsf{E}\{\chi(x(\omega) = 1)\} = p \end{aligned}$$

3.2.2 Continuous variables

Lemma 3.4. *If* $\xi(\omega)$ *is* ***a continuous random variable*** *(2.5), i.e., its distribution function* $F_\xi(x)$ *is a continuous function, then*

$$\boxed{\mathsf{E}\{\xi(\omega)\} = \int_{-\infty}^{+\infty} x F_\xi(dx)} \tag{3.28}$$

where

$$\boxed{F_\xi(dx) := F_\xi(x + dx) - F_\xi(x) = \mathsf{P}\{\omega : \xi(\omega) \in (x, x + dx]\}} \tag{3.29}$$

Proof. Consider a simple random variable $\{\xi_n(\omega)\}$ (3.9), which monotonically from below converges to $\xi(\omega)$. Then using the axioms by Definition 3.3:

$$\begin{aligned}
\mathsf{E}\{\xi(\omega)\} &= \lim_{n\to\infty} \mathsf{E}\{\xi_n(\omega)\} \\
&= \lim_{n\to\infty} \mathsf{E}\left\{\sum_{k=1}^{N(n)} x_{k,n} I_{A_{k,n}}(\{\omega \in A_{k,n}\} = \{\xi_n(\omega) = x_{k,n}\})\right\} \\
&= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \mathsf{E}\{I_{A_{k,n}}(\{\omega \in A_{k,n}\} = \{\xi_n(\omega) = x_{k,n}\})\} \\
&= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \mathsf{P}\{\omega : \xi_n(\omega) = x_{k,n}\}
\end{aligned}$$

Taking into account the relation

$$\mathsf{P}\{\omega : \xi_n(\omega) = x_{k,n}\} = \sum_j \mathsf{P}\{\omega : x_j \le \omega \le x_j + \Delta_{j,n} | \xi_n(\omega) = x_{k,n}\} \tag{3.30}$$

when $(\Delta_{j,n} \underset{n\to\infty}{\to} 0$ for all $j)$ we get:

$$\begin{aligned}
\mathsf{E}\{\xi(\omega)\} &= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \mathsf{P}\{\omega : \xi_n(\omega) = x_{k,n}\} \\
&= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \sum_j \mathsf{P}\{\omega : x_j \le \omega \le x_j + \Delta_{j,n} | \xi_n(\omega) = x_{k,n}\} \\
&= \int_{-\infty}^{+\infty} x \mathsf{P}\{\omega : \xi(\omega) \in (x, x + dx]\}
\end{aligned} \tag{3.31}$$

The relation (3.29) leads to the presentation (3.28). Lemma is proven. □

3.2.3 Absolutely continuous variables

Lemma 3.5. *If $\xi(\omega)$ is* ***an absolutely continuous random variable*** *(2.5),i.e., its distribution function $F_\xi(x)$ can be expressed in the form*

$$\boxed{F_\xi(x) = \int_{-\infty}^{x} f_\xi(u) du} \tag{3.32}$$

then

$$\boxed{\mathsf{E}\{\xi(\omega)\} = \int_{-\infty}^{+\infty} x f_\xi(x) dx} \tag{3.33}$$

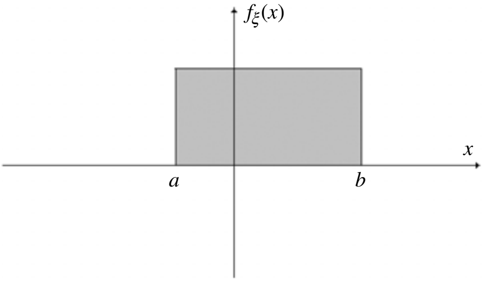

Fig. 3.1. A uniform distribution.

Proof. From (3.30) and (3.32) it follows that

$$\mathsf{P}\{\omega : \xi_n(\omega) = x_{k,n}\}$$
$$= \sum_j \mathsf{P}\{\omega : x_j \le \omega \le x_j + \Delta_{j,n} | \xi_n(\omega) = x_{k,n}\} = \sum_j \int_{x_j}^{x_j+\Delta_{j,n}} f_\xi(u)du$$

Substitution of this expression into (3.31) gives

$$\begin{aligned}
\mathsf{E}\{\xi(\omega)\} &= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \mathsf{P}\{\omega : \xi_n(\omega) = x_{k,n}\} \\
&= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \sum_j \mathsf{P}\{\omega : x_j \le \omega \le x_j + \Delta_{j,n} | \xi_n(\omega) = x_{k,n}\} \\
&= \lim_{n\to\infty} \sum_{k=1}^{N(n)} x_{k,n} \sum_j \int_{x_j}^{x_j+\Delta_{j,n}} f_\xi(u)du \\
&= \lim_{n\to\infty} \sum_j \sum_{k=1}^{N(n)} x_{k,n} f_\xi(x_{k,n}) \Delta_{j,n} = \int_{-\infty}^{+\infty} x f_\xi(x)dx
\end{aligned}$$

Lemma is proven. □

Example 3.3. (A uniform distribution) *In this situation a random variable $\xi(\omega)$ has the density (see Fig. 3.1)*

$$\boxed{f_\xi(x) = \frac{1}{b-a}, \quad x \in [a, b], a < b} \tag{3.34}$$

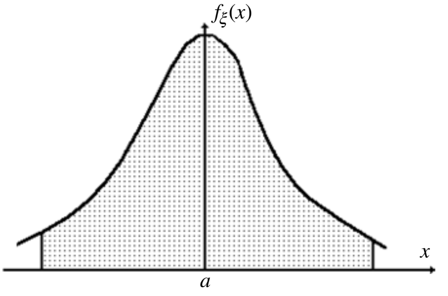

Fig. 3.2. A Gaussian distribution.

By (3.33) we directly obtain:

$$\mathsf{E}\{\xi(\omega)\} = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{b+a}{2}$$

$$\mathsf{E}\{\xi^2(\omega)\} = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3}$$

$$\begin{aligned} \operatorname{var} \xi &= \mathsf{E}\{(\xi(\omega) - \mathsf{E}\{\xi(\omega)\})^2\} \\ &= \frac{b^2 + ab + a^2}{3} - \frac{(b-a)^2}{4} = \frac{(b-a)^2}{12} \end{aligned}$$

3.2.3.1 Gaussian random variables

A Gaussian or normal random variable $\xi(\omega)$ has the density (see Fig. 3.2)

$$\boxed{f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right), \quad x \in [-\infty, \infty], \sigma > 0} \tag{3.35}$$

By (3.33) we directly obtain

$$\mathsf{E}\{\xi(\omega)\} = \int_{-\infty}^{\infty} x f_\xi(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx$$

Changing variables in this integral

$$\bar{x} := \frac{x-a}{\sigma}, \quad x = \sigma\bar{x} + a, \quad dx = \sigma d\bar{x}$$

we get

$$\mathsf{E}\{\xi(\omega)\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{\bar{x}=-\infty}^{\infty}(\sigma\bar{x}+a)\exp\left(-\frac{(\bar{x})^2}{2}\right)\sigma d\bar{x} = \frac{1}{\sqrt{2\pi}}(\sigma I_0 + a I_1) \quad (3.36)$$

where

$$I_0 := \int_{\bar{x}=-\infty}^{\infty}\bar{x}e^{-\frac{(\bar{x})^2}{2}}d\bar{x},\ I_1 := \int_{\bar{x}=-\infty}^{\infty}e^{-\frac{(\bar{x})^2}{2}}d\bar{x}$$

Taking into account that the function $g(\bar{x}) := \bar{x}e^{-\frac{(\bar{x})^2}{2}}$ is the odd function, i.e., $g(-\bar{x}) = -g(\bar{x})$, we conclude that $I_0 = 0$. To calculate I_1 let us represent it as $I_1 = 2\int_{\bar{x}=0}^{\infty} e^{-\frac{(\bar{x})^2}{2}}d\bar{x}$. Then changing variables as $\bar{x} = ut, u \geq 0, d\bar{x} = udt$ we obtain $I_1 = 2u\int_{t=0}^{\infty} e^{-\frac{u^2t^2}{2}}dt$ or, in another form,

$$I_1 e^{-\frac{u^2}{2}} = 2ue^{-\frac{u^2}{2}}\int_{t=0}^{\infty}e^{-\frac{u^2t^2}{2}}dt$$

Integrating then both sides of this inequality from $u = 0$ up to $u = \infty$ we derive:

$$\begin{aligned}\int_{u=0}^{\infty} I_1 e^{-\frac{u^2}{2}}du = \frac{1}{2}I_1^2 &= 2\int_{u=0}^{\infty}ue^{-\frac{u^2}{2}}\left[\int_{t=0}^{\infty}e^{-\frac{u^2t^2}{2}}dt\right]du \\ &= 2\int_{t=0}^{\infty}\left[\int_{u=0}^{\infty}ue^{-\frac{u^2(1+t)^2}{2}}du\right]dt \\ &= 2\int_{t=0}^{\infty}\left[\int_{v=u^2/2=0}^{\infty}e^{-v(1+t)^2}dv\right]dt \\ &= 2\int_{t=0}^{\infty}\frac{dt}{1+t^2} = 2\arctan\{t\}\mid_{t=0}^{t=\infty} = 2\frac{\pi}{2} = \pi\end{aligned}$$

Hence, $I_1^2 = 2\pi$ and therefore

$$I_1 = \sqrt{2\pi} \quad (3.37)$$

Substituting (3.37) into (3.36) we finally derive that

$$\boxed{\mathsf{E}\{\xi(\omega)\} = a} \quad (3.38)$$

To calculate the second moment $m_2 := \mathsf{E}\{\xi^2(\omega)\}$ of a Gaussian random variable $\xi(\omega)$ let us differentiate both sides of the formula (3.38) with respect to the parameter a:

$$\frac{d}{da}\mathsf{E}\{\xi(\omega)\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x\frac{(x-a)}{\sigma^2}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$= \frac{1}{\sigma^2}\int_{-\infty}^{\infty} x^2\frac{(x-a)}{\sigma^2}e^{-\frac{(x-a)^2}{2\sigma^2}}dx - \frac{a}{\sigma^2}\int_{-\infty}^{\infty} x\frac{(x-a)}{\sigma^2}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$= \frac{m_2}{\sigma^2} - \frac{a^2}{\sigma^2} = 1$$

from which it follows that

$$m_2 = \sigma^2 + a^2 \tag{3.39}$$

Combining then (3.38) and (3.39) we get

$$\boxed{\mathsf{E}\{(\xi(\omega)-a)^2\} = m_2 - a^2 = \sigma^2} \tag{3.40}$$

i.e., the parameters a and σ^2 in the Gaussian distribution (3.35) have 'the direct physical sense':

- a is a mathematical expectation of the corresponding Gaussian random variable or, in other words, the center of the Gaussian distribution;
- σ^2 is the variance of this Gaussian distribution or its *'variance'*.

Since the Gaussian density function $f_\xi(x)$ (3.35), even in the case $a = 0, \sigma^2 = 1$, is not a simple function, its integral

$$\boxed{\Phi(x) = \mathsf{P}\{\xi < x\} = \int_{-\infty}^{x} f_\xi(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt} \tag{3.41}$$

can be calculated only numerically. In many applications the following estimations turn out to be useful.

Lemma 3.6. (the Mill's ratio) *Let* $f_\xi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ *be the '**standard**' (normalized) normal density and* $\Phi(x)$ *the corresponding distribution function. Then for any* $x > 0$

$$\boxed{\left(1 - x^{-2}\right)\frac{f_\xi(x)}{x} < 1 - \Phi(x) < \frac{f_\xi(x)}{x}} \tag{3.42}$$

and, in particular,

$$\boxed{\lim_{x\to\infty}\frac{x\left[1-\Phi\left(x\right)\right]}{f_{\xi}\left(x\right)}=1} \tag{3.43}$$

Proof. Using the identity $\frac{d}{dx}f_{\xi}(x) = -xf_{\xi}(x)$, partial integration yields

$$0 < \int_{x}^{\infty} t^{-2} f_{\xi}(t)dt = \frac{f_{\xi}(x)}{x} - [1-\Phi(x)]$$

that proves the right-most inequality. Similarly,

$$\begin{aligned} 0 < \int_{x}^{\infty} 3t^{-4} f_{\xi}(t)dt &= \frac{f_{\xi}(x)}{x^3} - \int_{x}^{\infty} t^{-2} f_{\xi}(t)dt \\ &= \frac{f_{\xi}(x)}{x^3} - \frac{f_{\xi}(x)}{x} + [1-\Phi(x)] \end{aligned}$$

that gives the left-hand inequality. □

3.2.3.2 Resumé

Summary 3.1. *In general the mathematical expectation* $\mathsf{E}\{\xi(\omega)\}$ *can be presented by the unique formula*

$$\boxed{\mathsf{E}\{\xi(\omega)\} = \int_{-\infty}^{+\infty} x F_{\xi}(dx)} \tag{3.44}$$

where

- $$\boxed{F_{\xi}(dx) = \sum_{s} \delta(x - x_s)dx} \tag{3.45}$$

  for **discrete random variables** in the given points x_s;
- $$\boxed{F_{\xi}(dx) = \mathsf{P}\{\xi(\omega) \in (x; x+dx]\}} \tag{3.46}$$

  for **continuous random variables**;
- $$\boxed{F_{\xi}(dx) = p_{\xi}(x)dx} \tag{3.47}$$

  for **absolutely continuous random variables**.

3.3 Covariance, correlation and independence

3.3.1 Covariance

Definition 3.8. *Let ξ, η be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{E}\{\xi\}$, $\mathsf{E}\{\eta\}$ and $\mathsf{E}\{\xi\eta\}$ exist. Then the* ***covariance*** *of ξ and η is*

$$\boxed{\operatorname{cov}(\xi, \eta) := \mathsf{E}\{(\xi - \mathsf{E}\{\xi\})(\eta - \mathsf{E}\{\eta\})\}} \tag{3.48}$$

Proposition 3.1. *It is easy to check that*

1.

$$\boxed{\operatorname{cov}(\xi, \eta) = \mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\,\mathsf{E}\{\eta\}} \tag{3.49}$$

2. for any $a, b \in \mathbb{R}$

$$\boxed{\operatorname{cov}(a\xi, b\eta) = ab\operatorname{cov}(\xi, \eta)} \tag{3.50}$$

3.3.2 Correlation

Definition 3.9. *Let ξ, η be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that* $\operatorname{var}\xi \in (0, \infty)$ *and* $\operatorname{var}\eta \in (0, \infty)$. *Then the* ***correlation coefficient*** *of ξ and η is*

$$\boxed{\rho_{\xi,\eta} := \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\operatorname{var}\xi}\sqrt{\operatorname{var}\eta}}} \tag{3.51}$$

and the random variables ξ, η are said to be ***uncorrelated*** *if and only if*

$$\boxed{\rho_{\xi,\eta} = 0} \tag{3.52}$$

Proposition 3.2.

$$\boxed{-1 \le \rho_{\xi,\eta} \le 1} \tag{3.53}$$

that follows directly from the Cauchy–Bounyakovski–Schwartz inequality (see (4.16) below) if $(\xi - \mathsf{E}\{\xi\})$ is taken instead of ξ and $\mathsf{E}\{\eta - \mathsf{E}\{\eta\}\}$ is taken instead of η.

3.3.3 Relation with independence

Proposition 3.3. *If ξ, η are* ***independent****, then they are* ***uncorrelated****, that is, $\rho_{\xi,\eta} = 0$ since in this case $\mathsf{E}\{\xi\eta\} = \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}$ that follows from Fubini's theorem (see Poznyak (2008)).*

Proposition 3.4.

(a) *If ξ, η are uncorrelated then they are not obligatory independent.*
(b) ***Uncorrelation implies independence if both variables are Gaussian****.*

Proof. To prove (a) it is sufficient to construct an example. Let the pair (ξ, η) have the joint density given by

$$f_{\xi,\eta}(x, y) := \begin{cases} \pi^{-1} & \text{for } x^2 + y^2 \le 1 \\ 0 & \text{for otherwise} \end{cases}$$

from which it follows that

$$f_{\xi}(x) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x, y)\, dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \pi^{-1} dy = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & \text{for } |x| \le 1 \\ 0 & \text{for otherwise} \end{cases}$$

$$f_{\eta}(y) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x, y)\, dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \pi^{-1} dy = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2} & \text{for } |y| \le 1 \\ 0 & \text{for otherwise} \end{cases}$$

So the joint density $f_{\xi,\eta}(x, y)$ is not equal to the product of the marginal ones $f_{\xi}(x)$ and $f_{\eta}(y)$. To prove (b) let us show that this is true for Gaussian distributions. Indeed, in the general Gaussian case

$$f_{\xi,\eta}(x, y) := \frac{1}{2\pi\sqrt{\det R}} \exp\left(-\frac{1}{2}\begin{pmatrix} x - \mathsf{E}\{\xi\} \\ y - \mathsf{E}\{\eta\} \end{pmatrix}^{\mathsf{T}} R^{-1} \begin{pmatrix} x - \mathsf{E}\{\xi\} \\ y - \mathsf{E}\{\eta\} \end{pmatrix}\right)$$

where

$$\begin{aligned} R &= \mathsf{E}\left\{\begin{pmatrix} \xi - \mathsf{E}\{\xi\} \\ \eta - \mathsf{E}\{\eta\} \end{pmatrix}\begin{pmatrix} \xi - \mathsf{E}\{\xi\} \\ \eta - \mathsf{E}\{\eta\} \end{pmatrix}^{\mathsf{T}}\right\} \\ &= \mathsf{E}\left\{\begin{pmatrix} (\xi - \mathsf{E}\{\xi\})^2 & (\xi - \mathsf{E}\{\xi\})(\eta - \mathsf{E}\{\eta\}) \\ (\xi - \mathsf{E}\{\xi\})(\eta - \mathsf{E}\{\eta\}) & (\eta - \mathsf{E}\{\eta\})^2 \end{pmatrix}\right\} \\ &= \begin{pmatrix} \operatorname{var}\xi & \mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\} \\ \mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\} & \operatorname{var}\eta \end{pmatrix} \end{aligned}$$

If the random variables ξ, η are uncorrelated then $\mathsf{E}\{\xi\eta\} = \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}$ and

$$R = \begin{pmatrix} \operatorname{var}\xi & 0 \\ 0 & \operatorname{var}\eta \end{pmatrix}, \quad R^{-1} = \begin{pmatrix} \dfrac{1}{\operatorname{var}\xi} & 0 \\ 0 & \dfrac{1}{\operatorname{var}\eta} \end{pmatrix}$$

and hence,

$$
\begin{aligned}
f_{\xi,\eta}(x, y) &:= \frac{1}{2\pi\sqrt{(\operatorname{var}\xi)(\operatorname{var}\eta)}} \exp\left(-\frac{1}{2}\left[\frac{(x-\mathsf{E}\{\xi\})^2}{\operatorname{var}\xi} + \frac{(y-\mathsf{E}\{\eta\})^2}{\operatorname{var}\eta}\right]\right) \\
&= \frac{1}{\sqrt{2\pi\,(\operatorname{var}\xi)}} \exp\left(-\frac{(x-\mathsf{E}\{\xi\})^2}{2\operatorname{var}\xi}\right) \frac{1}{\sqrt{2\pi\,(\operatorname{var}\eta)}} \exp\left(-\frac{(y-\mathsf{E}\{\eta\})^2}{2\operatorname{var}\eta}\right) \\
&= f_{\xi}(x)\, f_{\eta}(y)
\end{aligned}
$$

that exactly means independence. Proposition is proven. □

4 Basic Probabilistic Inequalities

Contents

The inequalities discussed below play an important role in probability theory and are intensively used in the subsequent chapters of the book.

4.1 Moment-type inequalities

Each of the inequalities considered below is a special case of the corresponding integral inequalities from Section 16.4 in Poznyak (2008) when the Lebesgue measure within is a *probabilistic measure* (see Lemma 1.4).

4.1.1 Generalized Chebyshev inequality

Theorem 4.1. (the generalized Chebyshev inequality) *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a nonnegative non-decreasing function defined on the interval* $[0, \infty)$*, i.e.,*

$$\boxed{g(x) \geq 0 \ \forall x \in [0, \infty), \quad g(x_1) \geq g(x_2) \ \forall x_1 \geq x_2} \tag{4.1}$$

and $\xi = \xi(\omega)$ *be a random variable defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *such that*

$$\boxed{\mathsf{E}\{g(|\xi|)\} < \infty} \tag{4.2}$$

Then for any nonnegative value $a \geq 0$ *the following inequality holds:*

$$\boxed{\mathsf{E}\{g(|\xi|)\} \geq g(a)\mathsf{P}\{|\xi| \geq a\}} \tag{4.3}$$

Proof. By (3.44) and under the assumption of the theorem it follows that

$$\begin{aligned} \mathsf{E}\{g(|\xi|)\} &= \int_0^\infty g(x)\, F_{|\xi|}(dx) \geq \int_a^\infty g(x)\, F_{|\xi|}(dx) \\ &\geq \int_a^\infty g(a)\, F_{|\xi|}(dx) = g(a) \int_a^\infty F_{|\xi|}(dx) = g(a)\, \mathsf{P}\{|\xi| \geq a\} \end{aligned} \tag{4.4}$$

□

Claim 4.1. *The inequality (4.3) remains valid without the use of the module* sign *within, i.e.,*

$$\boxed{\mathsf{E}\{g(\xi)\} \geq g(a)\mathsf{P}\{\xi \geq a\}} \tag{4.5}$$

which can be proven by repeating (4.4) where the integral is taken over $(-\infty, \infty)$.

4.1.1.1 Cantelli's inequalities

Theorem 4.2. (Cantelli's inequalities) *For any random variable with bounded variance* $\operatorname{var}\xi := \mathsf{E}\left\{(\xi - \mathsf{E}\{\xi\})^2\right\} = \sigma^2$ *and any* $a \geq 0$ *it follows that*

$$\boxed{\begin{array}{c} \mathsf{P}\{|\xi - \mathsf{E}\{\xi\}| \geq a\} \leq \dfrac{2\sigma^2}{a^2 + \sigma^2} \\ \mathsf{P}\{\xi - \mathsf{E}\{\xi\} \geq a\} \leq \dfrac{\sigma^2}{a^2 + \sigma^2} \end{array}} \tag{4.6}$$

Proof. Letting in (4.3) $g(a) := a^2 + \sigma^2$ and using $[\xi - \mathsf{E}\{\xi\}]$ instead of ξ, one has

$$\begin{aligned} &\mathsf{E}\{g(|\xi|)\} = \mathsf{E}\left\{|\xi - \mathsf{E}\{\xi\}|^2 + \sigma^2\right\} = 2\sigma^2 \\ &\mathsf{P}\{\xi \geq a\} \leq \frac{\mathsf{E}\{g(\xi)\}}{g(a)} \\ &\mathsf{E}\{g(\xi)\} = \mathsf{E}\left\{(\xi - \mathsf{E}\{\xi\})^2 + \sigma^2\right\} = 2\sigma^2 \end{aligned}$$

that proves the first inequality in (4.6). The second one follows from the consideration

$$\begin{aligned} \mathsf{E}\{g(\xi)\} = 2\sigma^2 &= \int_{-\infty}^{0} g(x)\, F_\xi(dx) + \int_{0}^{\infty} g(x)\, F_\xi(dx) \\ &= \int_{0}^{\infty} g(-x)\, F_\xi(dx) + \int_{0}^{\infty} g(x)\, F_\xi(dx) = 2\int_{0}^{\infty} g(x)\, F_\xi(dx) \\ &\geq 2\int_{a}^{\infty} g(x)\, F_\xi(dx) \geq 2g(a)\mathsf{P}\{\xi \geq a\} \end{aligned}$$

that complete the proof. □

4.1.1.2 Markov and Chebyshev inequalities

Using the generalized Chebyshev inequality (4.3) one can obtain the following important and commonly used integral relations known as the *Markov* and *the Chebyshev inequalities*.

Theorem 4.3. (the Markov inequality) *Put in (4.3)*

$$\boxed{g(x) = x^r, \quad x \in [0, \infty), r > 0} \tag{4.7}$$

Then for any $a > 0$ *the inequality (4.3) becomes*

$$\boxed{\mathsf{P}\{|\xi| \geq a\} \leq a^{-r}\mathsf{E}\left\{|\xi|^r\right\}} \tag{4.8}$$

Two partial cases corresponding $r = 1, 2$ present a special interest.

Corollary 4.1. (the first Chebyshev inequality) *For* $r = 1$ *the Markov inequality (4.8) becomes*

$$\boxed{\mathsf{P}\{|\xi| \geq a\} \leq \frac{1}{a}\mathsf{E}\{|\xi|\}} \tag{4.9}$$

Corollary 4.2. (the second Chebyshev inequality) *For* $r = 2$ *the Markov inequality (4.8) becomes*

$$\boxed{\mathsf{P}\{|\xi| \geq a\} \leq \frac{1}{a^2}\mathsf{E}\left\{\xi^2\right\}} \tag{4.10}$$

4.1.2 Hölder inequality

Theorem 4.4. (the Hölder inequality) *Let* p *and* q *be positive values such that*

$$\boxed{p > 1, q > 1, \quad p^{-1} + q^{-1} = 1} \tag{4.11}$$

and ξ, η *be random variables defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *such that*

$$\boxed{\mathsf{E}\left\{|\xi|^p\right\} < \infty, \quad \mathsf{E}\left\{|\eta|^q\right\} < \infty} \tag{4.12}$$

Then the following inequality holds:

$$\boxed{\mathsf{E}\{|\xi\eta|\} \leq \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\eta|^q\right\}\right)^{1/q}} \tag{4.13}$$

Proof. If $\mathsf{E}\left\{|\xi|^p\right\} = \mathsf{E}\left\{|\eta|^q\right\} = 0$ then $\xi = \eta = 0$ almost everywhere and (13.73) looks trivial. Suppose that both $\mathsf{E}\left\{|\xi|^p\right\} > 0$ and $\int_{\mathcal{E}} \mathsf{E}\left\{|\eta|^q\right\} > 0$. Since the function $\ln(x)$ is concave for any $x, y, a, b > 0$ the following inequality holds ($a + b = 1$):

$$\boxed{\ln(ax + by) \geq a\ln(x) + b\ln(y)} \tag{4.14}$$

or, equivalently,

$$\boxed{ax + by \geq x^a y^b} \tag{4.15}$$

Taking $a := 1/p, b := 1/q$ and $x := \dfrac{|\xi|^p}{\mathsf{E}\{|\xi|^p\}}$, $y := \dfrac{|\eta|^p}{\mathsf{E}\{|\eta|^p\}}$ implies

$$1/p \frac{|\xi|^p}{\mathsf{E}\{|\xi|^p\}} + 1/q \frac{|\eta|^p}{\mathsf{E}\{|\eta|^p\}} \geq \frac{|\xi|}{\left(\mathsf{E}\{|\xi|^p\}\right)^{1/p}} \frac{|\eta|}{\left(\mathsf{E}\{|\eta|^p\}\right)^{1/q}}$$

Applying the operator $\mathsf{E}\{\cdot\}$ to both sides of this inequality and using the assumption that $p^{-1} + q^{-1} = 1$ proves (13.73). □

4.1.3 Cauchy–Bounyakovski–Schwartz inequality

The following particular case $p = q = 2$ of (13.73) is the most common in use.

Corollary 4.3. (The CBS inequality)

$$\boxed{\mathsf{E}\{|\xi\eta|\} \leq \sqrt{\mathsf{E}\{|\xi|^2\}}\sqrt{\mathsf{E}\{|\eta|^2\}}} \tag{4.16}$$

and the equality in (4.16) is reached if

$$\boxed{\xi(\omega) = k\eta(\omega) \text{ for any real } k} \tag{4.17}$$

and almost all $\omega \in \Omega$.

Proof. To prove (4.17) it is sufficient to substitute $\xi(\omega) = k\eta(\omega)$ in to (4.16). □

4.1.4 Jensen inequality

Theorem 4.5. (the Jensen inequality) *Let* $g_{\cup} : \mathbb{R} \to \mathbb{R}$ *and* $g_{\cap} : \mathbb{R} \to \mathbb{R}$ *be convex downward (or, simply,* ***convex****) and convex upward (or, simply,* ***concave****), respectively, and* ξ *be randomly variable defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *such that*

$$\boxed{\max\{g_{\cap}(\mathsf{E}\{\xi\}),\ \mathsf{E}\{g_{\cup}(\xi)\}\} < \infty} \tag{4.18}$$

Then

$$\boxed{g_{\cup}(\mathsf{E}\{\xi\}) \leq \mathsf{E}\{g_{\cup}(\xi)\}} \tag{4.19}$$

and

$$\boxed{g_{\cap}(\mathsf{E}\{\xi\}) \geq \mathsf{E}\{g_{\cap}(\xi)\}} \tag{4.20}$$

Proof. By the convexity (concavity) definition we may conclude that in both convexity and concavity cases there exists a number $\lambda(x_0)$ such that for any $x, x_0 \in \mathbb{R}$ the following inequalities are fulfilled:

$$\begin{aligned} g_{\cup}(x) &\geq g_{\cup}(x_0) + \lambda(x_0)(x - x_0) \\ g_{\cap}(x) &\leq g_{\cap}(x_0) + \lambda(x_0)(x - x_0) \end{aligned} \tag{4.21}$$

Taking $x := \xi$, $x_0 := \mathsf{E}\{\xi\}$ in (4.21) we obtain

$$g_{\cup}(\xi) \geq g_{\cup}(\mathsf{E}\{\xi\}) + \lambda(\mathsf{E}\{\xi\})(\xi - \mathsf{E}\{\xi\})$$

$$g_{\cap}(\xi) \leq g_{\cap}(\mathsf{E}\{\xi\}) + \lambda(\mathsf{E}\{\xi\}\,(\xi - \mathsf{E}\{\xi\})$$

The application of $\mathsf{E}\{\cdot\}$ to both sides of these inequalities leads to (4.19) and (4.20), respectively. Theorem is proven. □

Example 4.1. *For $g_{\cap}(x) := \ln(|x|)$ one has*

$$\boxed{\ln(\mathsf{E}\{|\xi|\}) \geq \mathsf{E}\{\ln(|\xi|)\}} \tag{4.22}$$

4.1.5 Lyapunov inequality

The inequality below is a particular case of the Jensen inequality (4.19).

Corollary 4.4. (the Lyapunov inequality) *For a random variable ξ defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{E}\left\{|\xi|^t\right\} < \infty$ $(t > 0)$ the following inequality holds:*

$$\boxed{\left(\mathsf{E}\left\{|\xi|^s\right\}\right)^{1/s} \leq \left(\mathsf{E}\left\{|\xi|^t\right\}\right)^{1/t}} \tag{4.23}$$

where $0 < s \leq t$.

Proof. Define $r := \dfrac{t}{s}$. Taking in (4.19) $\xi := |\xi|^s$ and $g_{\cup}(x) := |x|^r$ implies

$$\left(\mathsf{E}\left\{|\xi|^s\right\}\right)^{t/s} = \left(\mathsf{E}\left\{|\xi|^s\right\}\right)^r \leq \mathsf{E}\left\{\left(|\xi|^s\right)^r\right\} = \mathsf{E}\left\{|\xi|^t\right\}$$

that completes the proof. □

Corollary 4.5. *For any random variable ξ defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{E}\left\{|\xi|^k\right\} < \infty$ ($k > 2$ is integer) the following inequalities hold:*

$$\boxed{\mathsf{E}\{|\xi|\} \leq \left(\mathsf{E}\left\{|\xi|^2\right\}\right)^{1/2} \leq \cdots \leq \left(\mathsf{E}\left\{|\xi|^k\right\}\right)^{1/k}} \tag{4.24}$$

4.1.6 Kulbac inequality

Theorem 4.6. (the continuous version) *Suppose $p : \mathbb{R} \to \mathbb{R}$ and $q : \mathbb{R} \to \mathbb{R}$ be any positive density functions given on $\mathcal{E} \subset \mathbb{R}$ such that the Lebesgue integral*

$$\boxed{I_{\mathcal{E}}(p,q) := \int\limits_{\mathcal{E}} \ln\left(\frac{p(x)}{q(x)}\right) p(x)dx} \tag{4.25}$$

is finite, that is, $I_{\mathcal{E}}(p,q) < \infty$. Then

$$\boxed{I_{\mathcal{E}}(p,q) \geq 0} \tag{4.26}$$

and $I_{\mathcal{E}}(p,q) = 0$ if and only if $p(x) = q(x)$ almost everywhere on $\mathcal{E}$.

Proof. Notice that $(-\ln(x))$ is a convex function on $(0,\infty)$, i.e., $-\ln(x) = g_{\cup}(x)$. Hence, by the Jensen inequality (4.19) we have

$$\begin{aligned} I_{\mathcal{E}}(p,q) &= \int\limits_{\mathcal{E}} \ln\left(\frac{p(x)}{q(x)}\right) p(x)dx = \int\limits_{\mathcal{E}} \ln\left(-\left(\frac{q(x)}{p(x)}\right)\right) p(x)dx \\ &\geq -\ln \int\limits_{\mathcal{E}} \left(\frac{q(x)}{p(x)}\right) p(x)dx = -\ln \int\limits_{\mathcal{E}} q(x)dx = -\ln 1 = 0 \end{aligned}$$

that proves (4.26). Evidently, $I_{\mathcal{E}}(p,q) = 0$ if $p(x) = q(x)$ almost everywhere on $\mathcal{E}$. Suppose $I_{\mathcal{E}}(p,q) = 0$ and $p(x) \neq q(x)$ for some $x \in \mathcal{E}_0 \subset \mathcal{E}$ such that $\mu(\mathcal{E}_0) = \int\limits_{\mathcal{E}_0} dx > 0$. Then the Jensen inequality (4.19) implies

$$\begin{aligned} 0 = I_{\mathcal{E}}(p,q) &= -\int\limits_{\mathcal{E}_0} \ln\left(\frac{q(x)}{p(x)}\right) p(x)dx \\ &\geq -\ln\left(\int\limits_{\mathcal{E}_0} \left(\frac{q(x)}{p(x)}\right) p(x)\right) dx = -\ln\left(\int\limits_{\mathcal{E}_0} q(x)dx\right) = -\ln\alpha > 0 \end{aligned}$$

where $\alpha := \int\limits_{\mathcal{E}_0} q(x)dx < 1$ which always can be done selecting $\mathcal{E}_0$ to be small enough. The last inequality represents a contradiction. So, $\mu(\mathcal{E}_0) = 0$. Theorem is proven. □

4.1.7 Minkowski inequality

Theorem 4.7. (the Minkovski inequality) *Suppose ξ, η be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that*

$$\boxed{\mathsf{E}\left\{|\xi|^p\right\} < \infty, \quad \mathsf{E}\left\{|\eta|^p\right\} < \infty} \tag{4.27}$$

for some $p \in [1, \infty)$. *Then the following inequality holds:*

$$\boxed{\left(\mathsf{E}\left\{|\xi + \eta|^p\right\}\right)^{1/p} \le \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p} + \left(\mathsf{E}\left\{|\eta|^p\right\}\right)^{1/p}} \tag{4.28}$$

Proof. Consider the following inequality

$$|\xi + \eta|^p = |\xi + \eta|\,|\xi + \eta|^{p-1} \le |\xi|\,|\xi + \eta|^{p-1} + |\eta|\,|\xi + \eta|^{p-1}$$

which after the application $\mathsf{E}\{\cdot\}$ becomes

$$\mathsf{E}\left\{|\xi + \eta|^p\right\} \le \mathsf{E}\left\{|\xi|\,|\xi + \eta|^{p-1}\right\} + \mathsf{E}\left\{|\eta|\,|\xi + \eta|^{p-1}\right\} \tag{4.29}$$

Applying the Hölder inequality (13.73) to each term in the right-hand side of (4.29) we derive:

$$\begin{aligned}\mathsf{E}\left\{|\xi|\,|\xi + \eta|^{p-1}\right\} &\le \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\xi + \eta|^{(p-1)q}\right\}\right)^{1/q}\\ &= \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\xi + \eta|^{p}\right\}\right)^{1/q}\end{aligned}$$

since $p = (p-1)q$, and

$$\begin{aligned}\mathsf{E}\left\{|\eta|\,|\xi + \eta|^{p-1}\right\} &\le \left(\mathsf{E}\left\{|\eta|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\xi + \eta|^{(p-1)q}\right\}\right)^{1/q}\\ &= \left(\mathsf{E}\left\{|\eta|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\xi + \eta|^{p}\right\}\right)^{1/q}\end{aligned}$$

Using these inequalities for the right-hand side estimation in (4.29) we get

$$\mathsf{E}\left\{|\xi + \eta|^p\right\} \le \left[\left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p} + \left(\mathsf{E}\left\{|\eta|^p\right\}\right)^{1/p}\right]\left(\mathsf{E}\left\{|\xi + \eta|^p\right\}\right)^{1/q}$$

that implies

$$\begin{aligned}\left(\mathsf{E}\left\{|\xi + \eta|^p\right\}\right)^{1-1/q} &= \left(\mathsf{E}\left\{|\xi + \eta|^p\right\}\right)^{1/p}\\ &\le \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p} + \left(\mathsf{E}\left\{|\eta|^p\right\}\right)^{1/p}\end{aligned}$$

Theorem is proven. □

4.1.8 r-Moment inequality

First, prove the next simple lemma.

Lemma 4.1. *For* $x, y \ge 0$ *and* $r > 0$ *the following inequality holds:*

$$\boxed{(x + y)^r \le \begin{cases} 2^r\,(x^r + y^r) & \text{for } \quad r > 0\\ x^r + y^r & \text{for } \ 0 < r \le 1\\ 2^{r-1}\,(x^r + y^r) & \text{for } \quad r \ge 1\end{cases}} \tag{4.30}$$

Proof. The general case ($r > 0$) follows from the following chain of evident inequalities:

$$(x + y)^r \le (2 \max\{x, y\})^r = 2^r (\max\{x, y\})^r \le 2^r \left(x^r + y^r\right)$$

For $r \in (0, 1]$ in view of the inequality $x^{1/r} \le x$, valid for $x \in [0, 1]$, we have

$$\left(\frac{x^r}{x^r + y^r}\right)^{1/r} + \left(\frac{y^r}{x^r + y^r}\right)^{1/r} \le \frac{x^r}{x^r + y^r} + \frac{y^r}{x^r + y^r} = 1$$

or, equivalently,

$$x + y \le \left(x^r + y^r\right)^{1/r}$$

that implies the desired result. For $r \ge 1$ in view of the convexity of the function $|x|^r$ it follows that

$$\left(\frac{x + y}{2}\right)^r \le \frac{1}{2}x^r + \frac{1}{2}y^r$$

that implies the third inequality in (4.30). □

Theorem 4.8. (on the r-moments of a sum) *Let ξ and η be random variables with bounded r-moments, that is, $\mathsf{E}\left\{|\xi|^r\right\} < \infty$ and $\mathsf{E}\left\{|\eta|^r\right\} < \infty$ for some $r > 0$. Then*

$$\boxed{\mathsf{E}\left\{|\xi + \eta|^r\right\} \le \begin{cases} 2^r \left(\mathsf{E}\left\{|\xi|^r\right\} + \mathsf{E}\left\{|\eta|^r\right\}\right) & \text{for} \quad r > 0 \\ \mathsf{E}\left\{|\xi|^r\right\} + \mathsf{E}\left\{|\eta|^r\right\} & \text{for} \quad 0 < r \le 1 \\ 2^{r-1} \left(\mathsf{E}\left\{|\xi|^r\right\} + \mathsf{E}\left\{|\eta|^r\right\}\right) & \text{for} \quad r \ge 1 \end{cases}} \tag{4.31}$$

Proof. By the triangle inequality

$$\mathsf{E}\left\{|\xi + \eta|^r\right\} \le \mathsf{E}\left\{(|\xi| + |\eta|)^r\right\}$$

and, letting in (4.30) $x := |\xi|$, $y := |\eta|$ and taking then the mathematical expectation from both sides of (4.30) we get (4.31). □

4.1.9 Exponential inequalities

The theorem given below deals with bounded random variables.

Theorem 4.9. (on exponential estimates (Gut, 2005))

1. *Suppose that $\mathsf{P}\{|\xi| \le b\} = 1$ for some $b > 0$, $\mathsf{E}\{\xi\} = 0$ and $\operatorname{var}\xi = \mathsf{E}\{\xi^2\} = \sigma^2$. Then for any $t \in \left(0, b^{-1}\right)$ and $x > 0$ the following inequalities hold:*

$$\boxed{\begin{gathered} \mathsf{P}\{\xi > x\} \le \exp\left\{-tx + t^2\sigma^2\right\} \\ \mathsf{P}\{|\xi| > x\} \le 2\exp\left\{-tx + t^2\sigma^2\right\} \end{gathered}} \tag{4.32}$$

2. *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables with*

$$\mathsf{E}\{\xi_k\} = 0, \quad \operatorname{var}\xi_k = \mathsf{E}\{\xi_k^2\} = \sigma_k^2$$

and such that $\mathsf{P}\{|\xi_k| \le b\} = 1$ for all $k = 1, \ldots, n$, then for any $t \in (0, b^{-1})$, $x > 0$ and $S_n := \sum_{k=1}^n \xi_k$:

$$\boxed{\begin{aligned} \mathsf{P}\{S_n > x\} &\le \exp\left\{-tx + t^2 \sum_{k=1}^n \sigma_k^2\right\} \\ \mathsf{P}\{|S_n| > x\} &\le 2\exp\left\{-tx + t^2 \sum_{k=1}^n \sigma_k^2\right\} \end{aligned}} \tag{4.33}$$

3. *If in the previous item 2, in addition, $\xi_1, \xi_2, \ldots, \xi_n$ are identically distributed $(\sigma_k^2 = \sigma^2)$ then*

$$\boxed{\begin{aligned} \mathsf{P}\{S_n > x\} &\le \exp\left\{-tx + nt^2\sigma^2\right\} \\ \mathsf{P}\{|S_n| > x\} &\le 2\exp\left\{-tx + nt^2\sigma^2\right\} \end{aligned}} \tag{4.34}$$

Proof.

1. Applying the inequality (4.5) for $g(x) = e^{tx}$ and using the simple estimate (for $|x| \le 1$)

$$\begin{aligned} e^x &= 1 + x + x^2 \sum_{k=2}^{\infty} \frac{x^{k-2}}{k!} \le 1 + x + x^2 \sum_{k=2}^{\infty} \frac{1}{k!} \\ &= 1 + x + x^2 (e - 2) \le 1 + x + x^2 \end{aligned}$$

one has

$$\begin{aligned} \mathsf{P}\{\xi > x\} &\le \frac{\mathsf{E}\{e^{t\xi}\}}{e^{tx}} \le e^{-tx}\left(1 + t\mathsf{E}\{\xi\} + t^2\mathsf{E}\{\xi^2\}\right) \\ &= e^{-tx}\left(1 + t^2\mathsf{E}\{\xi^2\}\right) = e^{-tx}\left(1 + t^2\sigma^2\right) \end{aligned}$$

Finally, in view of the inequality $e^x \ge 1 + x$, it follows that

$$\mathsf{P}\{\xi > x\} \le e^{-tx}\left(1 + t^2\sigma^2\right) \le e^{-tx}e^{t^2\sigma^2} = e^{-tx + t^2\sigma^2}$$

that proves the first inequality in (4.32). The second inequality follows from the consideration

$$\begin{aligned} \mathsf{P}\{|\xi| > x\} &= \mathsf{P}\{(\xi > x) \cap (\xi \ge 0) \cup (\xi < -x) \cap (\xi < 0)\} \\ &\le \mathsf{P}\{(\xi > x) \cap (\xi \ge 0)\} + \mathsf{P}\{(\xi < -x) \cap (\xi < 0)\} \\ &= 2\mathsf{P}\{(\xi > x) \cap (\xi \ge 0)\} \le 2\mathsf{P}\{\xi > x\} \end{aligned} \tag{4.35}$$

2. Changing ξ by S_n in (4.32) and taking into account that all $\xi_1, \xi_2, \ldots, \xi_n$ are independent, we derive that $\operatorname{var} S_n = \mathsf{E}\{S_n^2\} = \sum_{k=1}^n \sigma_k^2$ that implies (4.33).

3. The inequalities (4.34) follow from (4.33) if we take $\sigma_k^2 = \sigma^2$ $(k = 1, \ldots, n)$. □

4.2 Probability inequalities for maxima of partial sums

4.2.1 Classical Kolmogorov-type inequalities

Theorem 4.10. (the Kolmogorov inequality) *Let $\xi_1, \xi_2, \ldots, \xi_n$ be* ***independent*** *random variables with*

$$\boxed{\mathsf{E}\{\xi_k\} = 0, \quad \operatorname{var} \xi_k = \mathsf{E}\{\xi_k^2\} = \sigma_k^2} \tag{4.36}$$

for all $k = 1, \ldots, n$ and

$$S_k := \sum_{s=1}^{k} \xi_s \tag{4.37}$$

be the ***partial sum****. Then for any $x > 0$*

$$\boxed{\mathsf{P}\left\{\max_{1 \le k \le n} |S_k| > x\right\} \le x^{-2} \sum_{k=1}^{n} \operatorname{var} \xi_k} \tag{4.38}$$

In particular, if $\xi_1, \xi_2, \ldots, \xi_n$ are identically distributed then

$$\boxed{\mathsf{P}\left\{\max_{1 \le k \le n} |S_k| > x\right\} \le x^{-2} n \operatorname{var} \xi_1} \tag{4.39}$$

Proof. Define the sets

$$A_k := \left\{\omega \in \Omega \mid \max_{1 \le s \le k-1} |S_s| \le x, \ |S_k| > x\right\} \quad (k = 1, \ldots, n) \tag{4.40}$$

Notice that

$$\left\{\omega \in \Omega \mid \max_{1 \le k \le n} |S_k| > x\right\} = \bigcup_{k=1}^{n} A_k$$

$$A_k \bigcap_{k \ne j} A_j = \varnothing \quad (k, j = 1, \ldots, n)$$

So, using disjoining of $\{A_k\}$, it follows that

$$\sum_{k=1}^{n} \operatorname{var} \xi_k = \mathsf{E}\{S_n^2\} \ge \mathsf{E}\left\{S_n^2 \sum_{k=1}^{n} \chi(A_k)\right\} = \sum_{k=1}^{n} \mathsf{E}\left\{S_n^2 \chi(A_k)\right\}$$
$$= \sum_{k=1}^{n} \mathsf{E}\left\{\left[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] \chi(A_k)\right\}$$

$$\begin{aligned}
&\geq \sum_{k=1}^{n} \mathsf{E}\left\{\left[S_k^2 + 2S_k\left(S_n - S_k\right)\right]\chi\left(A_k\right)\right\} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{S_k^2\chi\left(A_k\right)\right\} + 2\sum_{k=1}^{n} \mathsf{E}\left\{S_k\left(S_n - S_k\right)\chi\left(A_k\right)\right\} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{S_k^2\chi\left(A_k\right)\right\} + 2\sum_{k=1}^{n} \mathsf{E}\left\{\chi\left(A_k\right)S_k\right\}\underbrace{E\left\{\left(S_n - S_k\right)\right\}}_{0} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{S_k^2\chi\left(A_k\right)\right\} \geq \sum_{k=1}^{n} \mathsf{E}\left\{x^2\chi\left(A_k\right)\right\} \\
&\geq x^2\sum_{k=1}^{n} \mathsf{E}\left\{\chi\left(A_k\right)\right\} = x^2\sum_{k=1}^{n} \mathsf{P}\{A_k\} \\
&= x^2\mathsf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} = x^2\mathsf{P}\left\{\max_{1\leq k\leq n}\left|S_k\right| > x\right\}
\end{aligned}$$

that directly leads to (4.38). The inequality (4.39) is a trivial consequence of (4.38). □

4.2.2 Hájek–Renyi inequality

The following result is a generalization of the Kolmogorov theorem (4.10).

Theorem 4.11. **(Hájek and Renyi, 1955)** *If in the Kolmogorov theorem (4.10) we additionally consider a set $\{c_k, k = 1, \ldots, n\}$ of positive, non-increasing real numbers, then for any $x > 0$*

$$\boxed{\mathsf{P}\left\{\max_{1\leq k\leq n} c_k\left|S_k\right| > x\right\} \leq x^{-2}\sum_{k=1}^{n} c_k^2 \operatorname{var}\xi_k} \tag{4.41}$$

Proof. Notice that for monotonic sequence $\{c_k, k = 1, \ldots, n\}$

$$\max_{1\leq k\leq n} c_k\left|S_k\right| = \max_{1\leq k\leq n}\left|\sum_{k=1}^{n} c_k\xi_k\right| = \max_{1\leq k\leq n}\left|\sum_{k=1}^{n}\tilde{\xi}_k\right|$$

where $\tilde{\xi}_k := c_k\xi_k$. Then (4.41) follows directly from (4.38) being applied to $\tilde{S}_k := \sum_{r=1}^{k}\tilde{\xi}_r$ if we take into account that $\operatorname{var}\tilde{\xi}_k = c_k^2\operatorname{var}\xi_k$. □

4.2.3 Relation of a maxima of partial sum probabilities with a distribution of the last partial sum

The following results state the relation between tail probabilities and the distribution of the last partial sum.

Theorem 4.12. **(Lévy, 1937)** *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables, then for any $x \in \mathbb{R}$*

$$\boxed{\begin{array}{l} \mathsf{P}\left\{\max_{1\le k\le n}\left(S_k - \operatorname{med}(S_k - S_n)\right) > x\right\} \le 2\mathsf{P}\{S_n > x\} \\ \mathsf{P}\left\{\max_{1\le k\le n}\left|S_k - \operatorname{med}(S_k - S_n)\right| > x\right\} \le 2\mathsf{P}\{|S_n| > x\} \end{array}} \tag{4.42}$$

where S_k and med (ξ) *are defined in (4.37) and (3.22), respectively.*

Proof. Define sets

$$A_k := \left\{\max_{1\le j\le k-1}\left(S_j - \operatorname{med}\left(S_j - S_n\right)\right) \le x,\ S_k - \operatorname{med}(S_k - S_n) > x\right\}$$
$$B_k := \{S_n - S_k - \operatorname{med}(S_n - S_k) \ge 0\}$$

Notice that $\{A_k\}$ are disjoint, A_k and B_k (they are independent since they contain no common summand), $\mathsf{P}\{B_k\} \ge 1/2$ and $\{S_n > x\} \supset \bigcup_{k=1}^{n}\{A_k \cap B_k\}$. Hence

$$\begin{aligned} \mathsf{P}\{S_n > x\} &\ge \sum_{k=1}^{n}\mathsf{P}\{A_k \cap B_k\} = \sum_{k=1}^{n}\mathsf{P}\{A_k\}\mathsf{P}\{B_k\} \\ &\ge \sum_{k=1}^{n}\mathsf{P}\{A_k\}\frac{1}{2} = \frac{1}{2}\mathsf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} \\ &= \frac{1}{2}\mathsf{P}\left\{\max_{1\le k\le n}\left(S_k - \operatorname{med}(S_k - S_n)\right) > x\right\} \end{aligned}$$

that proves the first assertion in (4.42). The second one follows by considering the other tail and addition. □

Corollary 4.6. *For the independent random variables ξ_k $(1 \le k \le n)$ with symmetric distributions (that implies* $\operatorname{med}(S_k - S_n) = 0$*) it follows that*

$$\boxed{\begin{array}{l} \mathsf{P}\left\{\max_{1\le k\le n} S_k > x\right\} \le 2\mathsf{P}\{S_n > x\} \\ \mathsf{P}\left\{\max_{1\le k\le n} |S_k| > x\right\} \le 2\mathsf{P}\{|S_n| > x\} \end{array}} \tag{4.43}$$

and also

$$\boxed{\begin{array}{l} \mathsf{P}\{S_k > x\} \le 2\mathsf{P}\{S_n > x\} \\ \mathsf{P}\{|S_k| > x\} \le 2\mathsf{P}\{|S_n| > x\} \end{array}} \tag{4.44}$$

since $\{\max_{1\le k\le n} S_k > x\} \supset \{S_k > x\}$ *and* $\{\max_{1\le k\le n} |S_k| > x\} \supset \{|S_k| > x\}$.

Another efficient inequality, concerning the same class of estimates, was given by Kahane (1985) and extended by Hoffman-Jörgensen (1974) and Jain (1975).

Theorem 4.13. (KHJ inequality) *For the independent random variables* ξ_k $(1 \le k \le n)$ *with symmetric distributions and any* $x, y > 0$ *it follows that*

$$\boxed{\begin{aligned} \mathsf{P}\{|S_n| > 2x + y\} &\le \mathsf{P}\left\{\max_{1\le k\le n} |\xi_k| > y\right\} + 4\,[\mathsf{P}\{|S_n| > x\}]^2 \\ &\le \sum_{k=1}^{n} \mathsf{P}\{|\xi_k| > y\} + 4\,[\mathsf{P}\{|S_n| > x\}]^2 \end{aligned}} \tag{4.45}$$

Proof. Define $\zeta_n := \max\limits_{1\le k\le n} |\xi_k|$ and, as before, define the sets

$$A_k := \left\{\max_{1\le j\le k-1} \left|S_j\right| \le x, |S_k| > x\right\}, \quad k = 1, \dots, n$$

which are obviously disjoint. Since $\{|S_n| > 2x + y\} \subset \bigcup_{k=1}^{n} A_k$, it follows that

$$\begin{aligned} \mathsf{P}\{|S_n| > 2x + y\} &= \mathsf{P}\left\{\{|S_n| > 2x + y\} \cap \left\{\bigcup_{k=1}^{n} A_k\right\}\right\} \\ &= \sum_{k=1}^{n} \mathsf{P}\{\{|S_n| > 2x + y\} \cap A_k\} \end{aligned} \tag{4.46}$$

By the triangle inequality for any $k = 1, \dots, n$ one has

$$|S_k| = |S_{k-1} + \xi_k| \le |S_{k-1}| + |\xi_k| \le |S_{k-1}| + |\xi_k| + |S_n - S_k|$$

that on the set $\{\{|S_n| > 2x + y\} \cap A_k\}$ gives

$$|S_n - S_k| \ge |S_k| - |S_{k-1}| - |\xi_k| > 2x + y - x - \zeta_n = x + y - \zeta_n$$

The independence of $S_n - S_k$ and A_k implies

$$\begin{aligned} \mathsf{P}\{\{|S_n| > 2x + y\} \cap A_k\} &\le \mathsf{P}\{\{|S_n - S_k| > x + y - \zeta_n\} \cap A_k\} \\ &= \mathsf{P}\{\{|S_n - S_k| > x + y - \zeta_n\} \cap A_k \cap \{\zeta_n > y\}\} \\ &\quad + \mathsf{P}\{\{|S_n - S_k| > x + y - \zeta_n\} \cap A_k \cap \{\zeta_n \le y\}\} \\ &\le \mathsf{P}\{A_k \cap \{\zeta_n > y\}\} + \mathsf{P}\{\{|S_n - S_k| > x\} \cap A_k\} \\ &= \mathsf{P}\{A_k \cap \{\zeta_n > y\}\} + \mathsf{P}\{|S_n - S_k| > x\}\mathsf{P}\{A_k\} \end{aligned}$$

Applying then to the second term in the right-hand side the Lévy inequality (4.44) we derive

$$\mathsf{P}\{\{|S_n| > 2x + y\} \cap A_k\} \le \mathsf{P}\{A_k \cap \{\zeta_n > y\}\} + 2\mathsf{P}\{|S_n| > x\}\mathsf{P}\{A_k\}$$

Joining this inequality with (4.46) and in view of the Lévy inequality (4.43) yields

$$\begin{aligned}
\mathsf{P}\{|S_n| > 2x + y\} &\le \sum_{k=1}^{n} \mathsf{P}\{A_k \cap \{\zeta_n > y\}\} + 2\mathsf{P}\{|S_n| > x\} \sum_{k=1}^{n} \mathsf{P}\{A_k\} \\
&= \mathsf{P}\left\{\bigcup_{k=1}^{n} A_k \cap \{\zeta_n > y\}\right\} + 2\mathsf{P}\{|S_n| > x\}\mathsf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} \\
&\le \mathsf{P}\{\zeta_n > y\} + 2\mathsf{P}\{|S_n| > x\}\mathsf{P}\left\{\max_{1\le k\le n} |S_k| > x\right\} \\
&\le \mathsf{P}\{\zeta_n > y\} + 4\left[\mathsf{P}\{|S_n| > x\}\right]^2
\end{aligned}$$

The relation (4.45) follows from the last inequality if we take into account which $\mathsf{P}\{\zeta_n > y\} \le \sum_{k=1}^{n} \mathsf{P}\{|\xi_k| > y\}$ which completes the proof. □

Corollary 4.7. *In particular, for identically distributed ξ_k $(1 \le k \le n)$ (when $x = y$)*

$$\boxed{\mathsf{P}\{|S_n| > 3x\} \le n\mathsf{P}\{|\xi_1| > x\} + 4\left[\mathsf{P}\{|S_n| > x\}\right]^2} \tag{4.47}$$

Proof. Taking $x = y$ in (4.45) implies the desired result. □

Corollary 4.8. *Under the conditions of Theorem 4.13,*

$$\boxed{\begin{aligned}
\mathsf{P}\{|S_n| > 2x + y\} &\le 2\mathsf{P}\left\{\max_{1\le k\le n} |\xi_k| > y\right\} + 8\left[\mathsf{P}\{|S_n| > x\}\right]^2 \\
&\le 2\sum_{k=1}^{n} \mathsf{P}\{|\xi_k| > y\} + 8\left[\mathsf{P}\{|S_n| > x\}\right]^2
\end{aligned}} \tag{4.48}$$

Proof. It follows directly from (4.45) if we again apply the Lévy inequality (4.43) to the second term. □

Theorem 4.14. **(Jain, 1975)** *Under the conditions of Theorem 4.13 for any integer $j \ge 1$*

$$\boxed{\mathsf{P}\{|S_n| > 3^j x\} \le C_j \mathsf{P}\left\{\max_{1\le k\le n} |\xi_k| > x\right\} + D_j\left[\mathsf{P}\{|S_n| > x\}\right]^{2^j}} \tag{4.49}$$

where C_j, D_j are positive constants depending only on j. Particularly, for identically distributed ξ_k $(1 \le k \le n)$

$$\boxed{\mathsf{P}\{|S_n| > 3^j x\} \le C_j n\mathsf{P}\{|\xi_1| > y\} + D_j\left[\mathsf{P}\{|S_n| > x\}\right]^{2^j}} \tag{4.50}$$

Proof. It follows by induction if we take into account that iterating (4.45) implies (4.50) with $C_2 = 9$ and $D_2 = 128$. Continuing the same procedure for arbitrary $j > 2$ proves the final result. □

The next result is the avoidance of the symmetry assumption.

Theorem 4.15. **(Etemadi, 1981)** *If* ξ_k $(1 \le k \le n)$ *are independent random variables then for all* $x > 0$

$$\boxed{\mathsf{P}\left\{\max_{1\le k\le n}|S_k| > 3x\right\} \le 3\max_{1\le k\le n}\mathsf{P}\{|S_k| > x\}} \tag{4.51}$$

Proof. The proof follows the ideas of the previous Theorem 4.13. Again, analogously define the sets

$$A_k := \left\{\max_{1\le j\le k-1}\left|S_j\right| \le 3x,\ |S_k| > 3x\right\}, \quad k = 1,\dots,n$$

As before, A_k are disjoint, but now $\bigcup_{k=1}^{n} A_k = \{\max_{1\le k\le n}|S_k| > 3x\}$. In view of the fact that

$$\begin{aligned}
\mathsf{P}\left\{\max_{1\le k\le n}|S_k| > 3x\right\} &= \mathsf{P}\left\{\left\{\max_{1\le k\le n}|S_k| > 3x\right\}\cap\{|S_n| > x\}\right\} \\
&\quad + \mathsf{P}\left\{\left\{\max_{1\le k\le n}|S_k| > 3x\right\}\cap\{|S_n| \le x\}\right\} \\
&\le \mathsf{P}\{|S_n| > x\} + \sum_{k=1}^{n}\mathsf{P}\{A_k\cap\{|S_n - S_k| > 2x\}\} \\
&\le \mathsf{P}\{|S_n| > x\} + \sum_{k=1}^{n}\mathsf{P}\{A_k\}\mathsf{P}\{|S_n - S_k| > 2x\}
\end{aligned}$$

This implies

$$\begin{aligned}
\mathsf{P}\left\{\max_{1\le k\le n}|S_k| > 3x\right\} &\le \mathsf{P}\{|S_n| > x\} + \max_{1\le k\le n}\mathsf{P}\{\{|S_n - S_k| > 2x\}\}\mathsf{P}\left\{\bigcup_{k=1}^{n}A_k\right\} \\
&\le \mathsf{P}\{|S_n| > x\} + \max_{1\le k\le n}\mathsf{P}\{\{|S_n - S_k| > 2x\}\} \\
&\le \mathsf{P}\{|S_n| > x\} + \max_{1\le k\le n}\left(\mathsf{P}\{\{|S_n| > x\}\} + \mathsf{P}\{|S_k| > x\}\right) \\
&= 2\mathsf{P}\{|S_n| > x\} + \max_{1\le k\le n}\left(\mathsf{P}\{\{|S_k| > x\}\}\right) \\
&\le 3\max_{1\le k\le n}\left(\mathsf{P}\{\{|S_k| > x\}\}\right)
\end{aligned}$$

which completes the proof. □

The following inequality (see Skorohod, 1956) also seems to be useful.

Theorem 4.16. **(the Skorohod–Ottaviany inequality)** *For independent random variables* ξ_k $(1 \le k \le n)$ *and for all* $x, y > 0$

$$\boxed{\mathsf{P}\left\{\max_{1\le k\le n}|S_k| > x + y\right\} \le \frac{1}{1-\beta}\mathsf{P}\{|S_n| > x\}} \tag{4.52}$$

where

$$\boxed{\beta := \max_{1\le k\le n} \mathsf{P}\{|S_n - S_k| > y\}} \tag{4.53}$$

Proof. Define

$$A_k := \left\{ \max_{1\le j\le k-1} \left|S_j\right| \le x + y,\ |S_k| > x + y \right\}, \quad k = 1, \ldots, n$$

So, as in the proof above and in view of (4.53) it follows that

$$\begin{aligned}
\mathsf{P}\{|S_n| > x\} &= \sum_{k=1}^{n} \mathsf{P}\{\{|S_n| > x\} \cap A_k\} + \sum_{k=1}^{n} \mathsf{P}\{\{|S_n| > x\} \cap A_k^c\} \\
&\ge \sum_{k=1}^{n} \mathsf{P}\{\{|S_n| > x\} \cap A_k\} \ge \sum_{k=1}^{n} \mathsf{P}\{\{|S_n - S_k| \le y\} \cap A_k\} \\
&= \sum_{k=1}^{n} \mathsf{P}\{|S_n - S_k| \le y\}\mathsf{P}\{A_k\} \ge (1-\beta) \sum_{k=1}^{n} \mathsf{P}\{A_k\} \\
&= (1-\beta)\, \mathsf{P}\left\{ \bigcup_{k=1}^{n} A_k \right\} = (1-\beta)\, \mathsf{P}\left\{ \max_{1\le k\le n} |S_k| > x + y \right\}
\end{aligned}$$

that proves (4.52). □

4.2.3.1 Maxima of partial sums for bounded random variables

The next results deal with *bounded random variables*.

Theorem 4.17. (the 'joint' Kolmogorov inequality) *Let $\xi_1, \xi_2, \ldots, \xi_n$ be* ***independent*** *zero-mean bounded random variables such that*

$$\boxed{\mathsf{E}\{\xi_k\} = 0, \quad \sup_n |\xi_n| \le \xi^+} \tag{4.54}$$

Then

$$\boxed{\mathsf{P}\left\{ \max_{1\le k\le n} |S_k| \le x \right\} \sum_{k=1}^{n} \operatorname{var} \xi_k \ge \left(x + \xi^+\right)^2} \tag{4.55}$$

Proof. Let the events $\{A_k\}$ be defined as in (4.40) and

$$B_k := \left\{ \omega \in \Omega \mid \max_{1\le s\le k} |S_s| \le x \right\} \ (k = 1, \ldots, n) \tag{4.56}$$

Then it is evident that

$$A_k \cap B_k = \varnothing \ (k = 1, \ldots, n), \quad \bigcup_{s=1}^{k} A_s = B_k^c, \quad B_{k-1} = B_k \cup A_{k-1}$$

Thus

$$S_k \chi(B_{k-1}) = \left[S_{k-1} + \xi_k\right] \chi(B_{k-1}) = S_k \left[\chi(B_k) + \chi(A_{k-1})\right]$$

and squaring with the following mathematical expectation application gives

$$\begin{aligned} E\left\{(S_k \chi(B_{k-1}))^2\right\} &= E\left\{(S_{k-1} \chi(B_{k-1}))^2\right\} + E\left\{(\xi_k \chi(B_{k-1}))^2\right\} \\ &\quad + 2E\left\{S_{k-1}\xi_k \chi(B_{k-1})\right\} \\ &= E\left\{S_{k-1}^2 \chi(B_{k-1})\right\} + E\left\{\xi_k^2\right\} \\ &= E\left\{S_{k-1}^2 \chi(B_{k-1})\right\} + \operatorname{var}\xi_k \mathsf{P}\{B_k\} \end{aligned}$$

On the other hand,

$$\begin{aligned} E\left\{(S_k \chi(B_{k-1}))^2\right\} &= E\left\{(S_k [\chi(B_k) + \chi(A_k)])^2\right\} \\ &= E\left\{S_k^2 \chi(B_k)\right\} + E\left\{S_k^2 \chi(A_k)\right\} + 2E\left\{S_k^2 \chi(B_k)\chi(A_k)\right\} \\ &= E\left\{S_k^2 \chi(B_k)\right\} + E\left\{S_k^2 \chi(A_k)\right\} \\ &= E\left\{S_k^2 \chi(B_k)\right\} + E\left\{\left[S_{k-1} + \xi_k\right]^2 \chi(A_k)\right\} \\ &\le E\left\{S_k^2 \chi(B_k)\right\} + \left(x + \xi^+\right)^2 \mathsf{P}\{A_k\} \end{aligned}$$

Joining these inequalities implies, and taking into account that $B_k \supset B_n$ (or equivalently, $\mathsf{P}\{B_k\} \ge \mathsf{P}\{B_k\}$),

$$E\left\{S_{k-1}^2 \chi(B_{k-1})\right\} + \operatorname{var}\xi_k \mathsf{P}\{B_k\} \le E\left\{S_k^2 \chi(B_k)\right\} + \left(x + \xi^+\right)^2 \mathsf{P}\{A_k\}$$

and

$$\begin{aligned} \operatorname{var}\xi_k \mathsf{P}\{B_n\} \le \operatorname{var}\xi_k \mathsf{P}\{B_k\} &\le E\left\{S_k^2 \chi(B_k)\right\} \\ &\quad - E\left\{S_{k-1}^2 \chi(B_{k-1})\right\} + \left(x + \xi^+\right)^2 \mathsf{P}\{A_k\} \end{aligned}$$

that after summation leads to the following relation (here $B_0 = \varnothing$):

$$\begin{aligned} \mathsf{P}\{B_n\} \sum_{k=1}^{n} \operatorname{var}\xi_k &\le E\left\{S_n^2 \chi(B_n)\right\} + \left(x + \xi^+\right)^2 \sum_{k=1}^{n} \mathsf{P}\{A_k\} \\ &\le E\left\{S_n^2 \chi(B_n)\right\} + \left(x + \xi^+\right)^2 \mathsf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} \\ &\le x^2 \mathsf{P}\{B_n\} + \left(x + \xi^+\right)^2 \mathsf{P}\{B_n^c\} \end{aligned}$$

$$= x^2 \mathsf{P}\{B_n\} + \left(x + \xi^+\right)^2 (1 - \mathsf{P}\{B_n\})$$
$$= \left(x + \xi^+\right)^2 - \left(x + \xi^+\right)^2 \mathsf{P}\{B_n\} \left[1 - \frac{x^2}{\left(x + \xi^+\right)^2}\right] \leq \left(x + \xi^+\right)^2$$

that gives (4.55). □

Corollary 4.9. *Under the assumptions of Theorem 4.17 and when*

$$\boxed{\sum_{k=1}^{n} \operatorname{var} \xi_k > 0} \tag{4.57}$$

it follows that

$$\boxed{\mathsf{P}\left\{\max_{1 \leq k \leq n} |S_k| > x\right\} \geq 1 - \frac{\left(x + \xi^+\right)^2}{\sum_{k=1}^{n} \operatorname{var} \xi_k}} \tag{4.58}$$

Corollary 4.10. *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent with non zero-mean bounded random variables such that*

$$\boxed{\mathsf{E}\{\xi_k\} \neq 0, \quad \sup_n |\xi_n| \leq \xi^+} \tag{4.59}$$

and when (4.57) holds, one has

$$\boxed{\mathsf{P}\left\{\max_{1 \leq k \leq n} |S_k - E\{S_k\}| > x\right\} \geq 1 - \frac{\left(x + 2\xi^+\right)^2}{\sum_{k=1}^{n} \operatorname{var} \xi_k}} \tag{4.60}$$

Proof. The result follows from (4.58) if we take into account that $|S_k - E\{S_k\}| \leq 2\xi^+$. □

4.3 Inequalities between moments of sums and summands

The inequality below suggests a relation between quantities of the order p in the sense that it states an upper bound for the p-moment of a sum through the p-moments of its summands.

Theorem 4.18. **(Gut, 2005)** *Let $p \geq 1$ and $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables such that $\mathsf{E}\left\{|\xi_k|^p\right\} < \infty$ for all $k = 1, \ldots, n$. Then the following inequality for the summand $S_n := \sum_{k=1}^{n} \xi_k$ holds:*

$$\boxed{\mathsf{E}\left\{|S_n|^p\right\} \leq \max\left\{2^p \sum_{k=1}^{n} \mathsf{E}\left\{|\xi_k|^p\right\}, 2^{p^2} \left(\sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}\right)^p\right\}} \tag{4.61}$$

Proof. Letting

$$\bar{S}_n := \sum_{k=1}^{n} |\xi_k|, \quad S_n^{(j)} := \sum_{k=1, k \neq j}^{n} \xi_k$$

$$\begin{aligned}
\mathsf{E}\left\{|S_n|^p\right\} &\leq \mathsf{E}\left\{\bar{S}_n^p\right\} = \mathsf{E}\left\{\bar{S}_n^{p-1} \sum_{k=1}^{n} |\xi_k|\right\} = \sum_{k=1}^{n} \mathsf{E}\left\{\bar{S}_n^{p-1} |\xi_k|\right\} \\
&\leq 2^{p-1} \sum_{k=1}^{n} \mathsf{E}\left\{\left(\bar{S}_n^{p-1} + |\xi_k|^{p-1}\right) |\xi_k|\right\} \\
&= 2^{p-1} \sum_{k=1}^{n} \left(\mathsf{E}\left\{\bar{S}_n^{p-1}\right\} \mathsf{E}\{|\xi_k|\} + \mathsf{E}\left\{|\xi_k|^p\right\}\right) \\
&= 2^{p-1} \left(\sum_{k=1}^{n} \left[\mathsf{E}\left\{|\xi_k|^p\right\} + \mathsf{E}\left\{\bar{S}_n^{p-1}\right\}\right] \sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}\right) \\
&\leq 2^{p-1} \left(\sum_{k=1}^{n} \mathsf{E}\left\{|\xi_k|^p\right\} + \left[\mathsf{E}\left\{\bar{S}_n^p\right\}\right]^{(p-1)/p} \sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}\right) \\
&\leq 2^p \max\left\{\sum_{k=1}^{n} \mathsf{E}\left\{|\xi_k|^p\right\}, \left[\mathsf{E}\left\{\bar{S}_n^p\right\}\right]^{(p-1)/p} \sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}\right\}
\end{aligned}$$

Thus

$$\mathsf{E}\left\{\left|\bar{S}_n\right|^p\right\} \leq 2^p \sum_{k=1}^{n} \mathsf{E}\left\{|\xi_k|^p\right\}$$

and

$$\mathsf{E}\left\{\left|\bar{S}_n\right|^p\right\} \leq 2^p \left[\mathsf{E}\left\{\left(\bar{S}_n\right)^p\right\}\right]^{(p-1)/p} \sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}$$

or equivalently

$$\left(\mathsf{E}\left\{\left|\bar{S}_n\right|^p\right\}\right)^{1-(p-1)/p} = \left(\mathsf{E}\left\{\left|\bar{S}_n\right|^p\right\}\right)^{1/p} \leq 2^p \sum_{k=1}^{n} \mathsf{E}\{|\xi_k|\}$$

that completes the proof. □

This page intentionally left blank

5 Characteristic Functions

Contents

The *method of characteristic functions* is one of the tools in analytical probabilistic theory. Adding of independent variables is a frequent component of probabilistic calculations. Mathematically this corresponds to convolving functions. Analogous to the deterministic analysis, where the Laplace or Fourier transformations transform convolution into multiplication, the method of characteristic functions converts adding of independent variables into multiplications of transforms. Below it will be shown that the so-called continuity property permits us to determine limits of distributions based on limits of transforms. This will appear clearly in the proof of the *central limit theorem* which, in some sense, generalizes the De Moivre–Laplace theorem used frequently in complex analysis. The material of this chapter follows Shiryayev (1984) and Gut (2005).

5.1 Definitions and examples

Here we will use complex numbers where $\bar{x} := u - iv$ is the *complex conjugate* to the complex numbers $x = u + iv$, and $|x|^2 = u^2 + v^2$.

Definition 5.1.

1. *Let* $F = F(x_1, x_2 \ldots, x_n)$ *be an n-dimensional distribution function in* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *(2.3). Its* ***characteristic function*** $\varphi : \mathbb{R}^n \to \mathbb{C}$ *is defined by*

$$\boxed{\varphi(t) := \int_{\mathbb{R}^n} e^{i(t,x)} dF(x)} \tag{5.1}$$

2. *If* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *is a random vector defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *with values in* $\mathbb{R}^n$*, then its* ***characteristic function*** $\varphi_\xi : \mathbb{R}^n \to \mathbb{C}$ *is*

$$\boxed{\varphi_\xi(t) := \int_{\mathbb{R}^n} e^{i(t,x)} dF_\xi(x) = \mathsf{E}\{e^{i(t,\xi)}\}} \tag{5.2}$$

where $F_\xi(x) = F_\xi(x_1, x_2 \ldots, x_n)$ *is the distribution function of the random vector* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$.

Remark 5.1. *If* $F_\xi(x)$ *has the density* $f_\xi(x)$, *namely, if* $dF_\xi(x) = f_\xi(x)\,dx$, *then*

$$\boxed{\varphi_\xi(t) := \int_{\mathbb{R}^n} e^{i(t,x)} f_\xi(x)\,dx} \tag{5.3}$$

that is, in this case the characteristic function $\varphi_\xi(t)$ *is just the Fourier transform (with a minus* sign *in the exponent and a factor* $1/\sqrt{2\pi}$*) of the density function* $f_\xi(x)$.

5.1.1 Some examples of characteristic functions

5.1.1.1 Bernoulli random variable

Let ξ be a Bernoulli random variable taking two values: 0 and 1. The characteristics of this random variable are as follows:

$$\boxed{\mathsf{P}\{\xi = 1\} = p > 0, \quad \mathsf{P}\{\xi = 0\} = q, \quad p + q = 1} \tag{5.4}$$

Proposition 5.1. *For the Bernoulli random variable* ξ *satisfying (5.4)*

$$\boxed{\varphi_\xi(t) = pe^{it} + q} \tag{5.5}$$

Proof. It follows directly from Definition (5.2). Indeed,

$$\begin{aligned}\varphi_\xi(t) &= \mathsf{E}\{e^{it\xi}\} \\ &= e^{it\xi}\,|_{\xi=1}\,\mathsf{P}\{\xi = 1\} + e^{it\xi}\,|_{\xi=0}\,\mathsf{P}\{\xi = 0\} = pe^{it} + q\end{aligned}$$ □

5.1.1.2 Gaussian random variable

Proposition 5.2.

1. *If* ξ *is a* ***normalized Gaussian scalar*** *random variable, i.e.,* $\xi \sim \mathcal{N}(0, 1)$ *having* $\mathsf{E}\{\xi\} = 0$ *and* $\mathsf{E}\{\xi^2\} = 1$, *then*

$$\boxed{\varphi_\xi(t) = e^{-t^2/2}} \tag{5.6}$$

2. *If* η *is a* ***Gaussian scalar*** *random variable, i.e.,* $\eta \sim \mathcal{N}\left(m, \sigma^2\right)$ $\left(|m| < \infty,\ \sigma^2 > 0\right)$ *having* $\mathsf{E}\{\eta\} = m$ *and* $\mathsf{E}\{(\eta - m)^2\} = \sigma^2$, *then*

$$\boxed{\varphi_\eta(t) = e^{itm}\varphi_\xi(\sigma t) = e^{itm - t^2\sigma^2/2}} \tag{5.7}$$

3. *If* $\boldsymbol{\xi}$ *is a* ***normalized Gaussian random vector****, i.e.,* $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ *having* $\mathsf{E}\{\boldsymbol{\xi}\} = \mathbf{0}$ *and* $\mathsf{E}\{\boldsymbol{\xi}\boldsymbol{\xi}^\intercal\} = \mathbf{I} := \operatorname{diag}(1, 1, \ldots, 1) \in \mathbb{R}^{n \times n}$*, then*

$$\boxed{\varphi_{\boldsymbol{\xi}}(t) = e^{-\|t\|^2/2}, \quad t \in \mathbb{R}^n} \tag{5.8}$$

4. *If* $\boldsymbol{\eta}$ *is a* ***Gaussian random vector****, i.e.,* $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{m}, \mathbf{R})$ *having* $\mathsf{E}\{\boldsymbol{\eta}\} = \mathbf{0}$ *and* $\mathsf{E}\{(\boldsymbol{\eta} - \boldsymbol{m})(\boldsymbol{\eta} - \boldsymbol{m})^\intercal\} = \mathbf{R} = \mathbf{R}^\intercal > 0 \in \mathbb{R}^{n \times n}$*, then*

$$\boxed{\varphi_{\boldsymbol{\eta}}(t) = e^{i(t,\mathbf{m})}\varphi_{\boldsymbol{\xi}}\left(\mathbf{R}^{1/2}t\right) = e^{i(t,\mathbf{m})-(t,\mathbf{R}t)/2}, \quad t \in \mathbb{R}^n} \tag{5.9}$$

Proof.

1. From (5.3) for

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

it follows that

$$\begin{aligned}
\varphi_\xi(t) &:= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itx}e^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sum_{k=0}^{\infty}\frac{(itx)^k}{k!}e^{-x^2/2}dx = \sum_{k=0}^{\infty}\frac{(it)^k}{k!}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^k e^{-x^2/2}dx
\end{aligned}$$

Since

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^k e^{-x^2/2}dx = \begin{cases} 0 & \text{if} \quad k = 2l - 1, \quad (l = 1, 2, \ldots) \\ \dfrac{(2l)!}{l!} & \text{if} \quad k = 2l, \quad (l = 1, 2, \ldots) \\ 1 & \text{if} \quad k = 0 \end{cases}$$

one has

$$\begin{aligned}
\varphi_\xi(t) &= \sum_{l=0}^{\infty}\frac{(it)^{2l}}{(2l)!}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^{2l}e^{-x^2/2}dx = \sum_{l=0}^{\infty}\frac{\left(-t^2\right)^l}{(2l)!}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^{2l}e^{-x^2/2}dx \\
&= \sum_{l=0}^{\infty}\frac{\left(-t^2/2\right)^l}{l!}\left(\frac{l!}{(2l)!}\right)\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^{2l}e^{-x^2/2}dx = \sum_{l=0}^{\infty}\left(-\frac{t^2}{2}\right)^l\frac{1}{l!} = e^{-t^2/2}
\end{aligned}$$

2. Define $\xi := (\eta - m)/\sigma$. It is evident that $\xi \sim \mathcal{N}(0,1)$, since

$$\mathsf{E}\{\xi\} = \frac{1}{\sigma}\mathsf{E}\{\eta - m\} = \frac{m-m}{\sigma} = 0$$
$$\mathsf{E}\{\xi^2\} = \frac{1}{\sigma^2}\mathsf{E}\{(\eta - m)^2\} = \frac{\sigma^2}{\sigma^2} = 1$$

Hence

$$\varphi_\eta(t) = \mathsf{E}\{e^{it\eta}\} = \mathsf{E}\{e^{it(\sigma\xi+m)}\} = e^{itm}\varphi_\xi(\sigma t) = e^{itm - t^2\sigma^2/2}$$

3. For

$$f_{\boldsymbol{\xi}}(x) = \frac{1}{(2\pi)^{n/2}}e^{-\|x\|^2/2}, \quad x \in \mathbb{R}^n$$

it follows that

$$\begin{aligned}
\varphi_\xi(t) &= \int_{\mathbb{R}^n} e^{i(t,x)} f_\xi(x)\,dx = \int_{\mathbb{R}^n} e^{i(t,x)}\frac{1}{(2\pi)^{n/2}}e^{-\|x\|^2/2}dx \\
&= \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \exp\left(i\sum_{k=1}^n t_k x_k\right)\exp\left(-\sum_{k=1}^n x_k^2/2\right)dx \\
&= \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \exp\left(i\sum_{k=1}^n t_k x_k - \sum_{k=1}^n x_k^2/2\right)dx \\
&= \int_{\mathbb{R}^n}\prod_{k=1}^n \frac{1}{\sqrt{2\pi}}\exp\left(it_k x_k - x_k^2/2\right)dx_1\cdots dx_k \\
&= \prod_{k=1}^n \varphi_\xi(t_k) = \prod_{k=1}^n e^{-t_k^2/2} = \exp\left(-\sum_{k=1}^n t_k^2/2\right) = e^{-\|t\|^2/2}
\end{aligned}$$

4. If $x \in \mathbb{R}^n$ and

$$f_{\boldsymbol{\xi}}(x) = \frac{1}{(2\pi)^{n/2}(\det\mathbf{R})^{1/2}}\exp\left\{-\frac{1}{2}(x-\mathbf{m})^\intercal\mathbf{R}^{-1}(x-\mathbf{m})\right\}$$

then

$$\boldsymbol{\xi} := \mathbf{R}^{-1/2}(\boldsymbol{\eta} - \mathbf{m}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

since

$$\begin{aligned}
\mathsf{E}\{\boldsymbol{\xi}\} &= \mathbf{R}^{-1/2}\mathsf{E}\{\boldsymbol{\eta} - \mathbf{m}\} = \mathbf{R}^{-1/2}(\mathbf{m} - \mathbf{m}) = 0 \\
\mathsf{E}\{\boldsymbol{\xi}\boldsymbol{\xi}^\intercal\} &= \mathbf{R}^{-1/2}\mathsf{E}\{(\boldsymbol{\eta} - \mathbf{m})(\boldsymbol{\eta} - \mathbf{m})^\intercal\}\mathbf{R}^{-1/2} \\
&= \mathbf{R}^{-1/2}\mathbf{R}\mathbf{R}^{-1/2} = \mathbf{I}
\end{aligned}$$

and therefore

$$\varphi_{\boldsymbol{\eta}}(t) = \mathsf{E}\{e^{i(t,\boldsymbol{\eta})}\} = \mathsf{E}\{e^{i\left(t,\left(\mathbf{R}^{1/2}\boldsymbol{\xi}+\mathbf{m}\right)\right)}\}$$
$$= e^{i(t,\mathbf{m})}\mathsf{E}\{e^{i\left(\mathbf{R}^{1/2}t,\boldsymbol{\xi}\right)}\} = e^{i(t,\mathbf{m})}\varphi_{\boldsymbol{\xi}}\left(\mathbf{R}^{1/2}t\right) = e^{i(t,\mathbf{m})-(t,\mathbf{R}t)/2}$$

Proposition is proven. □

5.1.1.3 Poisson random variable

For a Poisson random variable ξ

$$\boxed{\mathsf{P}\{\xi = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots} \tag{5.10}$$

the following formula holds.

Proposition 5.3. *For a Poisson random variable ξ satisfying (5.10) it follows that*

$$\boxed{\varphi_{\xi}(t) = \exp\left\{\lambda\left(e^{it} - 1\right)\right\}} \tag{5.11}$$

Proof. The direct calculation implies

$$\varphi_{\xi}(t) = \mathsf{E}\{e^{it\xi}\} = \sum_{k=0}^{\infty} e^{itk}\frac{e^{-\lambda}\lambda^k}{k!}$$
$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^{it}\right)^k}{k!} = \exp\left\{\lambda\left(e^{it} - 1\right)\right\}$$ □

The characteristic functions for other random variables can be found in Chapter 4 of Gut (2005).

5.1.2 A linear transformed random vector and the sum of independent random vectors

5.1.2.1 A linear transformed random vector

Lemma 5.1. *If $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is a random vector defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with values in $\mathbb{R}^n$, then the **characteristic function** $\varphi_\eta : \mathbb{R}^n \to \mathbb{C}$ of the random vector*

$$\boxed{\begin{gathered}\eta = A\xi + b\\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m\end{gathered}} \tag{5.12}$$

is

$$\boxed{\varphi_\eta(t) := e^{i(t,b)}\varphi_\xi\left(A^\intercal t\right)} \tag{5.13}$$

where $F_{\xi}(x) = F_{\xi}(x_1, x_2 \ldots, x_n)$ *is the distribution function of the random vector* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$.

Proof. It follows from the relations

$$\begin{aligned}\varphi_{\eta}(t) &:= \int_{\mathbb{R}^n} e^{i(t,x)} dF_{\eta}(x) = \mathsf{E}\{e^{i(t,\eta)}\} \\ &= \mathsf{E}\{e^{i(t,A\xi+b)}\} = e^{i(t,b)}\mathsf{E}\{e^{i(A^{\mathsf{T}}t,\xi)}\} = e^{i(t,b)}\varphi_{\xi}\left(A^{\mathsf{T}}t\right)\end{aligned}$$

□

5.1.2.2 The sum of independent random vectors

Lemma 5.2. (on multiplication) *If* $\xi_1, \xi_2, \ldots, \xi_n$ *are* ***independent*** *random vectors defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *with values in* $\mathbb{R}^n$, *and* $S_n := \sum_{k=1}^{n} \xi_k$ *is their partial sum, then the characteristic function* $\varphi_{S_n} : \mathbb{R}^n \to \mathbb{C}$ *of the random vector* S_n *is*

$$\boxed{\varphi_{S_n}(t) = \prod_{k=1}^{n} \varphi_{\xi_k}(t)} \tag{5.14}$$

Proof. It is trivial since

$$\begin{aligned}\varphi_{S_n}(t) &:= \mathsf{E}\{e^{i(t,S_n)}\} = \mathsf{E}\left\{\exp\left\{i\left(t, \sum_{k=1}^{n}\xi_k\right)\right\}\right\} \\ &= \mathsf{E}\left\{\prod_{k=1}^{n}\exp\{i(t,\xi_k)\}\right\} = \prod_{k=1}^{n}\mathsf{E}\{\exp\{i(t,\xi_k)\}\} = \prod_{k=1}^{n}\varphi_{\xi_k}(t)\end{aligned}$$

□

5.2 Basic properties of characteristic functions

5.2.1 Simple properties

Several simple properties of the characteristic function $\varphi_{\xi}(t)$ (5.2) are presented below.

Lemma 5.3. *Let* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *be a random vector defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *with values in* $\mathbb{R}^n$. *Then*

(a)

$$\boxed{\left|\varphi_{\xi}(t)\right| \le \varphi_{\xi}(0) = 1} \tag{5.15}$$

(b)

$$\boxed{\overline{\varphi_{\xi}(t)} = \varphi_{\xi}(-t) = \varphi_{-\xi}(t)} \tag{5.16}$$

(c) $\varphi_{\xi}(t)$ *is uniformly continuous on* $\mathbb{R}^n$.

(d) $\varphi_\xi(t)$ *is real-valued in the case* $n = 1$ *if and only if* $F_\xi(x)$ *is symmetric, namely,*

$$\boxed{\int_B dF_\xi(x) = \int_{-B} dF_\xi(x)} \tag{5.17}$$

where $B \in \mathcal{B}(\mathbb{R})$ *and* $-B := \{-x : x \in B\}$.

Proof.

(a) The first fact is obvious since

$$\begin{aligned} |\varphi_\xi(t)| &= \left| \int_{\mathbb{R}^n} e^{i(t,x)} dF_\xi(x) \right| \leq \int_{\mathbb{R}^n} \left| e^{i(t,x)} \right| dF_\xi(x) \\ &= \int_{\mathbb{R}^n} 1 dF_\xi(x) = \int_{\mathbb{R}^n} \left| e^{i(0,x)} \right| dF_\xi(x) = 1 \end{aligned}$$

(b) The second property results from the identities

$$\begin{aligned} \overline{\varphi_\xi(t)} &= \int_{\mathbb{R}^n} e^{-i(t,x)} dF_\xi(x) = \int_{\mathbb{R}^n} e^{i(-t,x)} dF_\xi(x) \\ &= \varphi_\xi(-t) = \int_{\mathbb{R}^n} e^{i(t,-x)} dF_\xi(x) = \varphi_{-\xi}(t) \end{aligned}$$

(c) As for the third property, notice that for all $h \in \mathbb{R}^n$

$$\begin{aligned} |\varphi_\xi(t+h) - \varphi_\xi(t)| &= \left| \int_{\mathbb{R}^n} e^{i(t,x)} \left[e^{i(h,x)} - 1 \right] dF_\xi(x) \right| \\ &\leq \int_{\mathbb{R}^n} \left| e^{i(t,x)} \right| \left| e^{i(h,x)} - 1 \right| dF_\xi(x) \leq \int_{\mathbb{R}^n} \left| e^{i(h,x)} - 1 \right| dF_\xi(x) \\ &\leq \int_{\mathbb{R}^n} \left| e^{i(h,x)} - 1 \right| \chi(\|x\| \leq a)\, dF_\xi(x) \\ &\quad + \int_{\mathbb{R}^n} \left| e^{i(h,x)} - 1 \right| \chi(\|x\| > a)\, dF_\xi(x) \end{aligned}$$

Since by the Euler formula

$$\begin{aligned} \left| e^{iz} - 1 \right| &= \sqrt{(\cos z - 1)^2 + \sin^2 z} \\ &= \sqrt{2(1 - \cos z)} = \sqrt{4 \sin^2 \frac{z}{2}} = 2 \left| \sin \frac{z}{2} \right| \leq \min(2, |z|) \end{aligned} \tag{5.18}$$

then for any $a > 0$ it follows that

$$\left|\varphi_\xi(t+h) - \varphi_\xi(t)\right| \le \int_{\mathbb{R}^n} \left|e^{i(h,x)} - 1\right| \chi(\|x\| \le a)\, dF_\xi(x)$$

$$+ \int_{\mathbb{R}^n} \left|e^{i(h,x)}\text{-}1\right| \chi(\|x\| > a)\, dF_\xi(x) \le \int_{\mathbb{R}^n} |(h,x)|\, \chi(\|x\| \le a)\, dF_\xi(x)$$

$$2\int_{\mathbb{R}^n} \chi(\|x\| > a)\, dF_\xi(x) \le \|h\|\, a + 2\mathsf{P}\{\|x\| > a\}$$

Taking then a so large that $2\mathsf{P}\{\|x\| > a\} \le \varepsilon/2$ and h so small that $\|h\|\, a \le \varepsilon/2$ we state that $\left|\varphi_\xi(t+h) - \varphi_\xi(t)\right| \le \varepsilon$ independently on t which means exactly the uniform continuity.

(d) Let $F_\xi(x)$ be symmetric. Then for a bounded and odd Borel function $g(x)$ it follows that $\int_{\mathbb{R}} g(x)\, dF_\xi(x) = 0$. As a result, $\int_{\mathbb{R}^n} \sin(tx)\, dF_\xi(x) = 0$, and hence, $\varphi_\xi(t) = \mathsf{E}\{\cos(t,\xi)\}$ is real. Conversely, if $\varphi_\xi(t)$ is a real function, then by (5.16)

$$\varphi_{-\xi}(t) = \varphi_\xi(-t) = \overline{\varphi_\xi(t)} = \varphi_\xi(t)$$

Therefore, $F_\xi(x)$ and $F_{-\xi}(x)$ are the same, and for each $B \in \mathcal{B}(\mathbb{R})$

$$\mathsf{P}\{\xi \in B\} = \mathsf{P}\{-\xi \in B\} = \mathsf{P}\{\xi \in -B\}$$

that proves the result. □

5.2.2 Relations with moments

The next lemma describes not so trivial properties of characteristic functions stating the relation between the derivatives of $\varphi_\xi(t)$ $(n = 1)$ and the moments $\mathsf{E}\{\xi^k\}$ of the random variable ξ.

Lemma 5.4. **(Shiryayev, 1984)** *If ξ is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with values in $\mathbb{R}$ and $\varphi_\xi(t) = \mathsf{E}\{e^{it\xi}\}$ its characteristic function, then*

(a) If $\mathsf{E}\{|\xi|^n\} < \infty$ for some integer $n \ge 1$, then $\varphi_\xi^{(r)}(t)$ exists for every $r \le n$ and

$$\boxed{\varphi_\xi^{(r)}(t) = \int_{\mathbb{R}} (ix)^r e^{itx} dF_\xi(x)} \tag{5.19}$$

$$\boxed{\varphi_\xi^{(r)}(0) = i^r \mathsf{E}\{\xi^r\}} \tag{5.20}$$

$$\boxed{\varphi_\xi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \mathsf{E}\{\xi^r\} + \frac{(it)^n}{n!} O_n(t)} \tag{5.21}$$

where $|O_n(t)| \le 3\mathsf{E}\{|\xi|^n\}$ and $O_n(t) \to 0$ whenever $t \to 0$.

(b) If $\varphi_{\xi}^{(2n)}(0)$ *exists (finite) then* $\mathsf{E}\{\xi^{2n}\} < \infty$.

(c) If $\mathsf{E}\{|\xi|^{n}\} < \infty$ *for all integer* $n \geq 1$ *and*

$$\boxed{R^{-1} := \limsup_{n\to\infty} n^{-1}\left(\mathsf{E}\{|\xi|^{n}\}\right)^{1/n} < \infty} \tag{5.22}$$

then

$$\boxed{\varphi_{\xi}(t) = \sum_{n=0}^{\infty} \frac{(it)^{n}}{n!}\mathsf{E}\{\xi^{n}\}} \tag{5.23}$$

for all $t : |t| < R$.

Proof.

(a) If $\mathsf{E}\{|\xi|^{n}\} < \infty$, then by the Lyapunov inequality (4.23) it follows that $\mathsf{E}\{|\xi|^{r}\} < \infty$ for all integer positive $r < n$. Since

$$\frac{\varphi_{\xi}(t+h) - \varphi_{\xi}(t)}{h} = \mathsf{E}\left\{e^{it\xi}\left(\frac{e^{it\xi}-1}{h}\right)\right\}$$

and in view of the inequality (5.18) for $z := ix$ providing the estimate

$$\frac{\left|e^{ihx}-1\right|}{h} \leq |x|$$

it follows that $\lim_{t\to 0}\mathsf{E}\left\{e^{it\xi}\left(\frac{e^{it\xi}-1}{h}\right)\right\}$ exists and it is equal to

$$\lim_{t\to 0}\mathsf{E}\left\{e^{it\xi}\left(\frac{e^{it\xi}-1}{h}\right)\right\} = i\mathsf{E}\{\xi e^{it\xi}\} = i\int_{\mathbb{R}} xe^{itx}dF_{\xi}(x)$$

Hence, $\varphi_{\xi}'(t)$ exists and

$$\varphi_{\xi}'(t) = i\int_{\mathbb{R}} xe^{itx}dF_{\xi}(x) = i\mathsf{E}\{\xi e^{it\xi}\}$$

The existence of $\varphi_{\xi}^{(r)}(t)$ for $r \geq 2$ follows by induction. The formula (5.20) follows directly from (5.19). To establish (5.21) notice that

$$e^{iy} = \cos y + i\sin y = \sum_{k=1}^{n-1}\frac{(iy)^{k}}{k!} + \frac{(iy)^{n}}{n!}\left[\cos(\theta_1 y) + i\sin(\theta_2 y)\right]$$

for real $y \in \mathbb{R}$ and $|\theta_i| \leq 1$ $(i = 1, 2)$. Letting in this formula $y := t\xi$ one has

$$e^{it\xi} = \sum_{k=1}^{n-1} \frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!} [\cos(\theta_1(\omega) t\xi) + i \sin(\theta_2(\omega) t\xi)]$$

$$\mathsf{E}\left\{e^{it\xi}\right\} = \sum_{k=1}^{n-1} \frac{(it)^k}{k!} \mathsf{E}\{\xi^k\} + \frac{(it)^n}{n!} \mathsf{E}\{\xi^n\} + O_n(t)$$

where

$$O_n(t) := \frac{(it)^n}{n!} \mathsf{E}\{\xi^n [\cos(\theta_1(\omega) t\xi) + i \sin(\theta_2(\omega) t\xi) - 1]\}$$

Evidently $|O_n(t)| \leq 3\mathsf{E}\{|\xi|^n\}$ and $O_n(t) \to 0$ when $t \to 0$.

(b) This property also can be proven by induction. Suppose that $\varphi_\xi''(0)$ exist (finite). Show that $\mathsf{E}\{\xi^2\} < \infty$. By the L'Hôpital rule and by the Fatou lemma (Lemma 16.4 in Poznyak (2008)) it follows that

$$\begin{aligned}
\varphi_\xi''(0) &= \lim_{h\to 0} \frac{1}{2}\left[\frac{\varphi_\xi'(2h) - \varphi_\xi'(0)}{2h} + \frac{\varphi_\xi'(0) - \varphi_\xi'(-2h)}{2h}\right] \\
&= \lim_{h\to 0} \frac{2\varphi_\xi'(2h) - 2\varphi_\xi'(-2h)}{8h} = \lim_{h\to 0} \frac{\varphi_\xi(2h) - 2\varphi_\xi(0) + \varphi_\xi(-2h)}{4h^2} \\
&= \lim_{t\to 0} \int_{\mathbb{R}} \left(\frac{e^{ihx} - e^{-ihx}}{2h}\right)^2 dF_\xi(x) = -\lim_{t\to 0} \int_{\mathbb{R}} \left(\frac{\sin(hx)}{hx}\right)^2 x^2 dF_\xi(x) \\
&\leq -\int_{\mathbb{R}} \lim_{h\to 0} \left(\frac{\sin(hx)}{hx}\right)^2 x^2 dF_\xi(x) = -\int_{\mathbb{R}} x^2 dF_\xi(x) = -\mathsf{E}\{\xi^2\}
\end{aligned}$$

which implies $\mathsf{E}\{\xi^2\} \leq \varphi_\xi''(0) < \infty$. Let now $\varphi_\xi^{(2k+2)}(0)$ exists and $\mathsf{E}\{\xi^{2k}\} < \infty$. If

$$\mathsf{E}\{\xi^{2k}\} = \int_{\mathbb{R}} x^{2k} dF_\xi(x) = 0$$

then obviously $\int_{\mathbb{R}} x^{2k+2} dF_\xi(x) = 0$. So we may suppose that $\int_{\mathbb{R}} x^{2k} dF_\xi(x) > 0$. Then by (5.19)

$$\varphi_\xi^{(2k)}(t) = \int_{\mathbb{R}} (ix)^{2k} e^{itx} dF_\xi(x)$$

and hence

$$(-1)^k \varphi_\xi^{(2k)}(t) = \int_{\mathbb{R}} e^{itx} dG_\xi(x)$$

with $G_\xi(x) = \int_{-\infty}^{x} u^{2k} dF_\xi(x)$. This means that the function $(-1)^k \varphi_\xi^{(2k)}(t)\, G_\xi^{-1}(\infty)$ is the characteristic function of the probability distribution $G_\xi(x)\, G_\xi^{-1}(\infty)$, and moreover, we have proved that

$$G_\xi^{-1}(\infty) \int_{\mathbb{R}} x^2 dG_\xi(x) < \infty$$

But since $G_\xi^{-1}(\infty) > 0$ we have

$$\int_{\mathbb{R}} x^{2k+2} dF_\xi(x) = \int_{\mathbb{R}} x^2 dG_\xi(x) < \infty$$

that proves this item.

(c) Let $t_0 \in (0, R)$. Then by the assumption

$$\limsup_{n\to\infty} n^{-1} \left(\mathsf{E}\{|\xi|^n\}\right)^{1/n} < t_0^{-1}$$

or equivalently,

$$\limsup_{n\to\infty} n^{-1} \left(\mathsf{E}\{|\xi|^n t_0^n\}\right)^{1/n} < 1$$

Hence, by *Stirling's formula* (see the details in Khan (1974))

$$\lim_{n\to\infty} \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{n!} = 1$$

it follows that

$$\lim_{n\to\infty} \left(\frac{\mathsf{E}\{|\xi|^n t_0^n\}}{n!}\right)^{1/n} < 1$$

Therefore, by the Cauchy test (see Criterion 16.1 in Poznyak (2008)) the series $\sum_{r=0}^{\infty} \frac{\mathsf{E}\{|\xi|^r t_0^r\}}{r!}$ converges, which means that the series $\sum_{r=0}^{\infty} (it)^r \frac{\mathsf{E}\{\xi^r\}}{r!}$ converges too. But by (5.21)

$$\varphi_\xi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \mathsf{E}\{\xi^r\} + D_n(t)$$

where $D_n(t) := \frac{(it)^n}{n!} O_n(t)$, $|D_n(t)| \le 3\frac{|t|^n}{n!} \mathsf{E}\{|\xi|^n\}$. Therefore (5.23) holds for all $|t| < R$ which completes the proof. □

5.3 Uniqueness and inversion

5.3.1 Uniqueness

Here we will show that the characteristic function is uniquely determined by the distribution function. The proof will be done for a single-dimensional case.

Theorem 5.1. (on uniqueness) *Let F and G be distribution functions with the same characteristic function, i.e.,*

$$\boxed{\varphi(t) = \int_{\mathbb{R}} e^{i(t,x)} dF(x) = \int_{\mathbb{R}} e^{i(t,x)} dG(x)} \tag{5.24}$$

for all $t \in \mathbb{R}^n$. Then

$$\boxed{F(x) = G(x)} \tag{5.25}$$

Proof. From (5.24) it follows

$$\int_{\mathbb{R}} e^{itx} dM(x) = 0, \quad M(x) := F(x) - G(x)$$

Suppose that there exists $x_0 \in \mathbb{R}$ such that $M(x_0) \neq 0$. If x_0 is a unique isolated point, then

$$0 = \int_{\mathbb{R}} e^{itx} dM(x) = e^{itx_0}$$

that never can be fulfilled since $\left|e^{i(t,x_0)}\right| = 1$. If x_0 is a point of continuity, then suppose that there exists a small neighborhood $\Omega_\varepsilon := \{x : \|x - x_0\| \leq \varepsilon\}$ such that $M(x) \neq 0$ (for example, $M(x) > 0$) for all $x \in \Omega_\varepsilon$ and $\int_{\Omega_\varepsilon} dM(x) \neq 0$. Then, by the generalized mean-value theorem (see, for example, Poznyak (2008))

$$\begin{aligned} 0 &= \int_{\mathbb{R}} e^{itx} dM(x) = \int_{\mathbb{R}} [\cos(tx) + i \sin(tx)] dM(x) \\ &= \cos\left(tx_0'\right) \int_{\Omega_\varepsilon} dM(x) + i \sin\left(tx_0''\right) \int_{\Omega_\varepsilon} dM(x) \\ &= \left[\cos\left(tx_0'\right) + i \sin\left(tx_0''\right)\right] \int_{\Omega_\varepsilon} dM(x) \end{aligned}$$

that implies $\cos\left(tx_0'\right) + i \sin\left(tx_0''\right) = 0$ for all $t \in \mathbb{R}$, which is impossible (even $x_0' \neq x_0''$). The proof is complete. □

5.3.2 *Inversion formula*

The previous theorem declares that a distribution function $F = F(x)$ is uniquely determined by its characteristic function $\varphi = \varphi(t)$. It is an existence result only. Here we will present the theorem which gives a representation of F in terms of φ and provides a formula for explicitly computing the distribution F based on φ.

Theorem 5.2. (the general inverse theorem (Gut, 2005)) *Let $\xi \in \mathbb{R}$ be a random variable with the distribution function $F_\xi(x)$ and the characteristic function $\varphi_\xi(t)$. Then for all $a < b$*

(a)

$$\boxed{\begin{aligned} & F_\xi(b) - F_\xi(a) + \frac{1}{2}\mathsf{P}\{\xi = a\} - \frac{1}{2}\mathsf{P}\{\xi = b\} \\ & \quad = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itb} - e^{-ita}}{(-it)}\varphi_\xi(t)\,dt \end{aligned}} \tag{5.26}$$

(b) *In particular, if a and b are both the points of continuity of $F_\xi(x)$, then*

$$\boxed{F_\xi(b) - F_\xi(a) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itb} - e^{-ita}}{(-it)}\varphi_\xi(t)\,dt} \tag{5.27}$$

(c) *If $\int_{\mathbb{R}} \left|\varphi_\xi(t)\right| dt < \infty$, and the distribution function $F_\xi(x)$ has the density $f_\xi(x)$, i.e., $F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)dy$ is absolutely continuous, then*

$$\boxed{f_\xi(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_\xi(t)\,dt} \tag{5.28}$$

(d) *If the distribution $F_\xi(x)$ has point masses concentrated in the points a_i with $\mathsf{P}\{\xi = a_i\}$ $(i = 1, 2, \ldots)$, then they can be recovered as follows:*

$$\boxed{\mathsf{P}\{\xi = a_i\} = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} e^{-ita_i}\varphi_\xi(t)\,dt} \tag{5.29}$$

Proof.

(a) Again, using the inequality (5.18) for $z = -t(b-a)$, we have

$$\left|\frac{e^{-itb} - e^{-ita}}{t}\right| = \left|e^{-ita}\right| \left|\frac{e^{-it(b-a)} - 1}{t}\right| = \left|\frac{e^{-it(b-a)} - 1}{t}\right| \leq b - a$$

and hence,

$$\left|\int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{t} \varphi_{\xi}(t)\, dt\right| = \int_{-T}^{T} \left|\frac{e^{-itb} - e^{-ita}}{t}\right| \left|\varphi_{\xi}(t)\right| dt \leq \int_{-T}^{T} \left|\frac{e^{-itb} - e^{-ita}}{t}\right| 1 dt \leq 2T(b-a)$$

that shows that the integral in the right-hand side of (5.26) exists. Applying Fubini's theorem, the Euler formula (see both, for example, in Poznyak (2008)) and symmetry, we obtain

$$\begin{aligned}
\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{(-it)} \varphi_{\xi}(t)\, dt &= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{(-it)} \left(\int_{-\infty}^{\infty} e^{itx} dF_{\xi}(x)\right) dt \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{e^{-it(x-a)} - e^{-it(x-b)}}{2it} dt\right) dF_{\xi}(x) \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \left[\frac{\sin t(x-a)}{t} - \frac{\sin t(x-b)}{t}\right] dt\right) dF_{\xi}(x) \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} H(a, b, t, x, T)\, dF_{\xi}(x)
\end{aligned}$$

where $H(a, b, t, x, T)$ is the inner integral. Taking into account that

$$\int_{0}^{T} \frac{\sin x}{x} dx \begin{cases} \int_{0}^{\pi} \frac{\sin x}{x} dx \leq \pi & \text{for all} \quad T > 0 \\ \to \frac{\pi}{2} & \text{as} \quad T \to \infty \end{cases}$$

it follows that

$$\lim_{T\to\infty} H(a,b,t,x,T) = \begin{cases} 0 & \text{if} \quad x < a \\ \dfrac{\pi}{2} & \text{if} \quad x = a \\ \pi & \text{if} \quad a < x < b \\ \dfrac{\pi}{2} & \text{if} \quad x = b \\ 0 & \text{if} \quad x > b \end{cases}$$

So, by the dominating convergence theorem (see Theorem 16.21 in Poznyak (2008))

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{(-it)} \varphi_\xi(t)\, dt &= \frac{1}{\pi} \int_{-\infty}^{\infty} \lim_{T\to\infty} H(a,b,t,x,T)\, dF_\xi(x) \\ &= \frac{1}{2}\mathsf{P}\{\xi = a\} + \mathsf{P}\{a < \xi < b\} + \frac{1}{2}\mathsf{P}\{\xi = b\} \end{aligned}$$

which proves (5.26).

(b) If a and b both are the points of continuity of $F_\xi(x)$ then $\mathsf{P}\{\xi = a\} = \mathsf{P}\{\xi = b\} = 0$ that in view of (5.26) proves (5.27).

(c) Letting in (5.27) $a = x$ and $b = x + h$ $(h > 0)$ we get

$$\begin{aligned} & F_\xi(x+h) - F_\xi(x) + \frac{1}{2}\mathsf{P}\{\xi = x\} - \frac{1}{2}\mathsf{P}\{\xi = x+h\} \\ & \quad \le \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{e^{-itb} - e^{-ita}}{(-it)} \right| \left|\varphi_\xi(t)\right| dt \\ & \quad \le \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} h \left|\varphi_\xi(t)\right| dt \le \frac{h}{2\pi} \int_{-\infty}^{\infty} \left|\varphi_\xi(t)\right| dt = O(h) \underset{h\to 0}{\to} 0 \end{aligned}$$

that leads to the relation

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-it(x+h)} - e^{-itx}}{(-ith)} \varphi_\xi(t)\, dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1 - e^{-ith}}{(ith)} \varphi_\xi(t)\, dt \end{aligned}$$

Letting $h \to 0$ and observing that $\dfrac{1 - e^{-ith}}{(ith)} \to 1$ we obtain (5.28).

(d) By proceeding along the lines of the item (a) we have

$$\frac{1}{2T} \int_{-T}^{T} e^{-ita_i} \varphi_\xi(t)\, dt = \frac{1}{2T} \int_{-T}^{T} e^{-ita_i} \left(\int_{-\infty}^{\infty} e^{itx} dF(x) \right) dt$$

$$= \frac{1}{2T} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} e^{it(x-a_i)} dt \right) dF(x)$$

$$= \frac{1}{2T} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} [\cos t (x - a_i) + i \sin t (x - a_i)] \, dt \right) dF(x)$$

$$= \frac{1}{T} \int_{-\infty}^{\infty} \left(\frac{\sin T (x - a_i)}{(x - a_i)} + 0 \right) dF(x)$$

$$= \int_{\mathbb{R}/a} \frac{\sin T (x - a_i)}{T (x - a_i)} dF(x) + 1 \cdot \mathsf{P}\{\xi = a_i\} \underset{T \to \infty}{\to} \mathsf{P}\{\xi = a_i\}$$

Theorem is proven. □

5.3.3 Parseval's-type relation

Lemma 5.5. (on the Parseval's-type relation (Gut, 2005)) *Let ξ and η be random variables with the distributions $F_\xi(x)$ and $F_\eta(x)$ having the characteristic functions $\varphi_\xi(x)$ and $\varphi_\eta(x)$, respectively. Then*

1.

$$\boxed{\int_{-\infty}^{\infty} e^{-iyu} \varphi_\xi(y) \, dF_\eta(y) = \int_{-\infty}^{\infty} \varphi_\eta(x - u) \, dF_\xi(x)} \tag{5.30}$$

2.

$$\boxed{\int_{-\infty}^{\infty} \varphi_\xi(y) \, dF_\eta(y) = \int_{-\infty}^{\infty} \varphi_\eta(x) \, dF_\xi(x)} \tag{5.31}$$

Proof. Multiplying both sides of the relation $\varphi_\xi(y) = \int_{-\infty}^{\infty} e^{ixy} dF_\xi(x)$ by e^{-iyu} and integrating with respect to $F_\eta(y)$, and applying Fubini's theorem yields

$$\int_{-\infty}^{\infty} e^{-iyu} \varphi_\xi(y) \, dF_\eta(y) = \int_{-\infty}^{\infty} e^{-iyu} \left(\int_{-\infty}^{\infty} e^{ixy} dF_\xi(x) \right) dF_\eta(y)$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{ix(y-u)} dF_\eta(y) \right) F_\xi(x) = \int_{-\infty}^{\infty} \varphi_\eta(x - u) \, dF_\xi(x)$$

that proves (5.30). The formula (5.31) follows from (5.30) if we take $u = 0$. Lemma is proven. □

Remark 5.2. *The usefulness of the relation (5.31) is related to the idea to join two distributions: if the left-hand side is a 'difficult' integral for calculation, then the right-hand side is an 'easy' integral.*

This page intentionally left blank

PART II

Discrete Time Processes

This page intentionally left blank

6 Random Sequences

Contents

This chapter introduces the main definitions and discusses the main properties of *random sequences* which, in the engineering language, are, in fact, *discrete-time random processes.* Two concepts will be discussed below: the first concept tackles the properties of *infinitely often events* and is also closely connected with the *convergence concept*, and the second one is based on some properties of the *Lebesgue integral* with respect to a probabilistic measure and is related to the possibility to take limits under the sign of 'mathematical expectation'. Finally, we will discuss *various modes of convergence* of sequences of random variables and relate them into a hierarchical scheme.

6.1 Random process in discrete and continuous time

Definition 6.1. *Let $T \subseteq \mathbb{R}$ be a subset of the real line $\mathbb{R}$. A set of random variables*

$$\boxed{\{\xi_t(\omega), t \in T\}}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ for any fixed $t \in T$ is called a ***random process with time domain*** *T. If*

(a) $T = \{\ldots, 0, 1, 2, \ldots\}$, *then we call the collection*

$$\boxed{\{\ldots, \xi_0(\omega), \xi_1(\omega), \xi_2(\omega), \ldots\}}$$

a ***random sequence****, or a* ***discrete-time random process****;*

(b) $T = (-\infty, \infty)$ *or* $T \in [a, b]$ $(-\infty < a < b < \infty)$, *then we call*

$$\boxed{\xi_t(\omega), \quad t \in T}$$

a ***random function****, or a* ***continuous-time random process****.*

Possible realizations (corresponding different trajectories ω_1 and ω_2 both from Ω) of random discrete-time and continuous-time processes are shown in Fig. 6.1.

In this part of the book we will consider only **discrete-time processes**. Continuous-time processes generated by stochastic differential equations will be considered in the next part.

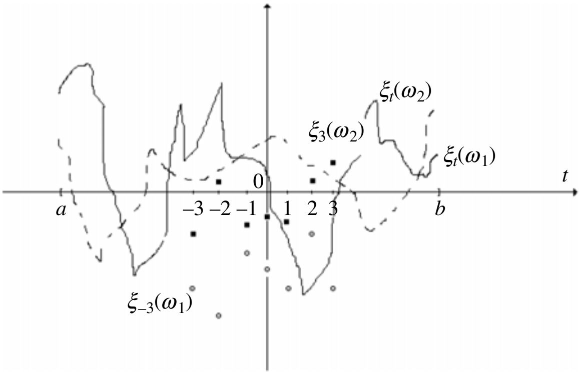

Fig. 6.1. Realizations of a continuous and discrete time process.

6.2 Infinitely often events

6.2.1 Main definition

Recall here the definitions (1.13) of the sets A_* and A^*:

$$\boxed{A_* = \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m, \quad A^* = \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m}$$

where $\{A_n, n \geq 1\}$ are a sequence of subsets from Ω.

Remark 6.1.

(a) *Directly from the definition (1.13) it follows that if $\omega \in A_*$ then $\omega \in \bigcap_{m=n}^{\infty} A_m$ for some n, that is, there exists n such that $\omega \in A_m$ for all $m \geq n$. In particular, if A_n is an event when something special occurs at time n, then $\liminf_{n\to\infty} A_n^c$ is the event that starting from some n this special property* ***never occurs****.*

(b) *Similarly, if $\omega \in A^*$, then $\omega \in \bigcup_{m=n}^{\infty} A_m$ for every n. This means that no matter how large n is, there is always some $m \geq n$ such that $\omega \in A_m$, or equivalently, for* ***infinitely many*** *values of m.*

A convenient way to express the fact that $\omega \in A^*$ is

$$\boxed{\omega \in A^* \text{ is equivalent to } \omega \in \{A_n \text{ i.o.}\} = \{A_n \text{ infinitely often}\}}$$

6.2.2 Tail events and the Kolmogorov zero–one law

One of the magic results on probability theory concerns situations where the probability of an event may be only 0 or 1. To state the corresponding theorem we need to define the σ-algebra containing information about '*what happens at infinity*'.

Let $\{A_k, k \geq 1\}$ be arbitrary events. Following Gut (2005), define the σ-algebra:

$$\boxed{\mathcal{A}_k^n := \sigma\left(A_k, A_{k+1}, \ldots, A_n\right), \quad k \leq n}$$

Definition 6.2. *The **tail-σ-algebra** (or **σ-algebra of remote events**) is defined by*

$$\boxed{\mathcal{T} := \bigcap_{k=0}^{\infty} \mathcal{A}_{k+1}^{\infty}} \tag{6.1}$$

which is a σ-algebra itself since it is an intersection of other σ-algebras.

Considering k as a time-index, we may conclude that $\mathcal{A}_{k+1}^{\infty}$ contains the information beyond time k, and therefore, $\mathcal{T}$ contains the information '*beyond time k for all k*'.

To prove the 'zero–one' statement we need the following intermediate approximation lemma which states that any set in a σ-algebra can be arbitrarily well approximated by another set that belongs to an algebra that generates the a σ-algebra.

Lemma 6.1. (on an approximation of events) *Let $\mathcal{F}_0$ be an algebra that generates the σ-algebra $\mathcal{F}$, that is, $\mathcal{F} = \sigma(\mathcal{F}_0)$. Then for any set $A \in \mathcal{F}$ and any $\varepsilon > 0$ there exists $A_\varepsilon \in \mathcal{F}_0$ such that*

$$\boxed{\mathsf{P}\{A \triangle A_\varepsilon\} < \varepsilon} \tag{6.2}$$

where

$$\boxed{A \triangle B := (A - B) \cup (B - A)} \tag{6.3}$$

*is the **symmetric set difference**.*

Proof. **(Gut, 2005)** For $\varepsilon > 0$ define

$$\mathcal{G} := \{A \in \mathcal{F} : \mathsf{P}\{A \triangle A_\varepsilon\} < \varepsilon \text{ for some } A_\varepsilon \in \mathcal{F}_0\}$$

1. If $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$ since $A^c \triangle (A_\varepsilon)^c = A \triangle A_\varepsilon$.
2. If $A_n \in \mathcal{G}$ $(n \geq 1)$, then so does the union. Namely, set $A = \bigcup_{n=1}^{\infty} A_n$, let ε be given and choose n^* such that $\mathsf{P}\left\{A \backslash \bigcup_{n=1}^{n^*} A_n\right\} < \varepsilon/2$. Next, let $\left\{A_{k,\varepsilon} \subset \mathcal{F}_0, k = 1, \ldots, n^*\right\}$ be such that

$$\mathsf{P}\left\{A \triangle A_{k,\varepsilon}\right\} < \varepsilon/\left(2n^*\right) \quad \text{for all } k = 1, \ldots, n^*$$

Since

$$\left(\bigcup_{k=1}^{n^*} A_k\right) \triangle \left(\bigcup_{k=1}^{n^*} A_{k,\varepsilon}\right) \subset \bigcup_{k=1}^{n^*} \left(A_k \triangle A_{k,\varepsilon}\right)$$

it follows that

$$\mathsf{P}\left\{\left(\bigcup_{k=1}^{n^*} A_k\right) \triangle \left(\bigcup_{k=1}^{n^*} A_{k,\varepsilon}\right)\right\} \leq \sum_{k=1}^{n^*} \mathsf{P}\left\{A_k \triangle A_{k,\varepsilon}\right\} < n^*\varepsilon/\left(2n^*\right) = \varepsilon/2$$

so that, finally,

$$\mathsf{P}\left\{A \triangle \left(\bigcup_{k=1}^{n^*} A_{k,\varepsilon}\right)\right\} < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

that completes the proof. □

Remark 6.2. *The general description of this result reveals that it reduces an infinite setting to a finite one.*

Now we are ready to formulate and prove the main result of this subsection.

Theorem 6.1. (the Kolmogorov zero–one law) *If* $\{A_k, k \geq 1\}$ *are independent events, then the tail-σ-algebra is trivial, i.e., it contains only the sets of probability* 0 *or one, that is, if* $A \in \mathcal{T}$ *then* $\mathsf{P}\{A\} = 0$ *or* $\mathsf{P}\{A\} = 1$.

Proof.
(a) The idea of the proof lies in the following concept. If $A \in \mathcal{T}$, then $A \in \mathcal{A}_{k+1}^{\infty}$ for all $k \geq 0$, and, hence, A is independent of $\mathcal{A}_1^k$, which is true for all k. This implies that in the same time $A \in \sigma(A_1, A_2, \ldots)$ and A is independent of $\sigma(A_1, A_2, \ldots)$. This means that A is '*independent of itself*', and therefore,

$$\mathsf{P}\{A\} = \mathsf{P}\{A \cap A\} = \mathsf{P}^2\{A\}$$

that is possible when $\mathsf{P}\{A\} = 0$ or $\mathsf{P}\{A\} = 1$.
(b) To realize the proof more rigorously (following Breiman (1992), Theorem 3.12), let us use the fact (see Lemma 6.1) that for every set $A \in \bigcup_{n \geq 1} \mathcal{A}_1^n$ and any $\varepsilon > 0$ there is a set $B \in \mathcal{A}_1^n$ for some n such that $\mathsf{P}\{A \triangle B\} \leq \varepsilon$. Let $E \in \mathcal{T}$. Then there exists $E_n \in \mathcal{A}_1^n$ such that $\mathsf{P}\{E \triangle E_n\} \to 0$ as $n \to \infty$. This implies

$$\mathsf{P}\{E_n\} \to \mathsf{P}\{E\}$$

and

$$\mathsf{P}\{E \cap E_n\} \to \mathsf{P}\{E\}$$

But $E \in \mathcal{A}_{n+1}^{\infty}$, So, E and E_n are in independent σ-fields. Thus

$$\mathsf{P}\{E \cap E_n\} = \mathsf{P}\{E\}\,\mathsf{P}\{E_n\}$$

Taking limits in this equations gives

$$\mathsf{P}\{E\} = \mathsf{P}^2\{E\}$$

that leads to the statement of this theorem. □

6.2.3 The Borel–Cantelli lemma

The results presented here deal with the same 'zero–one' law, but in terms of probabilities of current events, and therefore, provide much more wider possibilities of their applications.

Lemma 6.2. (Borel–Cantelli) *Let* $(\Omega, \mathcal{F}, \mathsf{P})$ *be a probability space and* $\{A_k, k \geq 1\}$ *is a sequence of events* $A_k \in \mathcal{F}$.

1. If

$$\boxed{\sum_{k=1}^{\infty} \mathsf{P}\{A_k\} < \infty} \tag{6.4}$$

then

$$\boxed{\mathsf{P}\{A_k \text{ i.o.}\} = 0} \tag{6.5}$$

or, equivalently,

$$\boxed{\mathsf{P}\left\{\omega : \sum_{k=1}^{\infty} \chi\{A_k\} < \infty\right\} = 1} \tag{6.6}$$

which means that ***there may occur only a finite number of events*** A_k.

2. If $\{A_k, k \geq 1\}$ *are* ***independent events*** *and*

$$\boxed{\sum_{k=1}^{\infty} \mathsf{P}\{A_k\} = \infty} \tag{6.7}$$

then

$$\boxed{\mathsf{P}\{A_k \text{ i.o.}\} = 1} \tag{6.8}$$

or, equivalently,

$$\boxed{\mathsf{P}\left\{\omega : \sum_{k=1}^{\infty} \chi\{A_k\} = \infty\right\} = 1} \tag{6.9}$$

which means that ***there may occur an infinite number of events*** A_k.

Proof.

1. By the definition (1.13) and in view of the property (1.29) it follows that

$$\begin{aligned}\mathsf{P}\{A_k \text{ i.o.}\} &= \mathsf{P}\left\{\limsup_{n\to\infty} A_n\right\} = \mathsf{P}\left\{\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\right\} \\ &\le \mathsf{P}\left\{\bigcup_{m=n}^{\infty} A_m\right\} \le \sum_{m=n}^{\infty} \mathsf{P}\{A_m\} \to 0 \text{ whereas } n \to \infty\end{aligned}$$

   that proves (6.5).
2. By independence of $\{A_k, k \ge 1\}$ we have

$$\begin{aligned}\mathsf{P}\{A_k \text{ i.o.}\} &= \mathsf{P}\left\{\limsup_{n\to\infty} A_n\right\} = \mathsf{P}\left\{\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\right\} \\ &\ge 1 - \mathsf{P}\left\{\bigcap_{m=n}^{\infty} A_m^c\right\} = 1 - \prod_{m=n}^{\infty} \mathsf{P}\left\{A_m^c\right\} = 1 - \prod_{m=n}^{\infty}\left(1 - \mathsf{P}\left\{A_m^c\right\}\right)\end{aligned}$$

   Applying the inequality $1 - x \le \exp(-x)$ valid for any $x \in \mathbb{R}$ yields

$$\begin{aligned}1 \ge \mathsf{P}\{A_k \text{ i.o.}\} &\ge 1 - \prod_{m=n}^{\infty}\left(1 - \mathsf{P}\left\{A_m^c\right\}\right) \\ \ge 1 - \prod_{m=n}^{\infty} \exp\left(-\mathsf{P}\left\{A_m^c\right\}\right) &= 1 - \exp\left(-\sum_{m=n}^{\infty} \mathsf{P}\left\{A_m^c\right\}\right) = 1 - 0\end{aligned}$$

   that completes the proof. □

Remark 6.3.

1. *In the case of independent events this lemma represents, in fact, a version of the 'zero–one' law since* $\mathsf{P}\{A_k \text{ i.o.}\}$ *may have only* 0 *or* 1 *values, and the convergence or divergence of the series* $\sum_{k=1}^{\infty} \mathsf{P}\{A_k\}$ *is the decisive factor.*
2. *From the computational point of view Lemma 6.2 turns out to be very useful since*
   - *it is sufficient to prove that* $\mathsf{P}\{A_k \text{ i.o.}\} > 0$ *to conclude that, in fact,* $\mathsf{P}\{A_k \text{ i.o.}\} = 1$;
   - *it is sufficient to prove that* $\mathsf{P}\{A_k \text{ i.o.}\} < 1$ *to conclude that, in fact,* $\mathsf{P}\{A_k \text{ i.o.}\} = 0$.

Corollary 6.1. *If* $\{A_k, k \ge 1\}$ *are arbitrary events,* $\left\{A_{k_s}, s \ge 1\right\}$ *is a subsequence of independent events and*

$$\boxed{\sum_{s=1}^{\infty} \mathsf{P}\left\{A_{k_s}\right\} = \infty}$$

then

$$\boxed{\mathsf{P}\{A_k \text{ i.o.}\} = 1}$$

or, equivalently,

$$\boxed{\mathsf{P}\left\{\omega : \sum_{k=1}^{\infty} \chi\left\{A_k\right\} = \infty\right\} = 1}$$

Proof. It follows from the fact that

$$\{A_k \text{ i.o.}\} \supset \{A_s \text{ i.o.}\}$$

and, as the result,

$$\mathsf{P}\left\{A_k \text{ i.o.}\right\} \geq \mathsf{P}\left\{A_{k_s} \text{ i.o.}\right\} = 1$$

□

Corollary 6.2. *Let* $\xi_1, \xi_2, \ldots$ *be a sequence of arbitrary random variables defined on* $(\Omega, \mathcal{F}, \mathsf{P})$.

1. *If for any* $\varepsilon > 0$

$$\boxed{\sum_{k=1}^{\infty} \mathsf{P}\left\{|\xi_k| > \varepsilon\right\} < \infty} \tag{6.10}$$

then with probability 1 (or, almost sure)

$$\boxed{\xi_k \overset{a.s.}{\to} 0} \tag{6.11}$$

2. *If* $\{A_k, k \geq 1\}$ *are independent events and there exists* $\varepsilon_0 > 0$ *such that*

$$\boxed{\sum_{k=1}^{\infty} \mathsf{P}\left\{|\xi_k| > \varepsilon\right\} = \infty} \tag{6.12}$$

then

$$\boxed{\xi_k \overset{a.s.}{\nrightarrow} 0} \tag{6.13}$$

Proof. It follows directly from Lemma 6.2 if we let

$$A_k := \{|\xi_k| > \varepsilon\}, \quad k = 1, 2, \ldots$$

In the first case we conclude that for any $\varepsilon > 0$ the number of the events $A_k := \{|\xi_k| > \varepsilon\}$ is finite with probability 1, which means the convergence to zero. In the second case, analogously, with probability 1 the events $A_k := \{|\xi_k| > \varepsilon\}$ occur infinitely many times, which means that $\{\xi_k\}$ does not converge to 0. □

6.3 Properties of Lebesgue integral with probabilistic measure

In this section we consider three fundamental theorems on 'taking limits' under the expectation sign, or in other words, under the Lebesgue integral with probabilistic measure. In fact, all three theorems which we are intending to discuss below can be found in Chapter 16 (Section 3) of Poznyak (2008) where they concern the Lebesgue integral with respect to a 'general' countably additive measure. Nevertheless, here we repeat the results discussed before using the 'probabilistic language' interpreting the Lebesgue integral as the mathematical expectation operator.

The following example shows that **it is not permitted in general to reverse the order of taking limit and computing an integral**, or in probabilistic terms, the mathematical expectation, that is, in general

$$\boxed{\lim_{n\to\infty} \mathsf{E}\{\xi_n\} \neq \mathsf{E}\left\{\lim_{n\to\infty} \xi_n\right\}} \tag{6.14}$$

Example 6.1. (Gut, 2005) *Let $\alpha > 0$, and consider a binary random variable*

$$\xi_n = \begin{cases} 0 & \text{with probability } 1 - n^{-2} \\ n^{\alpha} & \text{with probability } \quad n^{-2} \end{cases}, \quad n = 1, 2, \ldots$$

Notice that $\mathsf{P}\{\xi_n = 0\} \to 1$ as $n \to \infty$. Moreover, by the Borel–Cantelli Lemma 6.2 it follows that $\xi_n = n^{\alpha}$ only a finite number of times with probability one, since

$$\sum_{n=1}^{\infty} \mathsf{P}\left\{\xi_n = n^{\alpha}\right\} = \sum_{n=1}^{\infty} n^{-2} < \infty$$

Hence, $\xi_n \overset{a.s}{\to} 0$ as $n \to \infty$, and therefore

$$\mathsf{E}\left\{\lim_{n\to\infty} \xi_n\right\} = 0$$

On the other hand,

$$\lim_{n\to\infty} \mathsf{E}\{\xi_n\} = \lim_{n\to\infty} \left[0\left(1 - n^{-2}\right) + n^{\alpha}\left(n^{-2}\right)\right]$$
$$= \lim_{n\to\infty} n^{\alpha-2} = \begin{cases} 0 & \text{if } \alpha \in (0,2) \\ 1 & \text{if } \quad \alpha = 2 \\ \infty & \text{if } \quad \alpha > 2 \end{cases}$$

So the justification of the equality in (6.14) may vary and depends on the parameter α.

Below we discuss this problem in detail.

6.3.1 Lemma on monotone convergence

Lemma 6.3. (on monotone convergence) *Let η, $\{\xi_n\}_{n\geq 1}$ be random variables defined on $(\Omega, \mathcal{F}, \mathsf{P})$.*

1. *If*

$$\boxed{\xi_n \geq \eta \quad \textit{for all} \quad n \geq 1} \tag{6.15}$$

$$\boxed{\mathsf{E}\{\eta\} > -\infty} \tag{6.16}$$

and $\{\xi_n\}$ *monotonically (non-decreasing) converges to* ξ*, that is,*

$$\boxed{\xi_n \uparrow \xi} \tag{6.17}$$

then $\mathsf{E}\{\xi_n\}$ *also monotonically converges to* $\mathsf{E}\{\xi\}$*, that is,*

$$\boxed{\mathsf{E}\{\xi_n\} \uparrow \mathsf{E}\{\xi\}} \tag{6.18}$$

2. *If*

$$\boxed{\xi_n \leq \eta \quad \textit{for all} \quad n \geq 1} \tag{6.19}$$

$$\boxed{\mathsf{E}\{\eta\} < \infty} \tag{6.20}$$

and $\{\xi_n\}$ *monotonically (non-increasing) converges to* ξ*, that is,*

$$\boxed{\xi_n \downarrow \xi} \tag{6.21}$$

then $\mathsf{E}\{\xi_n\}$ *also monotonically converges to* $\mathsf{E}\{\xi\}$*, that is,*

$$\boxed{\mathsf{E}\{\xi_n\} \downarrow \mathsf{E}\{\xi\}} \tag{6.22}$$

Proof.

1. Suppose firstly that $\eta \geq 0$. By Theorem 2.3 for each $n \geq 1$ there exists a sequence $\{\xi_n^{(k)}\}_{k\geq 1}$ of simple random variables $\xi_n^{(k)}$ such that $\xi_n^{(k)} \underset{k\to\infty}{\uparrow} \xi_n$. Put $\zeta^{(k)} := \max\limits_{1\leq n\leq k} \xi_n^{(k)}$. Then

   $$\zeta^{(k-1)} \leq \zeta^{(k)} = \max_{1\leq n\leq k} \xi_n^{(k)} \leq \max_{1\leq n\leq k} \xi_n = \xi_k < \xi$$

   Let $\zeta := \lim\limits_{k\to\infty} \zeta^{(k)}$ which exists by the condition (6.17). Since

   $$\xi_n^{(k)} \leq \zeta^{(k)} \leq \xi_k$$

   for any $n \in [1, k]$, then taking limits as $k \to \infty$ we get

   $$\xi_n \leq \zeta \leq \xi$$

and, therefore, tending $n \to \infty$, implies $\zeta = \xi$. The random variables $\zeta^{(k)}$ are simple, nonnegative and $\zeta^{(k)} \underset{k\to\infty}{\uparrow} \zeta$. Therefore, by Definition (3.11),

$$\mathsf{E}\{\xi\} = \mathsf{E}\{\zeta\} = \lim_{k\to\infty} \mathsf{E}\left\{\zeta^{(k)}\right\} \le \lim_{k\to\infty} \mathsf{E}\{\xi_k\}$$

On the other hand, since $\xi_n \le \xi_{n+1} \le \xi$ it follows that

$$\lim_{n\to\infty} \mathsf{E}\{\xi_n\} \le \mathsf{E}\{\xi\}$$

that leads to the conclusion that $\lim_{n\to\infty} \mathsf{E}\{\xi_n\} = \mathsf{E}\{\xi\}$. Let now η be any random variable but satisfying (6.16). If $\mathsf{E}\{\eta\} = \infty$, then evidently $\mathsf{E}\{\xi_n\} = \mathsf{E}\{\xi\} = \infty$. Let $\mathsf{E}\{\eta\} < \infty$. Then $|\mathsf{E}\{\eta\}| < \infty$. Obviously

$$0 \le (\xi_n - \eta) \uparrow (\xi - \eta)$$

and therefore, by the previous consideration done for nonnegative variables,

$$\mathsf{E}\{\xi_n\}_n - \mathsf{E}\{\eta\} \uparrow \mathsf{E}\{\xi\} - \mathsf{E}\{\eta\}$$

Remembering that $|\mathsf{E}\{\eta\}| < \infty$, this gives $\mathsf{E}\{\xi_n\}_n \uparrow \mathsf{E}\{\xi\}$.

2. The proof of this item follows from the previous one by replacing the original variables with their negatives since $0 \le (2\xi - \xi_n) \uparrow \xi$. □

Corollary 6.3. *If* $\{\xi_n\}_{n\ge1}$ *are nonnegative random variables, then*

$$\boxed{\mathsf{E}\left\{\sum_{n=1}^{\infty} \xi_n\right\} = \sum_{n=1}^{\infty} \mathsf{E}\{\xi_n\}} \tag{6.23}$$

Proof. It follows from the fact that the sequence of partial sums $S_n := \sum_{k=1}^{n} \xi_k$ satisfies the conditions of Lemma 6.3. □

6.3.2 Fatou's lemma

Lemma 6.4. (Fatou's lemma) *Let* η, $\{\xi_n\}_{n\ge1}$ *be random variables defined on* $(\Omega, \mathcal{F}, \mathsf{P})$.

1. *If* $\xi_n \ge \eta$ *for all* $n \ge 1$, $\mathsf{E}\{\eta\} > -\infty$, *then*

$$\boxed{\mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\} \le \liminf_{n\to\infty} \mathsf{E}\{\xi_n\}} \tag{6.24}$$

2. *If* $\xi_n \le \eta$ *for all* $n \ge 1$, $\mathsf{E}\{\eta\} < \infty$, *then*

$$\boxed{\limsup_{n\to\infty} \mathsf{E}\{\xi_n\} \le \mathsf{E}\left\{\limsup_{n\to\infty} \xi_n\right\}} \tag{6.25}$$

3. *If* $|\xi_n| \leq \eta$ *for all* $n \geq 1$, $\mathsf{E}\{\eta\} < \infty$, *then*

$$\boxed{\begin{aligned}\mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\} \leq & \liminf_{n\to\infty} \mathsf{E}\{\xi_n\} \\ & \leq \limsup_{n\to\infty} \mathsf{E}\{\xi_n\} \leq \mathsf{E}\left\{\limsup_{n\to\infty} \xi_n\right\}\end{aligned}} \tag{6.26}$$

Proof.

1. Denote $\zeta_n := \inf_{m\geq n} \xi_m$. Then evidently

$$\liminf_{n\to\infty} \xi_n := \lim_{n\to\infty} \inf_{m\geq n} \xi_m = \lim_{n\to\infty} \zeta_n$$

It is also clear that $\zeta_n \uparrow \liminf_{n\to\infty} \xi_n$ and $\zeta_n \geq \eta$ for all $n \geq 1$. Therefore by Lemma 6.3 it follows that

$$\begin{aligned}\mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\} &= \mathsf{E}\left\{\lim_{n\to\infty} \zeta_n\right\} = \lim_{n\to\infty} \mathsf{E}\{\zeta_n\} \\ &\leq \liminf_{n\to\infty} \mathsf{E}\{\zeta_n\} \leq \liminf_{n\to\infty} \mathsf{E}\{\xi_n\}\end{aligned}$$

which establishes (6.24).
2. The relation (6.25) results from (6.24) if we take into account that $(2\eta - \xi_n) \geq \eta$ and, instead of ξ_n in item 1 consider $(2\eta - \xi_n)$.
3. The third item results from the first two. □

6.3.3 The Lebesgue dominated convergence theorem

Theorem 6.2. (the Lebesgue dominated convergence) *Let* η, $\{\xi_n\}_{n\geq 1}$ *be random variables defined on* $(\Omega, \mathcal{F}, \mathsf{P})$ *such that*

$$\boxed{|\xi_n| \leq \eta \quad \textit{for all} \quad n \geq 1, \quad \mathsf{E}\{\eta\} < \infty \quad \textit{and} \quad \xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi} \tag{6.27}$$

Then

$$\boxed{\mathsf{E}\{|\xi|\} < \infty, \quad \mathsf{E}\{\xi_n\} \to \mathsf{E}\{\xi\}} \tag{6.28}$$

and

$$\boxed{\mathsf{E}\{|\xi_n - \xi|\} \to 0} \tag{6.29}$$

as $n \to \infty$.

Proof. Since $|\xi_n| \leq \eta$ it follows that $|\xi_n - \xi| \leq 2\eta$. In view of the fact that by replacing ξ_n with $|\xi_n - \xi|$ we find that the proof reduces to showing the following fact: if $0 \leq \xi_n \leq \eta$

and $\xi_n \to 0$ almost surely, then $\mathsf{E}\{\xi_n\} \to 0$. But this follows from the Fatou's Lemma 6.4 (see (6.26)) since in this case

$$0 = \mathsf{E}\left\{\liminf_{n\to\infty} |\xi_n - \xi|\right\} = \mathsf{E}\left\{\lim_{n\to\infty} |\xi_n - \xi|\right\} = \lim_{n\to\infty} \mathsf{E}\{|\xi_n - \xi|\}$$
$$= \limsup_{n\to\infty} \mathsf{E}\{|\xi_n - \xi|\} = \mathsf{E}\left\{\limsup_{n\to\infty} |\xi_n - \xi|\right\} = 0$$

that proves the theorem. □

Corollary 6.4. *Let η, $\{\xi_n, n = 1, 2, \ldots\}$ be random variables such that*

$$\boxed{|\xi_n| \le \eta \quad \text{for all} \quad n \ge 1,\ \mathsf{E}\{\eta^p\} < \infty \quad \text{and} \quad \xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi} \tag{6.30}$$

for some $p > 0$. Then

$$\boxed{\mathsf{E}\left\{|\xi|^p\right\} < \infty, \quad \mathsf{E}\left\{|\xi_n - \xi|^p\right\} \underset{n\to\infty}{\to} 0} \tag{6.31}$$

Proof. The result follows from the relations

$$|\xi_n| \le \eta,\ |\xi_n - \xi|^p \le (|\xi_n| + |\xi|)^p \le (2\eta)^p$$

if instead of $|\xi_n|$ take $|\xi_n - \xi|^p$. □

6.3.4 Uniform integrability

In this subsection we will introduce a concept which permits us to use a somewhat '*weaker*' condition than $|\xi_n| \le \eta$ and $\mathsf{E}\{\eta\} < \infty$ to ensure the validity of the relations (6.26), (6.28) and (6.29). To do that, and following Shiryayev (1984), let us introduce one more definition.

Definition 6.3. *A family of random variables $\{\xi_n\}_{n\ge1}$ is said to be **uniformly integrable** if*

$$\boxed{\sup_n \int\limits_{\{\omega:|\xi_n|>c\}} |\xi_n(\omega)|\, \mathsf{P}(d\omega) = \sup_n \mathsf{E}\{|\xi_n|\, \chi(|\xi_n| > c)\} \to 0} \tag{6.32}$$

as $c \to \infty$.

Remark 6.4. *Obviously the conditions $|\xi_n| \le \eta$ $(n \ge 1)$ and $\mathsf{E}\{\eta\} < \infty$ imply the uniform integrability of the family $\{\xi_n\}_{n\ge1}$. Indeed,*

$$\sup_n \int\limits_{\{\omega:|\xi_n|>c\}} |\xi_n(\omega)|\, \mathsf{P}(d\omega) \le \sup_n \int\limits_{\{\omega:|\xi_n|>c\}} \eta(\omega)\, \mathsf{P}(d\omega)$$
$$= \sup_n \mathsf{E}\{\eta\chi(|\xi_n| > c)\} \le \mathsf{E}\{\eta\chi(\eta > c)\} \underset{c\to\infty}{\to} 0$$

The next theorem shows that both the Fatou Lemma 6.4 and the Lebesgue dominated convergence theorem 6.2 remain valid if instead of $|\xi_n| \leq \eta$ and $\mathsf{E}\{\eta\} < \infty$ they require only the uniform integrability property of the family $\{\xi_n\}_{n\geq 1}$.

Theorem 6.3. **(Shiryayev, 1984)** *Let $\{\xi_n\}_{n\geq 1}$ be a uniformly integrable family of random variables. Then the relations in (6.26) hold. If, additionally, $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$ then both properties (6.28) and (6.29) are valid too.*

Proof. For each $c > 0$ we have

$$\mathsf{E}\{\xi_n\} = \mathsf{E}\{\xi_n \chi(\xi_n < -c)\} + \mathsf{E}\{\xi_n \chi(\xi_n \geq -c)\} \tag{6.33}$$

By the assumed uniform integrability for any $\varepsilon > 0$ there exists c so large that

$$\sup_n |\mathsf{E}\{\xi_n \chi(\xi_n < -c)\}| < \varepsilon \tag{6.34}$$

By Fatou's Lemma 6.4 it follows that

$$\liminf_{n\to\infty} \mathsf{E}\{\xi_n \chi(\xi_n \geq -c)\} \geq \mathsf{E}\left\{\liminf_{n\to\infty} [\xi_n \chi(\xi_n \geq -c)]\right\}$$

But always

$$\xi_n \chi(\xi_n \geq -c) \geq \xi_n$$

and therefore

$$\liminf_{n\to\infty} \mathsf{E}\{\xi_n \chi(\xi_n \geq -c)\} \geq \mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\} \tag{6.35}$$

The relations (6.33)–(6.35) lead to the following inequality:

$$\liminf_{n\to\infty} \mathsf{E}\{\xi_n\} \geq \mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\} - \varepsilon$$

Since ε is arbitrarily small we obtain

$$\liminf_{n\to\infty} \mathsf{E}\{\xi_n\} \geq \mathsf{E}\left\{\liminf_{n\to\infty} \xi_n\right\}$$

The inequality

$$\liminf_{n\to\infty} \mathsf{E}\{\xi_n\} \leq \limsup_{n\to\infty} \mathsf{E}\{\xi_n\}$$

is evident by the definition. But the relation

$$\limsup_{n\to\infty} \mathsf{E}\{\xi_n\} \leq \mathsf{E}\left\{\limsup_{n\to\infty} \xi_n\right\}$$

may be proven similarly. Thus the relations in (6.26) are proven. The conclusions (6.28) and (6.29) are deduced from (6.26) as in Theorem 6.2. □

The significance of the uniform integrability concept is revealed by the following theorem which, in fact, is **the criterion for taking limits under the expectation operator**.

Theorem 6.4. *Let*

$$\boxed{0 \le \xi_n, \quad \mathsf{E}\{\xi_n\} < \infty}$$

and

$$\boxed{\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi}$$

Then

$$\boxed{\mathsf{E}\{\xi_n\} \underset{n\to\infty}{\to} \mathsf{E}\{\xi\}}$$

if and only if $\{\xi_n\}_{n\ge1}$ *is uniformly integrable.*

Proof. Notice, first, that *sufficiency* follows from the second part of Theorem 6.3. To prove *necessity* consider the set

$$A = \{a : \mathsf{P}\{\xi = a\} > 0\}$$

Then for any $a \notin A$

$$\xi_n \chi(\xi_n < a) \overset{a.s.}{\to} \xi\chi(\xi < a)$$

that implies

$$\mathsf{E}\{\xi_n \chi(\xi_n < a)\} \to \mathsf{E}\{\xi\chi(\xi < a)\}$$

For any $\varepsilon > 0$ there exists $a_0 \in A$ so large that $\mathsf{E}\{\xi\chi(\xi \ge a_0)\} < \varepsilon/2$. Choose N_0 also so large that

$$\mathsf{E}\{\xi_n \chi(\xi_n \ge a_0)\} \le \mathsf{E}\{\xi\chi(\xi \ge a_0/2)\} + \varepsilon/2$$

for all $n \ge N_0$. Then we get that

$$\mathsf{E}\{\xi_n \chi(\xi_n \ge a_0)\} \le \varepsilon$$

Choose then $a_1 \ge a_0$ so large that

$$\mathsf{E}\{\xi_n \chi(\xi_n \ge a_1)\} \le \varepsilon$$

for all $n \le N_0$. Hence, obviously, we have

$$\sup_n \mathsf{E}\{\xi_n \chi(\xi_n \ge a_1)\} \le \varepsilon$$

that proves the uniform integrability of $\{\xi_n\}_{n\ge1}$. □

The following proposition provides the simple test for checking the uniform integrability.

Theorem 6.5. (the uniform integrability test) *Let $\{\xi_n\}_{n\geq 1}$ be a family of absolutely integrable random variables and $G : \mathbb{R}_+ \to \mathbb{R}$ be a nonnegative increasing function such that*

$$\boxed{t^{-1}G(t) \to \infty \quad as \quad t \to \infty} \tag{6.36}$$

and

$$\boxed{M := \sup_n \mathsf{E}\{G(\xi_n)\} < \infty} \tag{6.37}$$

Then the family $\{\xi_n\}_{n\geq 1}$ is ***uniformly integrable****.*

Proof. For any $\varepsilon > 0$ one may define $a := M/\varepsilon$ and taking c so large that $t^{-1}G(t) \geq a$ for $t > c$, we derive

$$\begin{aligned} \mathsf{E}\{|\xi_n|\,\chi(|\xi_n| \geq c)\} &\leq \frac{1}{a}\mathsf{E}\{G(|\xi_n|)\,\chi(|\xi_n| \geq c)\} \\ &\leq \frac{1}{a}\mathsf{E}\{G(|\xi_n|)\} \leq \frac{1}{a}\sup_n \mathsf{E}\{G(\xi_n)\} \leq M/a \leq \varepsilon \end{aligned}$$

uniformly on $n \geq 1$. Theorem is proven. □

6.4 Convergence

6.4.1 Various modes of convergence

The following definition introduces **six principal convergence concepts** in probability theory.

Definition 6.4. *The sequence $\{\xi_n\}_{n\geq 1}$ of random variables, defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$,* ***converges*** *to a random variable ξ (also defined on $(\Omega, \mathcal{F}, \mathsf{P})$):*

1. ***in distribution****, that is denoted by*

$$\boxed{\xi_n \overset{d}{\underset{n\to\infty}{\to}} \xi} \tag{6.38}$$

if and only if

$$\boxed{\mathsf{E}\{f(\xi_n)\} \underset{n\to\infty}{\to} \mathsf{E}\{f(\xi)\}} \tag{6.39}$$

for any bounded continuous function $f(x)$, or equivalently, if and only if we have the convergence of distribution functions

$$\boxed{F_{\xi_n}(x) \underset{n\to\infty}{\to} F_{\xi}(x)} \tag{6.40}$$

at any point x where $F_{\xi}(x)$ is continuous.

2. ***in probability**, that is denoted by*

$$\boxed{\xi_n \overset{\mathsf{P}}{\underset{n\to\infty}{\to}} \xi} \tag{6.41}$$

if and only if for any $\varepsilon > 0$

$$\boxed{\mathsf{P}\{\omega : |\xi_n - \xi| > \varepsilon\} \underset{n\to\infty}{\to} 0} \tag{6.42}$$

3. ***with probability one** or **almost surely (a.s.)**, that is denoted by*

$$\boxed{\xi_n \overset{a.s.}{\underset{n\to\infty}{\to}} \xi} \tag{6.43}$$

if and only if

$$\boxed{\mathsf{P}\left\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\} = 1} \tag{6.44}$$

i.e., the set of sample points $\omega \in \Omega$ for which $\xi_n(\omega)$ does not converge to $\xi(\omega)$ has probability zero;

4. ***in mean of the order** p, that is denoted by*

$$\boxed{\xi_n \overset{L^p}{\underset{n\to\infty}{\to}} \xi} \tag{6.45}$$

if and only if for some $p > 0$

$$\boxed{\mathsf{E}\left\{|\xi_n - \xi|^p\right\} \underset{n\to\infty}{\to} 0} \tag{6.46}$$

*For $p = 2$ this is **mean-square convergence**, that is denoted by*

$$\boxed{\underset{n\to\infty}{\text{l.i.m.}} \xi_n = \xi} \tag{6.47}$$

5. ***pointly**, that is denoted by*

$$\boxed{\xi_n \overset{\forall\omega\in\Omega}{\underset{n\to\infty}{\to}} \xi} \tag{6.48}$$

if and only if

$$\boxed{\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega) \quad \text{for all} \quad \omega \in \Omega} \tag{6.49}$$

6. ***completely***, *that is denoted by*

$$\boxed{\xi_n \underset{n\to\infty}{\overset{\text{c.c.}}{\to}} \xi} \tag{6.50}$$

if and only if for any $\varepsilon > 0$

$$\boxed{\sum_{n=1}^{\infty} \mathsf{P}\{|\xi_n - \xi| > \varepsilon\} < \infty} \tag{6.51}$$

or equivalently, by the Borel–Cantelli Lemma 6.2, if and only if $|\xi_n - \xi| \le \varepsilon$ $(a.s.)$ *after a finite (maybe random) number* $n_0(\omega)$ *of steps.*[1]

6.4.2 Fundamental sequences

A set of random sequences represents in one or another sense a sequence of functions from some 'functional space'. But the analysis of a convergence in any functional space is closely related with the concept of a *fundamental* (or *Cauchy*) *sequence* (see, for example, Chapter 14 in Poznyak (2008)). So, as can be seen from below, it would be useful to introduce similar concepts for some of the convergence modes and to formulate the corresponding criteria of convergence using these concepts. Here we will do it for the first five kinds of convergence.

Definition 6.5. *We say that a sequence* $\{\xi_n\}_{n\ge1}$ *of random variables is a* ***fundamental (or Cauchy) sequence***

1. ***in distribution*** *if*

$$\boxed{|\xi_n - \xi_m| \underset{m,n\to\infty}{\overset{d}{\to}} 0} \tag{6.52}$$

2. ***in probability*** *if for any* $\varepsilon > 0$

$$\boxed{\mathsf{P}\{\omega : |\xi_n - \xi_m| > \varepsilon\} \underset{m,n\to\infty}{\to} 0} \tag{6.53}$$

[1] This concept was introduced in Hsu and Robbins (1947).

3. ***with probability one*** *or* ***almost surely (a.s.)*** *if* $\{\xi_n(\omega)\}_{n\geq 1}$ *is fundamental for almost all* $\omega \in \Omega$*, that is, when*

$$\boxed{|\xi_n - \xi_m| \underset{m,n\to\infty}{\overset{a.s.}{\to}} 0} \tag{6.54}$$

4. ***in*** L^P *sense if*

$$\boxed{\mathsf{E}\left\{|\xi_n - \xi_m|^p\right\} \underset{m,n\to\infty}{\to} 0} \tag{6.55}$$

5. ***pointly****, if*

$$\boxed{|\xi_n - \xi_m| \underset{m,n\to\infty}{\overset{\forall\omega\in\Omega}{\to}} 0} \tag{6.56}$$

The importance of the fundamental concept lies in the Cauchy criterion (see Theorem 14.8 in Poznyak (2008)):

(a) Every convergent sequence $\{x_n\}$ given in a metric space $\mathcal{X}$ is a Cauchy sequence.
(b) If $\mathcal{X}$ is a compact metric space and if $\{x_n\}$ is a Cauchy sequence in $\mathcal{X}$ then $\{x_n\}$ converges to some point in $\mathcal{X}$.
(c) In $\mathbb{R}^n$ a sequence converges if and only if it is a Cauchy sequence.

Since the space of random variables (for example with bounded second moments such that $\sum_{n=1}^{\infty} \mathsf{E}\left\{\xi_n^2\right\} < \infty$) is a Hilbert space with the scalar product $\left(\bar{\xi}, \bar{\eta}\right) := \sum_{n=1}^{\infty} \mathsf{E}\{\xi_n \eta_n\}$, then it becomes evident that the concept of a fundamental sequence serves as a principal instrument for the convergence criteria design. The theorem below illustrates this fact.

Theorem 6.6. **(Shiryayev, 1984)**

(a) A ***necessary and sufficient condition*** *that* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$ *is that for every* $\varepsilon > 0$

$$\boxed{\mathsf{P}\left\{\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\right\} \underset{n\to\infty}{\to} 0} \tag{6.57}$$

(b) The sequence $\{\xi_n\}_{n\geq 1}$ *of random variables is a* ***fundamental (or Cauchy) sequence with probability one*** *if and only if for every* $\varepsilon > 0$

$$\boxed{\mathsf{P}\left\{\sup_{l,k\geq n} |\xi_k - \xi_l| \geq \varepsilon\right\} \underset{n\to\infty}{\to} 0} \tag{6.58}$$

or equivalently,

$$\boxed{\mathsf{P}\left\{\sup_{k\geq 0}|\xi_{n+k}-\xi_n|\geq\varepsilon\right\}\underset{n\to\infty}{\to}0} \tag{6.59}$$

Proof.

(a) Define two sets

$$A_n^{\varepsilon}=\{\omega:|\xi_n-\xi|\geq\varepsilon\},\quad A^{\varepsilon}=\limsup_{n\to\infty}A_n^{\varepsilon}:=\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}A_k^{\varepsilon}$$

Then one has

$$\{\omega:\xi_n\nrightarrow\xi\}=\bigcup_{\varepsilon>0}A^{\varepsilon}=\bigcup_{m=1}^{\infty}A^{1/m}$$

$$\mathsf{P}\left\{A^{\varepsilon}\right\}=\lim_{n\to\infty}\mathsf{P}\left\{\bigcup_{k\geq n}A_k^{\varepsilon}\right\}$$

The necessity in (a) follows from the following chain of implications:

$$0=\mathsf{P}\{\omega:\xi_n\nrightarrow\xi\}=\mathsf{P}\left\{\bigcup_{\varepsilon>0}A^{\varepsilon}\right\}=\mathsf{P}\left\{\bigcup_{m=1}^{\infty}A^{1/m}\right\}$$

and therefore, $\mathsf{P}\left\{A^{1/m}\right\}=0$ $(m\geq 1)$, that gives $\mathsf{P}\{A^{\varepsilon}\}=0$, and hence,

$$\mathsf{P}\left\{\bigcup_{k\geq n}A_k^{\varepsilon}\right\}\underset{n\to\infty}{\to}0$$

that implies

$$\mathsf{P}\left\{\sup_{k\geq n}|\xi_k-\xi|\geq\varepsilon\right\}\underset{n\to\infty}{\to}0$$

The sufficiency results from the back consideration.

(b) Define now the sets

$$B_{k,t}^{\varepsilon}:=\{\omega:|\xi_k-\xi_l|\geq\varepsilon\},\quad B^{\varepsilon}:=\bigcap_{n=1}^{\infty}\bigcup_{l,k\geq n}B_{k,l}^{\varepsilon}$$

Then

$$\left\{\omega:\{\xi_n(\omega)\}_{n\geq 1}\text{ is not fundamental}\right\}=\bigcup_{\varepsilon>0}B^{\varepsilon}$$

Repeating the considerations as in (a) one can show that

$$\mathsf{P}\left\{\omega : \{\xi_n(\omega)\}_{n\geq 1} \text{ is not fundamental}\right\} = 0$$

that implies (6.58). The conclusion (6.59) follows from the inequality

$$\sup_{k\geq 0} |\xi_{n+k} - \xi_n| \leq \sup_{l,k\geq 0} |\xi_{n+k} - \xi_{n+l}|$$

that completes the proof. □

Corollary 6.5. (the relation with complete convergence) *A* ***sufficient condition*** *that* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$ *is that it* ***completely converges****, namely,*

$$\sum_{n=1}^{\infty} \mathsf{P}\{|\xi_n - \xi| > \varepsilon\} < \infty$$

for every $\varepsilon < 0$.

Proof. It follows from the relations

$$\mathsf{P}\left\{\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\right\} = \mathsf{P}\left\{\omega : \bigcup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\right\} \leq \sum_{k\geq n} \mathsf{P}\{|\xi_n - \xi| > \varepsilon\}$$ □

6.4.3 Distributional convergence

Definition 6.6.

(a) The ***variational distance between the distributions*** *F and G is*

$$\boxed{d(F, G) := \sup_{A\in\mathbb{R}} |F(A) - G(A)|} \tag{6.60}$$

(b) The ***distributional distance between two random variables*** ξ *and* η*, both defined on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*, is*

$$\boxed{d(\xi, \eta) := \sup_{A\in\mathbb{R}} |\mathsf{P}\{\xi \in A\} - \mathsf{P}\{\eta \in A\}|} \tag{6.61}$$

(c) If ξ *and* $\{\xi_n\}_{n\geq 1}$ *are random variables such that*

$$\boxed{d(\xi_n, \xi) \underset{n\to\infty}{\to} 0} \tag{6.62}$$

we say that ξ_n ***converges*** *to* ξ ***in total variation*** *as* $n \to \infty$.

Lemma 6.5. *If* ξ_n *converges to* ξ *in total variation then* $\xi_n \underset{n\to\infty}{\overset{d}{\to}} \xi$.

Proof. Indeed,

$$\begin{aligned}\left|F_{\xi_n}(x) - F_{\xi}(x)\right| &= \left|\mathsf{P}\{\xi_n \le x\} - \mathsf{P}\{\xi \le x\}\right| \\ &\le \sup_{x\in\mathbb{R}} \left|\mathsf{P}\{\xi_n \le x\} - \mathsf{P}\{\xi \le x\}\right|\end{aligned}$$

which establishes the desired result. □

The next theorem deals with convergence in probability for absolutely continuous random sequences.

Proposition 6.1. (Scheffé's Lemma) *Suppose $\{\xi_n\}_{n\ge1}$ is a sequence of absolutely continuous random variables with density functions $\left\{f_{\xi_n}\right\}_{n\ge1}$. Then*

$$\boxed{d(\xi_n, \xi) \le \int_{-\infty}^{\infty} \left|f_{\xi_n}(x) - f_{\xi}(x)\right| dx} \tag{6.63}$$

and, hence, if $f_{\xi_n}(x) \underset{n\to\infty}{\to} f_{\xi}(x)$ almost everywhere, then

$$d(\xi_n, \xi) \underset{n\to\infty}{\to} 0$$

and in particular,

$$\xi_n \overset{d}{\underset{n\to\infty}{\to}} \xi$$

Proof. By the triangle inequality

$$\begin{aligned}\left|\mathsf{P}\{\xi_n \le x\} - \mathsf{P}\{\xi \le x\}\right| &= \left|\int_{-\infty}^{x} \left[f_{\xi_n}(x) - f_{\xi}(x)\right] dx\right| \\ &\le \int_{-\infty}^{x} \left|f_{\xi_n}(x) - f_{\xi}(x)\right| dx \le \int_{-\infty}^{\infty} \left|f_{\xi_n}(x) - f_{\xi}(x)\right| dx\end{aligned}$$

Applying $\sup_{x\in\mathbb{R}}$ to both sides of the last inequality proves (6.63). The convergence follows automatically from Lemma 6.5. □

6.4.4 Relations between convergence concepts

In this subsection we will discuss the relations between the convergence concepts defined before.

Theorem 6.7. *The following implications hold (see Fig. 6.2).*

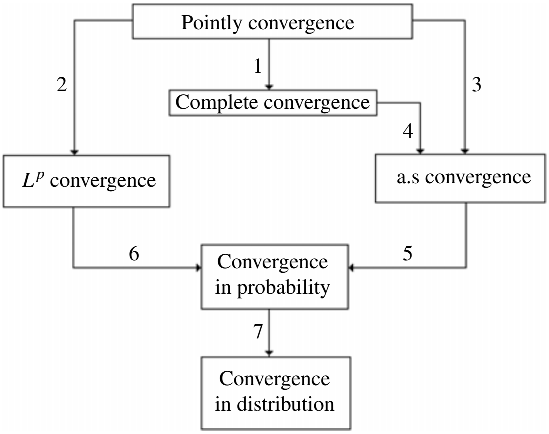

Fig. 6.2. Various modes of convergence and their interconnection.

Proof. The implications 1–3 are trivial. 4 follows from Corollary 6.5. 5 results from the relations

$$\lim_{n\to\infty} \mathsf{P}\{|\xi_n - \xi| > \varepsilon\} = 1 - \lim_{n\to\infty} \mathsf{P}\{|\xi_n - \xi| \le \varepsilon\} = 1 - 1 = 0$$

The implication 6 follows from the Markov inequality (4.8), namely,

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon > 0\} \le \varepsilon^{-p}\mathsf{E}\left\{|\xi_n - \xi|^p\right\}$$

To prove 7 let us consider a function $f(x)$ which is continuous and bounded: $|f(x)| \le C$. Let ε and N be such that $\mathsf{P}\{|\xi_n| > N\} \le \varepsilon/4C$. Taking $\delta > 0$ satisfying

$$|f(x) - f(y)| \le \varepsilon/2C \quad \text{for any} \quad x : |x| \le N, |x - y| \le \delta$$

we obtain

$$\begin{aligned}
\mathsf{E}\{|f(\xi_n) - f(\xi)|\} = & \int\limits_{\{\omega:|\xi_n(\omega)-\xi(\omega)|\le\delta,\ |\xi_n(\omega)|\le N\}} |f(\xi_n) - f(\xi)|\,\mathsf{P}(d\omega) \\
& + \int\limits_{\{\omega:|\xi_n(\omega)-\xi(\omega)|\le\delta,\ |\xi_n(\omega)|> N\}} |f(\xi_n) - f(\xi)|\,\mathsf{P}(d\omega) \\
& + \int\limits_{\{\omega:|\xi_n(\omega)-\xi(\omega)|>\delta\}} |f(\xi_n) - f(\xi)|\,\mathsf{P}(d\omega) \\
& \le \varepsilon/2C + (\varepsilon/2C)(\varepsilon/4C) + 2C\mathsf{P}\{|\xi_n - \xi| > \delta\} \\
& = \varepsilon \text{ Const} + o(1)
\end{aligned}$$

Since ε is arbitrary the required implication follows. □

The following two examples clearly show that $a.s.$-convergence does not imply L^p-convergence, and, inversely, L^p-convergence does not obligatorily imply $a.s.$-convergence.

Example 6.2. *Here we will show that even though* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$, *nevertheless* $\xi_n \underset{n\to\infty}{\overset{L^P}{\nrightarrow}} \xi$. *Let*

$$\xi_n(\omega) := \begin{cases} n^2 & \text{if } \omega \in \left[0, \frac{1}{n}\right] \\ 0 & \text{if } \omega \in \left(\frac{1}{n}, 1\right] \end{cases}, \quad \omega \in \Omega = [0,1]$$

Evidently

$$\mathsf{P}\left\{\omega : \lim_{n\to\infty} \xi_n(\omega) = 0\right\} = 1$$

But

$$\mathsf{E}\left\{|\xi_n|^p\right\} = n^{2p}\left(\frac{1}{n}\right) + 0\left(1 - \frac{1}{n}\right) = n^{2p-1} \underset{n\to\infty}{\nrightarrow} 0$$

if

$$p \geq 1/2$$

Example 6.3. *This example shows that even though* $\xi_n \underset{n\to\infty}{\overset{L^P}{\to}} \xi$, *nevertheless* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\nrightarrow}} \xi$. *Let*

$$\xi_n(\omega) := \begin{cases} 1 & \text{with probability} \quad p_n \underset{n\to\infty}{\to} 0 \\ 0 & \text{with probability} \quad q_n := 1 - p_n \end{cases}$$

and $\{\xi_n\}_{n\geq 1}$ *be a sequence of independent random variables. Then*

$$\mathsf{E}\left\{|\xi_n|^p\right\} = 1^p p_n + 0^p q_n = p_n \underset{n\to\infty}{\to} 0$$

But, on the other hand, by the Borel–Cantelli Lemma, 6.2, if

$$\sum_{n=1}^{\infty} p_n = \infty$$

then

$$\mathsf{P}\left\{\omega : \sum_{n=1}^{\infty} \chi\left(\xi_n(\omega) = 1\right) = \infty\right\} = 1$$

which means that 'almost always' there exists an infinite subsequence $\xi_{n_k}(\omega) = 1$ $(k \geq 1)$, *or equivalently,* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\nrightarrow}} \xi$.

6.4.5 Some converses

Under some additional conditions there exist the converses to some of the arrows in Theorem 6.7. Below we will present some of them.

6.4.5.1 When p-mean convergence implies a.s. convergence

Theorem 6.8. *If* $\xi_n \underset{n\to\infty}{\overset{L^p}{\to}} \xi$ $(p > 0)$ *such that*

$$\boxed{\sum_{n=1}^{\infty} \mathsf{E}\left\{|\xi_n - \xi|^p\right\} < \infty} \tag{6.64}$$

then $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$.

Proof. Using Markov's inequality (4.8) it follows that

$$\sum_{n=1}^{n^*} \mathsf{P}\left\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon > 0\right\}$$
$$\leq \sum_{n=1}^{n^*} \varepsilon^{-p} \mathsf{E}\left\{|\xi_n - \xi|^p\right\} \leq \varepsilon^{-p} \sum_{n=1}^{\infty} \mathsf{E}\left\{|\xi_n - \xi|^p\right\} < \infty$$

Taking then $n^* \to \infty$ we get

$$\sum_{n=1}^{n^*} \mathsf{P}\left\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon > 0\right\} < \infty$$

and hence, by the Borel–Cantelli Lemma, 6.2

$$\mathsf{P}\left\{\omega : \sum_{n=1}^{\infty} \chi\left(|\xi_n(\omega) - \xi(\omega)| \geq \varepsilon > 0\right) < \infty\right\} = 1$$

In other words, for any $\varepsilon > 0$ there exists $n_0 = n_0(\omega)$ such that for all $n \geq n_0(\omega)$

$$|\xi_n(\omega) - \xi(\omega)| \overset{a.s.}{<} \varepsilon$$

which exactly means that $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$. Theorem is proven. □

Theorem 6.9. *If* $0 \leq \xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} \xi$ *and*

$$\boxed{\mathsf{E}\{\xi_n\} \underset{n\to\infty}{\to} \mathsf{E}\{\xi\}} \tag{6.65}$$

then

$$\boxed{\mathsf{E}\{|\xi_n - \xi|\} \underset{n\to\infty}{\to} 0} \tag{6.66}$$

that is, $\xi_n \underset{n\to\infty}{\overset{L^1}{\to}} \xi$.

Proof. In view of the identity

$$|a-b| = b-a+2\,[a-b]\,\chi\,(a \geq b)$$

valid for any $a, b \in \mathbb{R}$, it follows that

$$\mathsf{E}\{|\xi_n - \xi|\} = \mathsf{E}\{\xi - \xi_n\} + 2\mathsf{E}\{(\xi_n - \xi)\,\chi\,(\xi_n \geq \xi)\} \tag{6.67}$$

But

$$0 \leq (\xi_n - \xi)\,\chi\,(\xi_n \geq \xi) := \zeta_n$$

and, by lemma on monotone convergence 6.3, it follows that

$$0 = \mathsf{E}\left\{\lim_{n\to\infty} \zeta_n\right\} = \lim_{n\to\infty} \mathsf{E}\{\zeta_n\}$$

So, taking limits in (6.67) and in view of (6.65), we obtain

$$\lim_{n\to\infty} \mathsf{E}\{|\xi_n - \xi|\} = \lim_{n\to\infty} \mathsf{E}\{\xi - \xi_n\} + 2\lim_{n\to\infty} \mathsf{E}\{\zeta_n\} = 0$$

that completes the proof. □

Exercise 6.1. *Let*

$$\xi_n(\omega) = \begin{cases} 1 & \textit{with probability} \quad p_n \leq 1/2 \\ 0 & \textit{with probability} \quad 1-2p_n \\ -1 & \textit{with probability} \quad p_n \leq 1/2 \end{cases}$$

and

$$p_n = \frac{1}{2n^\alpha}, \quad \alpha \geq 0, n = 1, 2, \ldots$$

Define for which α *we have* $\xi_n \underset{n\to\infty}{\overset{L^2}{\to}} 0$ *and for which* $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$.

Solution

(a) One has

$$\mathsf{E}\left\{\xi_n^2\right\} = 1^2 p_n + 0^2(1-2p_n) + (-1)^2 p_n = 2p_n = 1/n^\alpha \underset{n\to\infty}{\to} 0$$

for any $\alpha > 0$.

(b) But

$$\sum_{n=1}^{\infty} \mathsf{E}\left\{\xi_n^2\right\} = \sum_{n=1}^{\infty} 1/n^\alpha < \infty$$

for any $\alpha > 1$, which by Theorem 6.8 implies $\xi_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. □

6.4.5.2 When convergence in probability implies convergence in p-mean: uniform integrability

If one knows that there takes place the convergences in probability then a natural question is whether there exist some additional conditions guaranteeing the convergence in p-mean. The following concept turns out to be adequate to answer correctly this question.

Definition 6.7. *A sequence $\{\xi_n\}_{n\geq 1}$ of random variables is called* ***uniformly integrable*** *if*

$$\boxed{\epsilon_n(a) := \mathsf{E}\{|\xi_n| \chi(|\xi_n| > a)\} = \int\limits_{|x|>a}^{\infty} |x| \, dF_{\xi_n}(x) \underset{a\to\infty}{\to} 0} \tag{6.68}$$

uniformly in n.

The requirement that a sequence is uniformly integrable exactly means that the contributions in the tails of the integrals tend to zero uniformly for all members of this sequence.

Claim 6.1. *If a sequence $\{\xi_n\}_{n\geq 1}$ is uniformly integrable then*

$$\boxed{\mathsf{E}\{|\xi_n|\} \leq a + \epsilon_n(a)}$$

for a large enough ($\epsilon_n(a) \underset{a\to\infty}{\to} 0$ uniformly in n), and, hence, the first absolute moments are uniformly bounded.

Proof. Indeed,

$$\begin{aligned}\mathsf{E}\{|\xi_n|\} &= \mathsf{E}\{|\xi_n| \chi(|\xi_n| \leq a)\} + \mathsf{E}\{|\xi_n| \chi(|\xi_n| > a)\} \\ &\leq a\mathsf{P}\{\omega : |\xi_n(\omega)| \leq a\} + \epsilon_n(a) \leq a + \epsilon_n(a)\end{aligned}$$

which proves the claim. □

The following theorem gives the criterion of uniform integrability.

Theorem 6.10. (the criterion of uniform integrability) *The sequence $\{\xi_n\}_{n\geq 1}$ of random variables is uniformly integrable* ***if and only if***

(a)

$$\sup_n \mathsf{E}\{|\xi_n|\} < \infty \tag{6.69}$$

(b) for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon)$ such that for any set A with the measure $\mathsf{P}\{A\} < \delta$

$$\mathsf{E}\{|\xi_n| \chi(A)\} < \varepsilon \tag{6.70}$$

uniformly in n.

Proof.

Necessity. Let $\{\xi_n\}_{n\geq 1}$ be uniformly integrable. By Claim 6.1

$$\mathsf{E}\{|\xi_n|\} \leq a + \epsilon_n(a) \leq a + \sup_n \epsilon_n(a) < \infty$$

Taking $\sup_n$ in the left-hand side of this inequality implies (6.69). So, the necessity of (a) is proven. Let now $\varepsilon > 0$ be given and a set A be such that $\mathsf{P}\{A\} < \delta$. Then

$$\begin{aligned}\mathsf{E}\{|\xi_n|\chi(A)\} &= \mathsf{E}\{|\xi_n|\chi(A) \cap \chi(|\xi_n| \leq a)\} + \mathsf{E}\{|\xi_n|\chi(A) \cap \chi(|\xi_n| > a)\} \\ &\leq a\mathsf{P}\{\omega : |\xi_n(\omega)| \leq a\} + \mathsf{E}\{|\xi_n|\chi(|\xi_n| > a)\} \leq a\delta + \sup_n \epsilon_n(a) \leq \varepsilon\end{aligned}$$

if one chooses a large enough to make $\mathsf{E}\{|\xi_n|\chi(|\xi_n| > a)\} \leq \varepsilon/2$ with δ ensuring $a\delta \leq \varepsilon/2$ that proves (b).

Sufficiency. Suppose that both conditions (6.69) and (6.70) of the theorem are fulfilled. Define the sets

$$A_n := \{\omega : |\xi_n(\omega)| > a\}$$

Then by Markov's inequality (4.8) it follows that

$$\mathsf{P}\{A_n\} \leq a^{-1}\mathsf{E}\{|\xi_n|\} \leq a^{-1}\sup_n \mathsf{E}\{|\xi_n|\} < \delta$$

uniformly in n for sufficiently large a that by (6.70) shows that

$$\mathsf{E}\{|\xi_n|\chi(|\xi_n| > a)\} = \mathsf{E}\{|\xi_n|\chi(A_n)\} < \varepsilon$$

uniformly in n. This establishes the uniform integrability. □

In fact this is not a simple task for verifying uniform integrability for some given sequences. The theorem below gives the convenient and constructive condition sufficient for guaranteeing uniform integrability.

Theorem 6.11. *Let $\{\xi_n\}_{n\geq 1}$ be a sequence of random variables and $g : \mathbb{R} \to \mathbb{R}$ be a non-negative increasing function such that*

$$\boxed{\frac{g(t)}{t} \to \infty \quad as \ t \to \infty} \tag{6.71}$$

If

$$\boxed{\sup_n \mathsf{E}\{g(\xi_n)\} < \infty} \tag{6.72}$$

*then $\{\xi_n\}_{n\geq 1}$ is **uniformly integrable**.*

Proof. By the assumption of the theorem $\frac{g(t)}{t} > b$ for all $t > a$ where $a = a(b)$ is large enough. Hence, given $\varepsilon > 0$,

$$\begin{aligned}\mathsf{E}\{|\xi_n|\,\chi(|\xi_n| \le a)\} &\le b^{-1}\mathsf{E}\{g(\xi_n)\,\chi(|\xi_n| \le a)\} \\ &\le b^{-1}\mathsf{E}\{g(\xi_n)\} \le b^{-1}\sup_n \mathsf{E}\{g(\xi_n)\} < \varepsilon\end{aligned}$$

independently on ε for a large enough b that proves uniform integrability. □

Corollary 6.6. *If for some* $p > 1$

$$\boxed{\sup_n \mathsf{E}\left\{|\xi_n|^p\right\} < \infty} \tag{6.73}$$

then $\{\xi_n\}_{n\ge1}$ *is* ***uniformly integrable***.

Proof. It results from Theorem 6.11 if we let $g(t) := |t|^p$. □

Now we are ready to answer the question when the convergences in probability guarantee the convergence in p-mean.

Theorem 6.12. *Let* ξ *and* $\{\xi_n\}_{n\ge1}$ *be random variables. Suppose that* $\xi_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} \xi$ *and* $p > 0$. *If* $\left\{|\xi_n|^p\right\}_{n\ge1}$ *is uniformly integrable, that is,*

$$\mathsf{E}\left\{|\xi_n|^p\,\chi(|\xi_n| > a)\right\} \underset{a\to\infty}{\to} 0$$

uniformly in n, then $\xi_n \underset{n\to\infty}{\overset{L^p}{\to}} \xi$ *and* $\mathsf{E}\left\{|\xi_n|^p\right\} \underset{n\to\infty}{\to} \mathsf{E}\left\{|\xi|^p\right\}$.

Proof. First notice that if $\left\{|\xi_n|^p\right\}_{n\ge1}$ is uniformly integrable then $\left\{|\xi_n - \xi|^p\right\}_{n\ge1}$ is uniformly integrable too. Indeed, by the inequality (4.30),

$$|\xi_n - \xi|^p \le 2^p\left(|\xi_n|^p + |\xi|^p\right)$$

and hence,

$$\begin{aligned}|\xi_n - \xi|^p\,\chi(|\xi_n - \xi| > a) &\le 2^p\left(|\xi_n|^p + |\xi|^p\right)\chi(|\xi_n - \xi| > a) \\ &\le 2^p \max\left\{|\xi_n|^p, |\xi|^p\right\}\chi(2\max\{|\xi_n|, |\xi|\} > a) \\ &\le 2^p\left[|\xi_n|^p\,\chi(|\xi_n| > a/2) + |\xi|^p\,\chi(|\xi| > a/2)\right]\end{aligned}$$

Taking expectation and letting $a \to \infty$ implies

$$\mathsf{E}\left\{|\xi_n - \xi|^p\,\chi(|\xi_n - \xi| > a)\right\} \underset{a\to\infty}{\to} 0$$

uniformly in n. Hence one has

$$\begin{aligned}&\mathsf{E}\left\{|\xi_n - \xi|^p\right\} = \mathsf{E}\left\{|\xi_n - \xi|^p\,\chi(|\xi_n - \xi| \le \varepsilon)\right\} \\ &\quad + \mathsf{E}\left\{|\xi_n - \xi|^p\,\chi(|\xi_n - \xi| > \varepsilon)\right\} \le \varepsilon^p + \mathsf{E}\left\{|\xi_n - \xi|^p\,\chi(|\xi_n - \xi| > \varepsilon)\right\}\end{aligned}$$

By Theorem 6.10 and by the assumption on uniform integrability we have

$$\mathsf{E}\left\{|\xi_n - \xi|^p \chi\left(|\xi_n - \xi| > \varepsilon\right)\right\} \underset{n\to\infty}{\to} 0$$

that means

$$\limsup_{n\to\infty} \mathsf{E}\left\{|\xi_n - \xi|^p\right\} \le \varepsilon^p$$

By the arbitrariness of ε we have $\xi_n \overset{L^p}{\underset{n\to\infty}{\to}} \xi$. The convergence $\mathsf{E}\left\{|\xi_n|^p\right\} \underset{n\to\infty}{\to} \mathsf{E}\left\{|\xi|^p\right\}$ follows from the inequality

$$\left|\mathsf{E}\left\{|\xi_n|^p\right\} - \mathsf{E}\left\{|\xi|^p\right\}\right| \le \mathsf{E}\left\{|\xi_n - \xi|^p\right\} \underset{n\to\infty}{\to} 0$$

that proves the theorem. □

6.4.5.3 When a.s. convergence implies convergence in p-mean

Here also the concept of uniform integrability works.

Theorem 6.13. *Let ξ and $\{\xi_n\}_{n\ge1}$ be random variables. Suppose that $\xi_n \overset{\text{a.s.}}{\underset{n\to\infty}{\to}} \xi$ and $p > 0$. If $\left\{|\xi_n|^p\right\}_{n\ge1}$ is uniformly integrable, then $\xi_n \overset{L^p}{\underset{n\to\infty}{\to}} \xi$ and $\mathsf{E}\left\{|\xi_n|^p\right\} \underset{n\to\infty}{\to} \mathsf{E}\left\{|\xi|^p\right\}$.*

Proof. It follows directly from the previous theorem if we take into account that $a.s.$ convergence implies convergence in probability. □

This page intentionally left blank

7 Martingales

Contents

The term '*martingale*' in a non-mathematical context means '*a horse's harness*'. It also independently originated in *Gambling Theory*.[1] The first appearance of a martingale as a mathematical term was due to Ville (1939). The major breakthrough was related in the classical book '*Stochastic Processes*' (Doob, 1953). Other more recent books are by Nevue (1975), Williams (1991) and Liptser and Shiryayev (1989). Martingales are probably the most inventive and generalized sums of independent random variables with zero-mean. In some sense they occupy the intermediate place between independent and dependent sequences. The independence assumption has proved inadequate for handling contemporary developments in many fields. On the other hand, relevant martingales can almost always be constructed, for example by devices such as centering by subtracting conditional mathematical expectations given by past and then summing. Below we will discuss this construction in detail. But first, to do that mathematically rigorously, we should start with a *conditional mathematical expectation* introduction which is a corner-stone in Martingale Theory.

7.1 Conditional expectation relative to a sigma-algebra

7.1.1 Main definition

In this section we will introduce the so-called, conditional mathematical expectations with respect to *σ-algebras* which naturally arise whenever one needs to consider mathematical expectations in relation to increasing information patterns.

Definition 7.1. *A random variable* $\eta = \eta(\omega)$*, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$ *and considered within the equivalence class of random variables,*[2] *is called the* ***conditional mathematical expectation*** *of a random variable* $\xi = \xi(\omega)$ *(also defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*)* ***relative to a sub-σ-algebra*** $\mathcal{F}_0 \subseteq \mathcal{F}$ *if*

[1] The famous 'gambling strategy' is to double one's stake as long as one loses and leaves as soon as one wins. This strategy was often called *a martingale*. Unfortunately, the gambler will have spent an infinite amount of money 'on average' when he (she) finally wins.

[2] The *equivalence class* of a random variable is the collection of random variables that differ from this variable on a null set.

1. *$\eta(\omega)$ is $\mathcal{F}_0$-measurable, i.e., for any $c \in \mathbb{R}$*

$$\boxed{\{\omega : \eta(\omega) \leq c\} \in \mathcal{F}_0} \tag{7.1}$$

2. *for any $A \in \mathcal{F}_0$*

$$\boxed{\int_A \eta(\omega)\, \mathsf{P}(d\omega) = \int_A \xi(\omega)\, \mathsf{P}(d\omega)} \tag{7.2}$$

The conditional mathematical expectation $\eta(\omega)$ is denoted by

$$\boxed{\eta(\omega) = \mathsf{E}\{\xi/\mathcal{F}_0\}} \tag{7.3}$$

7.1.2 Some properties of conditional expectation

7.1.2.1 Basic properties

The lemma below shows that practically all properties valid for usual (complete) mathematical expectation remain valid for conditional expectations.

Lemma 7.1. *Let ξ and θ be integrable random variables, $\mathcal{F}_0 \subset \mathcal{F}$ and c, c_1, c_2 be real numbers. The following properties for $\mathsf{E}\{\xi/\mathcal{F}_0\}$ hold:*

1. *If $\xi \overset{a.s.}{=} c$, then*

$$\boxed{\mathsf{E}\{\xi/\mathcal{F}_0\} \overset{a.s.}{=} c} \tag{7.4}$$

2. *If $\xi \overset{a.s.}{\leq} \theta$, then*

$$\boxed{\mathsf{E}\{\xi/\mathcal{F}_0\} \overset{a.s.}{\leq} \mathsf{E}\{\theta/\mathcal{F}_0\}} \tag{7.5}$$

3.

$$\boxed{|\mathsf{E}\{\xi/\mathcal{F}_0\}| \overset{a.s.}{\leq} \mathsf{E}\{|\xi|/\mathcal{F}_0\}} \tag{7.6}$$

4.

$$\boxed{\mathsf{E}\{c_1\xi + c_2\theta/\mathcal{F}_0\} \overset{a.s.}{=} c_1\mathsf{E}\{\xi/\mathcal{F}_0\} + c_2\mathsf{E}\{\theta/\mathcal{F}_0\}} \tag{7.7}$$

5. *If $\xi \in \mathcal{F}_0$ (ξ is measurable with respect to $\mathcal{F}_0$), then*

$$\boxed{\mathsf{E}\{\xi/\mathcal{F}_0\} \overset{a.s.}{=} \xi} \tag{7.8}$$

6. *If $\mathcal{F}_0 = \mathcal{F}_* := \{\emptyset, \Omega\}$ is the* ***trivial*** *sigma-algebra, then*

$$\boxed{\mathsf{E}\{\xi/\mathcal{F}_*\} \overset{a.s.}{=} \mathsf{E}\{\xi\}} \tag{7.9}$$

7.

$$\boxed{\mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_0\}\} = \mathsf{E}\{\xi\}} \tag{7.10}$$

8. *If $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$, then the following* ***smoothing property*** *holds:*

$$\boxed{\begin{aligned} \mathsf{E}\{\xi/\mathcal{F}_0\} &\overset{a.s.}{=} \mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_0\}/\mathcal{F}_1\} \\ &\overset{a.s.}{=} \mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_1\}/\mathcal{F}_0\} \end{aligned}} \tag{7.11}$$

9. *If $\theta \in \mathcal{F}_0$, and such that $\mathsf{E}\{|\theta|\} < \infty$, $\mathsf{E}\{|\theta\xi|\} < \infty$, then*

$$\boxed{\mathsf{E}\{\theta\xi/\mathcal{F}_0\} \overset{a.s.}{=} \theta\mathsf{E}\{\xi/\mathcal{F}_0\}} \tag{7.12}$$

Proof.

1. As c is always measurable with respect to any sigma algebra $\mathcal{F}_0$, it is sufficient to check that the property (7.2):

$$\int_A \xi(\omega)\,\mathsf{P}(d\omega) = \int_A c\mathsf{P}(d\omega)$$

which is true for any $A \subset \mathcal{F}_0$ since $\xi \overset{a.s.}{=} c$.

2. As $\xi \overset{a.s.}{\leq} \theta$ then for any $A \subset \mathcal{F}_0$

$$\int_A \xi(\omega)\,\mathsf{P}(d\omega) \leq \int_A \theta(\omega)\,\mathsf{P}(d\omega)$$

and hence, by (7.2),

$$\begin{aligned} \int_A \mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega) &= \int_A \xi(\omega)\,\mathsf{P}(d\omega) \\ &\leq \int_A \theta(\omega)\,\mathsf{P}(d\omega) = \int_A \mathsf{E}\{\theta/\mathcal{F}_0\}\,\mathsf{P}(d\omega) \end{aligned}$$

Since this inequality is valid for any $A \subset \mathcal{F}_0$ we obtain (7.5).

3. Taking into account the representation (2.12) we have

$$\begin{aligned} |\mathsf{E}\{\xi/\mathcal{F}_0\}| &= \left|\mathsf{E}\left\{\xi^+/\mathcal{F}_0\right\} - \mathsf{E}\left\{\xi^-/\mathcal{F}_0\right\}\right| \\ &\overset{a.s.}{\leq} \left|\mathsf{E}\left\{\xi^+/\mathcal{F}_0\right\} + \mathsf{E}\left\{\xi^-/\mathcal{F}_0\right\}\right| = |\mathsf{E}\{|\xi|/\mathcal{F}_0\}| = \mathsf{E}\{|\xi|/\mathcal{F}_0\} \end{aligned}$$

that proves (7.6).

4. The formula (7.7) follows from (7.2) and the identity

$$\begin{aligned}\int_A \mathsf{E}\{c_1\xi + c_2\theta/\mathcal{F}_0\}\,\mathsf{P}(d\omega) &\overset{a.s.}{=} \int_A (c_1\xi + c_2\theta)\,\mathsf{P}(d\omega)\\ &= c_1\int_A \xi\mathsf{P}(d\omega) + c_2\int_A \theta\mathsf{P}(d\omega)\\ &= c_1\int_A \mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega) + c_2\int_A \mathsf{E}\{\theta/\mathcal{F}_0\}\,\mathsf{P}(d\omega)\end{aligned}$$

   valid for any $A \subset \mathcal{F}_0$.
5. If $\xi \in \mathcal{F}_0$ then by (7.2)

$$\int_A \mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega) = \int_A \xi(\omega)\,\mathsf{P}(d\omega)$$

   for any $A \subset \mathcal{F}_0$ that means (7.8).
6. By (7.2) it follows that for any $A \subset \mathcal{F}_* = \{\emptyset, \Omega\}$

$$\begin{aligned}\int_A \mathsf{E}\{\xi/\mathcal{F}_*\}\,\mathsf{P}(d\omega) &\overset{a.s.}{=} \int_A \xi(\omega)\,\mathsf{P}(d\omega)\\ &= \underbrace{\int_{A:A=\emptyset} \xi(\omega)\,\mathsf{P}(d\omega)}_{0} + \int_{A:A=\Omega} \xi(\omega)\,\mathsf{P}(d\omega)\\ &= \int_{A:A=\Omega} \xi(\omega)\,\mathsf{P}(d\omega) = \mathsf{E}\{\xi\}\end{aligned}$$

   and, since the constant $\mathsf{E}\{\xi\}$ is measurable with respect to any $\mathcal{F}_0$ (in particular with respect to $\mathcal{F}_*$), applying the property (7.8) we get

$$\begin{aligned}\int_A \mathsf{E}\{\xi\}\,\mathsf{P}(d\omega) &= \int_A \left\{\int_A \mathsf{E}\{\xi/\mathcal{F}_*\}\,\mathsf{P}(d\omega)\right\}\mathsf{P}(d\omega)\\ &= \left\{\int_A \mathsf{E}\{\xi/\mathcal{F}_*\}\,\mathsf{P}(d\omega)\right\}\end{aligned}$$

   that implies (7.9).
7. It follows from (7.2) setting $A = \Omega$.
8. Since $\mathsf{E}\{\xi/\mathcal{F}_0\} \in \mathcal{F}_1$ the first equality in (7.11) follows directly from the property (7.8). Let $A \in \mathcal{F}_0$; then evidently $A \in \mathcal{F}_1$. Applying then (7.2) three times we obtain $(a.s.)$

for any $A \in \mathcal{F}_0 \subseteq \mathcal{F}_1$

$$\int_{A\in\mathcal{F}_0\subseteq\mathcal{F}_1} \mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_1\}/\mathcal{F}_0\}\,\mathsf{P}(d\omega) = \int_{A\in\mathcal{F}_0\subseteq\mathcal{F}_1} \mathsf{E}\{\xi/\mathcal{F}_1\}\,\mathsf{P}(d\omega)$$
$$= \int_{A\in\mathcal{F}_0\subseteq\mathcal{F}_1} \xi\,\mathsf{P}(d\omega) = \int_{A\in\mathcal{F}_0\subseteq\mathcal{F}_1} \mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega)$$

that leads to the second identity in (7.11).

9. Suppose, first, that both θ and ξ are nonnegative. If additionally we assume that θ is a simple variable, namely, $\theta = \chi(\Lambda)$ where $\Lambda \in \mathcal{F}_0$, then by (7.2) we get

$$\int_{A\in\mathcal{F}_0} \theta\mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega) \overset{a.s.}{=} \int_{A\cap\Lambda\in\mathcal{F}_0} \mathsf{E}\{\xi/\mathcal{F}_0\}\,\mathsf{P}(d\omega)$$
$$\overset{a.s.}{=} \int_{A\cap\Lambda\in\mathcal{F}_0} \xi\,\mathsf{P}(d\omega) \overset{a.s.}{=} \int_{A\in\mathcal{F}_0} \theta\xi\,\mathsf{P}(d\omega)$$
$$= \int_{A\in\mathcal{F}_0} \mathsf{E}\{\theta\xi\}/\mathcal{F}_0\mathsf{P}(d\omega)$$

which corresponds to (7.12). Then, if $\{\theta_n, n \geq 1\}$ is simple random variables such that $\theta_n \uparrow \theta$ almost surely whenever $n \to \infty$, then one has $(a.s.)$

$$\theta_n\xi \uparrow \theta\xi \text{ and } \theta_n\mathsf{E}\{\xi/\mathcal{F}_0\} \uparrow \theta\mathsf{E}\{\xi/\mathcal{F}_0\}$$

which, by Theorem 2.3 on monotone approximation, implies (7.12). The general case follows from the decompositions (2.12), i.e., $\theta = \theta^+ - \theta^-$ and $\xi = \xi^+ - \xi^-$. □

Remark 7.1. *By the property (7.6),* $\mathsf{E}\{\xi/\mathcal{F}_0\}$ *exists if* ξ *is* ***integrable****, i.e., when* $\mathsf{E}\{|\xi|\} < \infty$.

7.1.2.2 Sigma-algebras generated by a sequence of measurable data

Definition 7.2. *A sigma-algebra*

$$\boxed{\mathcal{F}_n := \sigma(x_1, x_2, \ldots, x_n)}$$

constructed from all sets

$$\boxed{\{\omega : x_i(\omega) \leq c_i, \quad i = 1, 2, \ldots, n\}}$$

where $x_i = x_i(\omega)$ $(i = 1, 2, \ldots, n)$ *are random variables, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*, and* $c_i \in \mathbb{R}$ *are any constants, is called* ***the sigma-algebras generated by a sequence of measurable data*** $(x_1, x_2, \ldots, x_n)$.

Evidently

$$\boxed{\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n} \tag{7.13}$$

and $\mathsf{E}\{\xi/\mathcal{F}_n\} \in \mathcal{F}_n$ can be treated as a *conditional expectation under the given prehistory* $(x_1, x_2, \ldots, x_n)$.

Example 7.1. *Consider the recursion*

$$x_{n+1} = Ax_n + \xi_n, \quad n = 1, 2, \ldots, n_f$$

where $\left\{\xi_n, 1 \leq n \leq n_f\right\}$ *is a sequence of integrable random vectors from* $\mathbb{R}^n$ *independent of* $(x_1, x_2, \ldots, x_n)$*,* $A \in \mathbb{R}^{n\times n}$ *is a deterministic matrix and* x_1 *is an integrable random variable.*

(a) If $\mathcal{F}_n = \sigma(x_1, x_2, \ldots, x_n)$ *then*

$$\mathsf{E}\{x_{n+1}/\mathcal{F}_n\} \overset{a.s.}{=} \mathsf{E}\{Ax_n + \xi_n/\mathcal{F}_n\}$$
$$\overset{a.s.}{=} A\mathsf{E}\{x_n/\mathcal{F}_n\} + \mathsf{E}\{\xi_n/\mathcal{F}_n\} \overset{a.s.}{=} Ax_n + \mathsf{E}\{\xi_n\}$$

since $x_n \in \mathcal{F}_n$ *and* ξ_n *is independent on* $(x_1, x_2, \ldots, x_n)$*;*

(b) If $\mathcal{F}_n = \sigma(x_1, \xi_1, \xi_2, \ldots, \xi_n)$ *then*

$$\mathsf{E}\{x_{n+1}/\mathcal{F}_n\} \overset{a.s.}{=} \mathsf{E}\{Ax_n + \xi_n/\mathcal{F}_n\} \overset{a.s.}{=} Ax_n + \xi_n$$

since $x_n = Ax_{n-1} + \xi_{n-1} \in \mathcal{F}_n$ *and* $\xi_n \in \mathcal{F}_n$*.*

7.2 Martingales and related concepts

In this section we will study a class of random sequences for which the dependence property is defined in terms of conditional mathematical expectations.

7.2.1 Martingales, submartingales and supermartingales

Definition 7.3. *A sequence of random variables* $\{x_n\}_{n\geq 1}$*, given on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*, is said to be* ***adapted to a sequence of increasing*** σ***-algebras*** $\{\mathcal{F}_n\}_{n\geq 1}$*, or, in other words, to the* ***filtration (flow)***

$$\boxed{\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots} \tag{7.14}$$

if x_n *is* $\mathcal{F}_n$ *measurable, i.e.,* $x_n \in \mathcal{F}_n$ *for every* $n = 1, 2, \ldots$*. The sequence* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *of the pairs is called a* ***stochastic sequence*** *and* $\{\mathcal{F}_n\}_{n\geq 1}$ *– a* ***natural filtration****.*

Definition 7.4. *If a stochastic sequence* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *has the property*

$$\boxed{x_n \in \mathcal{F}_{n-1}} \tag{7.15}$$

for every $n = 1, 2, \ldots$ *then it is called a* ***predictable sequence****.*

Here is the main definition of this chapter.

Definition 7.5. *A stochastic sequence* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *of absolutely integrable random variables (such that* $\mathsf{E}\{|\ x_n\ |\} \overset{a.s.}{<} \infty$ *for all* $n \geq 1$*) is called*

(a) a ***martingale*** *if for all* $n \geq 1$

$$\boxed{\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{=} x_n} \tag{7.16}$$

(b) a ***submartingale*** *if*

$$\boxed{\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\geq} x_n} \tag{7.17}$$

(c) a ***supermartingale*** *if the sequence* $\{-x_n, \mathcal{F}_n\}_{n\geq 1}$ *is a submartingale, that is, if*

$$\boxed{E\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\leq} x_n} \tag{7.18}$$

(d) a ***martingale-difference*** *if for all* $n \geq 1$

$$\boxed{\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{=} 0} \tag{7.19}$$

Remark 7.2. *Sometimes, the sequences* $\{x_n\}_{n\geq 1}$ *themselves (but not only* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$*) are called martingales, submartingales, supermartingales or martingale-difference if the properties (7.16), (7.17), (7.18) and (7.19) are fulfilled, respectively.*

Based on the properties (see Lemma 7.1) of the conditional mathematical expectation it follows that the **definition equivalent to the previous one** is as given below.

Definition 7.6. *A stochastic sequence* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *of absolutely integrable random variables (such that* $\mathsf{E}\{|\ x_n\ |\} \overset{a.s.}{<} \infty$ *for all* $n \geq 1$*) is called*

(a) a ***martingale*** *if for all* $A \subset \mathcal{F}_n$ *and all* $n \geq 1$

$$\boxed{\int_A x_{n+1}(\omega)\,\mathsf{P}(d\omega) = \int_A x_n(\omega)\,\mathsf{P}(d\omega)} \tag{7.20}$$

or, equivalently,

$$\boxed{\int_A x_n(\omega)\,\mathsf{P}(d\omega) = \int_A x_m(\omega)\,\mathsf{P}(d\omega), \quad 0 \leq m \leq n} \tag{7.21}$$

(b) *a **submartingale** if for all $A \subset \mathcal{F}_n$ and all $n \geq 1$*

$$\boxed{\int_A x_{n+1}(\omega)\,\mathsf{P}(d\omega) \geq \int_A x_n(\omega)\,\mathsf{P}(d\omega)} \tag{7.22}$$

or, equivalently,

$$\boxed{\int_A x_n(\omega)\,\mathsf{P}(d\omega) \geq \int_A x_m(\omega)\,\mathsf{P}(d\omega), \quad 0 \leq m \leq n} \tag{7.23}$$

(c) *a **supermartingale** if the sequence $\{-x_n, \mathcal{F}_n\}_{n\geq 1}$ is a submartingale, that is, if for all $A \subset \mathcal{F}_n$ and all $n \geq 1$*

$$\boxed{\int_A x_{n+1}(\omega)\,\mathsf{P}(d\omega) \leq \int_A x_n(\omega)\,\mathsf{P}(d\omega)} \tag{7.24}$$

or, equivalently,

$$\boxed{\int_A x_n(\omega)\,\mathsf{P}(d\omega) \leq \int_A x_m(\omega)\,\mathsf{P}(d\omega), \quad 0 \leq m \leq n} \tag{7.25}$$

(d) *a **martingale-difference** if for all $A \subset \mathcal{F}_n$ and all $n \geq 1$*

$$\boxed{\int_A x_n(\omega)\,\mathsf{P}(d\omega) = 0} \tag{7.26}$$

7.2.2 *Some examples*

Example 7.2.

(a) *Consider a sequence $\{\xi_n\}_{n\geq 1}$ of independent integrable random variables ξ_n with zero-mean. Define $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ and*

$$S_n := \sum_{t=1}^{n} \xi_t \tag{7.27}$$

Then $\{S_n, \mathcal{F}_n\}_{n\geq 1}$ is a martingale. Indeed,

$$\begin{aligned} \mathsf{E}\{S_{n+1} \mid \mathcal{F}_n\} &= \mathsf{E}\{S_n + x_{n+1} \mid \mathcal{F}_n\} \\ &\overset{a.s.}{=} S_n + \mathsf{E}\{x_{n+1}\} = S_n \end{aligned}$$

(b) If additionally $\mathsf{E}\left\{\xi_t^2\right\} = \sigma_t^2$ $(t = 1, \ldots, n)$ *then* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *with*

$$x_n := S_n^2 - s_n^2, \; s_n^2 := \sum_{t=1}^{n} \sigma_t^2$$

is a martingale too since

$$\begin{aligned} \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} &= \mathsf{E}\left\{S_{n+1}^2 - s_{n+1}^2 \mid \mathcal{F}_n\right\} \\ &= \mathsf{E}\left\{S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2 - s_n^2 - \sigma_{n+1}^2 \mid \mathcal{F}_n\right\} \\ &\overset{a.s.}{=} \left(S_n^2 - s_n^2\right) + 2S_n\mathsf{E}\{\xi_{n+1} \mid \mathcal{F}_n\} + E\left\{\xi_{n+1}^2 \mid \mathcal{F}_n\right\} - \sigma_{n+1}^2 \\ &= x_n + 2S_n\mathsf{E}\{\xi_{n+1}\} + \mathsf{E}\left\{\xi_{n+1}^2\right\} - \sigma_{n+1}^2 = x_n \end{aligned}$$

(c) $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *with*

$$x_n := \frac{e^{tS_n}}{\prod_{k=1}^{n} \psi_k(t)} = \prod_{k=1}^{n} \frac{e^{t\xi_k}}{\psi_k(t)}, \quad \psi_k(t) := E\left\{e^{t\xi_k}\right\}, \quad t \in \mathbb{R}$$

is called an ***exponential martingale*** *since*

$$\begin{aligned} \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} &= \mathsf{E}\left\{\prod_{k=1}^{n+1}\left(\frac{e^{t\xi_k}}{\psi_k(t)}\right) \mid \mathcal{F}_n\right\} \\ &= \mathsf{E}\left\{\prod_{k=1}^{n}\left(\frac{e^{t\xi_k}}{\psi_k(t)}\right)\frac{e^{t\xi_{n+1}}}{\psi_{n+1}(t)} \mid \mathcal{F}_n\right\} \overset{a.s.}{=} x_n\mathsf{E}\left\{\frac{e^{t\xi_{n+1}}}{\psi_{n+1}(t)} \mid \mathcal{F}_n\right\} \\ &= x_n\mathsf{E}\left\{\frac{e^{t\xi_{n+1}}}{\psi_{n+1}(t)}\right\} = x_n \end{aligned}$$

Example 7.3. *If* $\{\xi_n, \mathcal{F}_n\}_{n\geq 1}$ *is a martingale-difference and* $\{y_n\}_{n\geq 1}$ *is* $\mathcal{F}_n$*-predictable, i.e.,* $y_n \in \mathcal{F}_{n-1}$*, then* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *with*

$$x_n := \sum_{t=1}^{n} \xi_t y_t$$

is a martingale because of the relation

$$\begin{aligned} \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} &= \mathsf{E}\{x_n + \xi_{n+1}y_{n+1} \mid \mathcal{F}_n\} \\ &\overset{a.s.}{=} x_n + y_{n+1}\mathsf{E}\{\xi_{n+1} \mid \mathcal{F}_n\} = x_n \end{aligned}$$

Example 7.4. *If* $\left\{\xi_n', \mathcal{F}_n\right\}_{n\geq 1}$ *and* $\left\{\xi_n'', \mathcal{F}_n\right\}_{n\geq 1}$ *are both submartingales (martingales or supermartingales) then* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *where*

$$x_n = c'\xi_n' + c''\xi_n'' \quad \text{with} \quad c', c'' \geq 0$$

is a submartingale (martingale or supermartingale). Indeed, in the case of submartingales we have

$$\begin{aligned}\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} &= \mathsf{E}\left\{c'\xi'_{n+1} + c''\xi''_{n+1} \mid \mathcal{F}_n\right\} \\ &= c'\mathsf{E}\left\{\xi'_{n+1} \mid \mathcal{F}_n\right\} + c''\mathsf{E}\left\{\xi''_{n+1} \mid \mathcal{F}_n\right\} \\ &\overset{a.s.}{\geq} c'\xi'_n + c''\xi''_n = x_n\end{aligned}$$

Other cases (martingales or supermartingales) are proved analogously.

Example 7.5. *If* $\left\{\xi'_n, \mathcal{F}_n\right\}_{n\geq 1}$ *and* $\left\{\xi''_n, \mathcal{F}_n\right\}_{n\geq 1}$ *are both martingales then* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *is*

(a) a submartingale if

$$x_n := \max\left\{\xi'_n, \xi''_n\right\}$$

(b) a supermartingale if

$$x_n := \min\left\{\xi'_n, \xi''_n\right\}$$

Indeed, since

$$\max\left\{\xi'_{n+1}, \xi''_{n+1}\right\} \geq \xi'_{n+1},\ \max\left\{\xi'_{n+1}, \xi''_{n+1}\right\} \geq \xi''_{n+1}$$

it follows that

$$\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} = \mathsf{E}\left\{\max\left\{\xi'_{n+1}, \xi''_{n+1}\right\} \mid \mathcal{F}_n\right\} \geq \mathsf{E}\left\{\xi'_{n+1} \mid \mathcal{F}_n\right\} \overset{a.s.}{=} \xi'_n$$

and

$$\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} = \mathsf{E}\left\{\max\left\{\xi'_{n+1}, \xi''_{n+1}\right\} \mid \mathcal{F}_n\right\} \geq \mathsf{E}\left\{\xi''_{n+1} \mid \mathcal{F}_n\right\} \overset{a.s.}{=} \xi''_n$$

Therefore,

$$\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\geq} \max\left\{\xi'_n, \xi''_n\right\} = x_n$$

Analogously, for $x_n := \min\left\{\xi'_n, \xi''_n\right\}$ *it follows*

$$\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\leq} \min\left\{\xi'_n, \xi''_n\right\} = x_n$$

Example 7.6. *If* $g_{\cup} : \mathbb{R} \to \mathbb{R}$ *is a convex function and* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *is a martingale such that* $\mathsf{E}\{|g_{\cup}(x_n)|\} < \infty$ *then* $\{g_{\cup}(x_n), \mathcal{F}_n\}_{n\geq 1}$ *is a submartingale since, by the Jensen inequality, it follows that*

$$\mathsf{E}\{g_{\cup}(x_{n+1}) \mid \mathcal{F}_n\} \overset{a.s.}{\geq} g_{\cup}(E\{x_{n+1} \mid \mathcal{F}_n\}) \overset{a.s.}{=} g_{\cup}(x_n)$$

For example, this is true if

$$g_{\cup}(x) = |x|^r, \quad r \geq 1 \tag{7.28}$$

7.2.3 Decompositions of submartingales and quadratic variation

7.2.3.1 Doob's decomposition

Theorem 7.1. (Doob, 1953) *Any submartingale $\{x_n, \mathcal{F}_n\}_{n\ge 0}$ can be uniquely decomposed into a sum of a martingale $\{M_n, \mathcal{F}_n\}_{n\ge 0}$ and a predictable non-decreasing process $\{A_n, \mathcal{F}_{n-1}\}_{n\ge 0}$ ($\mathcal{F}_{-1} := \mathcal{F}_0$, $A_0 = 0$), i.e.,*

$$\boxed{\begin{array}{ll} x_n = M_n + A_n, & n \ge 0 \\ A_n \in \mathcal{F}_{n-1}, & A_{n+1} \ge A_n \end{array}} \tag{7.29}$$

Proof. Since $A_0 = 0$ let $M_0 = x_0$ and

$$\begin{aligned} M_n &= M_0 + \sum_{j=0}^{n-1} \left(x_{j+1} - \mathsf{E}\left\{x_{j+1} \mid \mathcal{F}_j\right\}\right) \\ A_n &= x_n - M_n = \sum_{j=0}^{n-1} \left(\mathsf{E}\left\{x_{j+1} \mid \mathcal{F}_j\right\} - x_j\right) \end{aligned} \tag{7.30}$$

Evidently $\{M_n, \mathcal{F}_n\}_{n\ge 0}$ is a martingale. Indeed,

$$\mathsf{E}\{M_n \mid \mathcal{F}_{n-1}\} = \mathsf{E}\{M_{n-1} + (x_n - \mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\}) \mid \mathcal{F}_{n-1}\} \overset{a.s.}{=} M_{n-1}$$

It is evident also that $\{A_n, \mathcal{F}_{n-1}\}_{n\ge 0}$ is a predictable process since

$$A_n = \sum_{j=0}^{n-1} \left(\mathsf{E}\left\{x_{j+1} \mid \mathcal{F}_j\right\} - x_j\right) \in \mathcal{F}_{n-1}$$

Let us prove that $\{A_n\}_{n\ge 0}$ is a non-decreasing sequence. One has

$$\begin{aligned} A_{n+1} - A_n &= (x_{n+1} - x_n) - (M_{n+1} - M_n) \\ &= (x_{n+1} - x_n) - (x_{n+1} - \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\}) \\ &= \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} - x_n \overset{a.s.}{\ge} 0 \end{aligned}$$

by the submartingale property. This proves the existence of the decomposition. It remains to prove the uniqueness. Suppose that there exists another decomposition $x_n = M'_n + A'_n$. Then

$$\begin{aligned} A'_{n+1} - A'_n &= \mathsf{E}\left\{A'_{n+1} - A'_n \mid \mathcal{F}_n\right\} \\ &= \mathsf{E}\left\{(x_{n+1} - x_n) - \left(M'_{n+1} - M'_n\right) \mid \mathcal{F}_n\right\} \\ &= \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} - x_n - \left(\mathsf{E}\left\{M'_{n+1} \mid \mathcal{F}_n\right\} - M'_n\right) \\ &= \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} - x_n = A_{n+1} - A_n \end{aligned}$$

Taking into account that $A_0 = A'_0 = 0$ we obtain the uniqueness. Theorem is proven. □

Remark 7.3. *From the decomposition (7.29) it follows that the sequence $\{A_n, \mathcal{F}_{n-1}\}_{n\ge 0}$ compensates $\{x_n, \mathcal{F}_n\}_{n\ge 0}$ so that it becomes a martingale $\{M_n, \mathcal{F}_n\}_{n\ge 0}$. That is why $\{A_n, \mathcal{F}_{n-1}\}_{n\ge 0}$ is called a predictable* ***compensator*** *of the submartingale $\{x_n, \mathcal{F}_n\}_{n\ge 0}$.*

7.2.3.2 Quadratic variation of a martingale

Consider a square integrable martingale $\{M_n, \mathcal{F}_n\}_{n\geq 0}$ ($\mathsf{E}\{M_n\} < \infty, n \geq 0$). Notice that the sequence $\left\{M_n^2, \mathcal{F}_n\right\}_{n\geq 0}$ is a submartingale since, by the Jensen inequality,

$$\mathsf{E}\left\{M_{n+1}^2 \mid \mathcal{F}_n\right\} \overset{a.s.}{\geq} (\mathsf{E}\{M_{n+1} \mid \mathcal{F}_n\})^2 = M_n^2$$

Therefore, by Theorem 7.1 it follows that there exists a martingale $\{m_n, \mathcal{F}_n\}_{n\geq 0}$ such that

$$\boxed{M_n^2 = m_n + \langle M \rangle_n} \tag{7.31}$$

where the sequence $\left\{\langle M \rangle_n, \mathcal{F}_{n-1}\right\}_{n\geq 0}$ is a predictable non-decreasing sequence.

Definition 7.7. *The predictable non-decreasing sequence $\left\{\langle M \rangle_n, \mathcal{F}_{n-1}\right\}_{n\geq 0}$ satisfying (7.31) is called the* ***quadratic variation*** *of $\{M_n, \mathcal{F}_n\}_{n\geq 0}$.*

This definition is justified by the following property.

Lemma 7.2. *For a square integrable martingale $\{M_n, \mathcal{F}_n\}_{n\geq 0}$ one has*

$$\boxed{\begin{gathered}\langle M \rangle_n = \sum_{j=1}^{n} \mathsf{E}\left\{\left(\Delta M_j\right)^2 \mid \mathcal{F}_{j-1}\right\} \\ \Delta M_j := M_j - M_{j-1}\end{gathered}} \tag{7.32}$$

Proof. Indeed, from (7.30) it follows that

$$\begin{aligned}\langle M \rangle_n &= \sum_{j=0}^{n-1}\left(\mathsf{E}\left\{M_{j+1}^2 \mid \mathcal{F}_j\right\} - M_j^2\right) = \sum_{j=0}^{n-1}\mathsf{E}\left\{M_{j+1}^2 - M_j^2 \mid \mathcal{F}_j\right\} \\ &= \sum_{j=0}^{n-1}\mathsf{E}\left\{M_{j+1}^2 - 2M_{j+1}M_j + M_j^2 \mid \mathcal{F}_j\right\} = \sum_{j=0}^{n-1}\mathsf{E}\left\{\left(M_{j+1} - M_j\right)^2 \mid \mathcal{F}_j\right\}\end{aligned}$$

which proves the desired result. □

Example 7.7. *If $M_n = \sum_{t=1}^{n} \xi_t$ $(M_0 = 0)$ where $\{\xi_n\}_{n\geq 1}$ is square integrable $\left(\mathsf{E}\left\{\xi_t^2\right\} = \sigma_t^2 < \infty\right)$ independent zero-mean ($\mathsf{E}\{\xi_t\} = 0$) random variable, then*

$$\boxed{\langle M \rangle_n = \sum_{t=1}^{n} \sigma_t^2} \tag{7.33}$$

Indeed,

$$\begin{aligned}\langle M\rangle_n &= \sum_{j=0}^{n-1} \mathsf{E}\left\{M_{j+1}^2 - M_j^2 \mid \mathcal{F}_j\right\} = \sum_{j=0}^{n-1}\left(\mathsf{E}\left\{M_{j+1}^2\right\} - \mathsf{E}\left\{M_j^2\right\}\right) \\ &= \mathsf{E}\left\{M_n^2\right\} - \mathsf{E}\left\{M_0^2\right\} = \mathsf{E}\left\{M_n^2\right\} = \sum_{t=1}^{n} \mathsf{E}\left\{\xi_t^2\right\} = \sum_{t=1}^{n} \sigma_t^2\end{aligned}$$

Example 7.8. *If* $M_n = \sum_{t=1}^{n} \xi_t$ $(M_0 = 0)$ *and* $N_n = \sum_{t=1}^{n} \eta_t$ $(N_0 = 0)$ *where* $\{\xi_n\}_{n\geq 1}$ *and* $\{\eta_n\}_{n\geq 1}$ *are square integrable martingale-differences, that is,*

$$\begin{aligned}&\mathsf{E}\{\xi_t \mid \mathcal{F}_{t-1}\} = \mathsf{E}\{\eta_t \mid \mathcal{F}_{t-1}\} \overset{a.s.}{=} 0 \\ &\mathsf{E}\left\{\xi_t^2 \mid \mathcal{F}_{t-1}\right\} \overset{a.s.}{<} \infty, \quad \mathsf{E}\left\{\eta_t^2 \mid \mathcal{F}_{t-1}\right\} \infty\end{aligned}$$

then for the random variable $\langle M, N\rangle$ *defined by*

$$\boxed{\langle M, N\rangle := \frac{1}{4}\left(\langle M+N\rangle_n - \langle M-N\rangle_n\right)} \tag{7.34}$$

the following property holds:

$$\boxed{\langle M, N\rangle = \sum_{t=1}^{n} \mathsf{E}\{\xi_t\eta_t \mid \mathcal{F}_{t-1}\}} \tag{7.35}$$

Indeed, using the relations of the previous example, it follows

$$\begin{aligned}\langle M, N\rangle &= \frac{1}{4}\left(\langle M+N\rangle_n - \langle M-N\rangle_n\right) \\ &= \frac{1}{4}\sum_{j=0}^{n-1} \mathsf{E}\left\{\left(M_{j+1}+N_{j+1}\right)^2 - \left(M_j+N_j\right)^2\right. \\ &\qquad \left. - \left(M_{j+1}-N_{j+1}\right)^2 + \left(M_j-N_j\right)^2 \mid \mathcal{F}_j\right\} \\ &= \frac{1}{4}\left(\sum_{j=0}^{n-1} \mathsf{E}\left\{4M_{j+1}N_{j+1} - 4M_jN_j \mid \mathcal{F}_j\right\}\right) \\ &= \sum_{j=0}^{n-1} \mathsf{E}\left\{M_{j+1}N_{j+1} - M_jN_j \mid \mathcal{F}_j\right\} \\ &= \sum_{j=0}^{n-1} \mathsf{E}\left\{\left(M_j+\xi_{j+1}\right)\left(N_j+\eta_{j+1}\right) - M_jN_j \mid \mathcal{F}_j\right\} \\ &= \sum_{j=0}^{n-1} \mathsf{E}\left\{M_j\eta_{j+1} + \xi_{j+1}\eta_{j+1} + \xi_{j+1}N_j \mid \mathcal{F}_j\right\}\end{aligned}$$

$$= \sum_{j=0}^{n-1} \left(M_j \mathsf{E} \left\{ \eta_{j+1} \mid \mathcal{F}_j \right\} + N_j \mathsf{E} \left\{ \xi_{j+1} \mid \mathcal{F}_j \right\} + \mathsf{E} \left\{ \xi_{j+1} \eta_{j+1} \mid \mathcal{F}_j \right\} \right)$$

$$= \sum_{j=0}^{n-1} \mathsf{E} \left\{ \xi_{j+1} \eta_{j+1} \mid \mathcal{F}_j \right\}$$

7.2.4 Markov and stopping times

In this subsection we will consider the properties of the conditional mathematical expectation taken with respect to a sigma-algebra at a random time (defined by a random event depending on a prehistory of the process). We also will discuss the conditions when the preservation of the martingale property takes place under time-change at a random time.

7.2.4.1 Definition of a Markov and stopping times

Definition 7.8. *A random variable* $\tau = \tau(\omega)$ *defined on* $(\Omega, \mathcal{F}, \mathsf{P})$ *and taking values in the set* $\mathbb{N} := \{0, 1, \ldots\}$ *is called a* ***Markov time*** *with respect to a* σ*-algebra* $\mathcal{F}_n$ *(or, in other words, a random variable independent of the future) if for each* $n \in \mathbb{N}$

$$\boxed{\{\tau = n\} \in \mathcal{F}_n} \tag{7.36}$$

If, additionally,

$$\boxed{\mathsf{P}\{\omega : \tau(\omega) < \infty\} = 1} \tag{7.37}$$

then this Markov time is called a ***stopping time****.*

For any stochastic sequence $\{x_n, \mathcal{F}_n\}_{n \geq 0}$ and a Markov time τ (with respect to a σ-algebra $\mathcal{F}_n$) we can represent x_τ as

$$\boxed{x_\tau = \sum_{n=0}^{\infty} x_n \chi(\tau(\omega) = n)} \tag{7.38}$$

Proposition 7.1. *Obviously by (7.38)*

$$x_\tau = 0$$

on the set

$$\{\omega : \tau(\omega) = \infty\}$$

The following lemma clarifies the definition above and gives a criterion when a random time is, in fact, a Markov time.

Lemma 7.3. *A random time* $\tau = \tau(\omega)$ *is a Markov time if and only if one of the two properties holds: for all* $n \in \mathbb{N}$

$$\boxed{\begin{array}{c} \{\tau \leq n\} \in \mathcal{F}_n \\ \{\tau > n\} \in \mathcal{F}_n \end{array}} \tag{7.39}$$

Proof. It results directly from the relations

$$\{\tau \le n\} = \bigcup_{k=0}^{n} \{\tau \le k\}$$
$$\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}$$
$$\{\tau > n\} = \{\tau \le n\}^c$$

Example 7.9. *The typical stopping times are first entrance times, that is, the first time of a 'random walk' when the random trajectory reaches a certain region. Formally, this can be written as*

$$\boxed{\tau_B := \inf\{n \in \mathbb{N} : x_n \in B\}} \tag{7.40}$$

where $\{x_n, \mathcal{F}_n\}_{n\ge 0}$ is a stochastic sequence and $B \in \mathcal{B}(\mathbb{R})$.

The next definition will be central in this subsection.

Definition 7.9. *Every system $\{\mathcal{F}_n\}_{n\ge 0}$ and the corresponding Markov time τ generates together a collection of sets*

$$\boxed{\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}} \tag{7.41}$$

*which is called the **pre-τ-sigma-algebra**.*

Proposition 7.2. *$\mathcal{F}_\tau$ is a σ-algebra.*

Proof. To prove it we have to show that Ω complements countable unions of sets from $\mathcal{F}_\tau$ belonging to $\mathcal{F}_\tau$. Evidently $\Omega \in \mathcal{F}_\tau$ and $\mathcal{F}_\tau$ is closed under countable unions. Moreover, if $A \in \mathcal{F}_\tau$ then

$$\begin{aligned} A^c \cap \{\tau = n\} &= \{\Omega \setminus A\} \cap \{\tau = n\} \\ &= (\Omega \cap \{\tau = n\}) \setminus (A \cap \{\tau = n\}) \\ &= \{\tau = n\} \setminus (A \cap \{\tau = n\}) \in \mathcal{F}_\tau \end{aligned}$$

and therefore $A^c \in \mathcal{F}_\tau$. Hence, $\mathcal{F}_\tau$ is a σ-algebra. □

7.2.4.2 Stopping times properties

The next proposition gives some general facts on stopping times.

Proposition 7.3. *If τ is a Markov (stopping) time then the following facts are true:*

1. *every positive integer k is a Markov (stopping) time;*
2. *if $\tau = k$ then $\mathcal{F}_\tau = \mathcal{F}_k$;*
3. *$\mathcal{F}_\tau \subset \mathcal{F}_\infty$, $\tau \in \mathcal{F}_\tau$ and $\tau \in \mathcal{F}_\infty$;*
4. *$\{\tau = +\infty\} \in \mathcal{F}_\infty$*

Proof. Suppose $\tau = k$. We have

$$\{\tau = n\} = \Omega\chi\,(n = k) + \emptyset\chi\,(n \neq k)$$

that proves 1 since both Ω and $\emptyset$ belong to $\mathcal{F}_n$. For $A \in \mathcal{F}_\infty$ we have

$$A \cap \{\tau = n\} = A\chi\,(n = k) + \emptyset\chi\,(n \neq k)$$

that in view of (7.41) proves 2. As $\mathcal{F}_\tau$ are defined as sets of $\mathcal{F}_\infty$ it follows that $\mathcal{F}_\tau \subset \mathcal{F}_\infty$. For all integers m and n

$$\{\tau = m\} \cap \{\tau = n\} = \{\tau = n\}\,\chi\,(m = n) + \emptyset\chi\,(n \neq k) \in \mathcal{F}_n$$

that implies $\{\tau = m\} \in \mathcal{F}_\tau$. This exactly means that $\tau \in \mathcal{F}_\tau$. The fact $\tau \in \mathcal{F}_\infty$ results from the previous two. The fact 4 follows from the observation that

$$\{\tau = +\infty\} = \left(\bigcup_n \{\tau = n\}\right)^c \in \mathcal{F}_\infty$$

Proposition is proven. □

The following lemma (Gut, 2005) concerns the relation between Markov times and their convergent sequences.

Lemma 7.4. *If τ, τ_1 and τ_2 are Markov times then the following properties hold:*

1. $\tau_1 + \tau_2$, $\min\{\tau_1, \tau_2\}$ *and* $\max\{\tau_1, \tau_2\}$ *are Markov times;*
2. $\tau_M := \min\{\tau, M\}$ *is a Markov time (evidently, bounded);*
3. $\tau_1 - \tau_2$ *is not a Markov time;*
4. *if* $\{\tau_k\}_{k\geq 1}$ *are Markov (stopping) times then so are* $\sum_{k\geq 1} \tau_k$, $\min_k\{\tau_k\}$ *and* $\max_k\{\tau_k\}$*;*
5. *if* $\{\tau_k\}_{k\geq 1}$ *are Markov (stopping) times then* $\tau_k \uparrow \tau$ *and* $\tau_k \downarrow \tau$ *are stopping times too;*
6. *if* $A \in \mathcal{F}_{\tau_1}$ *then*

$$A \cap \{\tau_1 \leq \tau_2\} \in \mathcal{F}_{\tau_2}$$

7. *if* $\tau_1 \leq \tau_2$ *then*

$$\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$$

Proof.

1. For all $n \in \mathbb{N}$ in view of Lemma 7.3 we have

$$\{\tau_1 + \tau_2 = n\} = \bigcup_{k=0}^{n} (\{\tau_1 = k\} \cap \{\tau_2 = n - k\}) \in \mathcal{F}_n$$
$$\{\min\{\tau_1, \tau_2\} > n\} = \{\tau_1 > n\} \cap \{\tau_2 > n\} \in \mathcal{F}_n$$
$$\{\max\{\tau_1, \tau_2\} \leq n\} = \{\tau_1 \leq n\} \cap \{\tau_2 \leq n\} \in \mathcal{F}_n$$

   that proves the first item.
2. Since M is a Markov (evidently, stopping) time and $\tau_M \leq M$ it follows that τ_M is a stopping time too.

3. $\tau_1 - \tau_2$ is not a Markov time since it may be negative.
4. The result follows by deduction from 1.
5. It results from monotonicity since

$$\{\tau = n\} = \bigcap_k \{\tau_k = n\} \in \mathcal{F}_n$$
$$\{\tau = n\} = \bigcup_k \{\tau_k = n\} \in \mathcal{F}_n$$

respectively.
6. We have

$$A \cap \{\tau_1 \leq \tau_2\} \cap \{\tau_2 = n\} = (A \cap \{\tau_1 \leq n\}) \cap \{\tau_2 = n\} \in \mathcal{F}_n$$

that proves the item.
7. Since

$$A \cap \{\tau_2 = n\} = \underbrace{(A \cap \{\tau_1 \leq n\})}_{\in \mathcal{F}_{\tau_1}} \cap \{\tau_2 = n\} \in \mathcal{F}_{\tau_2}$$

we get the item. □

7.2.4.3 Theorems on martingale properties for stopping times

The next question seems to be natural: *if* $\{x_n, \mathcal{F}_n\}_{n \geq 0}$ *is a martingale and* $\tau_1 \overset{a.s.}{\leq} \tau_2$ *are both stopping times, is it true that* $\mathsf{E}\left\{x_{\tau_2} \mid \mathcal{F}_{\tau_1}\right\} \overset{a.s.}{=} x_{\tau_1}$*?* The theorem given below describes the typical situation when this is true.

Theorem 7.2. **(Doob, 1953)** *Let* $\{x_n, \mathcal{F}_n\}_{n \geq 0}$ *be a martingale (or submartingale) and*

$$\boxed{\tau_1 \overset{a.s.}{\leq} \tau_2} \tag{7.42}$$

are both stopping times for which

$$\boxed{\mathsf{E}\left\{\left|x_{\tau_i}\right|\right\} < \infty, \quad i = 1, 2} \tag{7.43}$$

and

$$\boxed{\liminf_{n \to \infty} \int\limits_{\{\tau_1 > n\}} |x_n| \, d\mathsf{P} = 0} \tag{7.44}$$

Then in the martingale case

$$\boxed{\mathsf{E}\left\{x_{\tau_2} \mid \mathcal{F}_{\tau_1}\right\} \overset{a.s.}{=} x_{\tau_1}} \tag{7.45}$$

and in the submartingale case

$$\boxed{\mathsf{E}\left\{x_{\tau_2} \mid \mathcal{F}_{\tau_1}\right\} \overset{a.s.}{\geq} x_{\tau_1}} \tag{7.46}$$

Proof. It is sufficient to show that for any $A \in \mathcal{F}_{\tau_1}$ with probability 1 we have

$$\int\limits_{A\cap\{\tau_1\leq\tau_2\}} x_{\tau_2} d\mathsf{P} \underset{(\geq)}{=} \int\limits_{A\cap\{\tau_1\leq\tau_2\}} x_{\tau_1} d\mathsf{P}$$

In turn, to demonstrate this it is sufficient to show that for any $n \geq 0$ and $B := A \cap \{\tau_1 = n\} \in \mathcal{F}_n$ with probability 1

$$\int\limits_{B\cap\{\tau_2\geq n\}} x_{\tau_2} d\mathsf{P} \underset{(\geq)}{=} \int\limits_{B\cap\{\tau_2\geq n\}} x_{\tau_1} d\mathsf{P} = \int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P} \tag{7.47}$$

In fact, we have

$$\begin{aligned}
\int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P} &= \int\limits_{B\cap\{\tau_2=n\}} x_n d\mathsf{P} + \int\limits_{B\cap\{\tau_2>n\}} x_n d\mathsf{P} \overset{a.s.}{=} \int\limits_{B\cap\{\tau_2=n\}} x_{\tau_2} d\mathsf{P} \\
&\quad + \int\limits_{B\cap\{\tau_2>n\}} E\left\{x_{n+1} \mid \mathcal{F}_n\right\} d\mathsf{P} \\
&\overset{a.s.}{=} \int\limits_{B\cap\{\tau_2=n\}} x_{\tau_2} d\mathsf{P} + \int\limits_{B\cap\{\tau_2>n+1\}} x_{n+1} d\mathsf{P} \\
&\underset{(\leq)}{=} \int\limits_{B\cap\{n\leq\tau_2\leq n+1\}} x_{\tau_2} d\mathsf{P} + \int\limits_{B\cap\{\tau_2\geq n+2\}} x_{n+2} d\mathsf{P} \\
&\underset{(\leq)}{=} \cdots \underset{(\leq)}{=} \int\limits_{B\cap\{n\leq\tau_2\leq m\}} x_{\tau_2} d\mathsf{P} + \int\limits_{B\cap\{\tau_2\geq m\}} x_m d\mathsf{P}
\end{aligned}$$

Hence,

$$\int\limits_{B\cap\{n\leq\tau_2\leq m\}} x_{\tau_2} d\mathsf{P} \underset{(\geq)}{=} \int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P} - \int\limits_{B\cap\{\tau_2\geq m\}} x_m d\mathsf{P}$$

and

$$\begin{aligned}
\int\limits_{B\cap\{n\leq\tau_2\}} x_{\tau_2} d\mathsf{P} &\underset{(\geq)}{=} \limsup_{m\to\infty} \left(\int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P} - \int\limits_{B\cap\{\tau_2\geq m\}} x_m d\mathsf{P} \right) \\
&= \int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P} - \liminf_{m\to\infty} \int\limits_{B\cap\{\tau_2\geq m\}} x_m d\mathsf{P} = \int\limits_{B\cap\{\tau_2\geq n\}} x_n d\mathsf{P}
\end{aligned}$$

in view of the assumption (7.44) and the presentation $x_m = 2x_m^+ - |x_m|$. So, (7.47) is established that proves (7.45) and (7.46). □

Corollary 7.1. *Under the assumptions of Theorem 7.2 the following relations hold:*

$$\boxed{\mathsf{E}\left\{x_{\tau_2}\right\} \underset{(\geq)}{=} \mathsf{E}\left\{x_{\tau_1}\right\}} \tag{7.48}$$

Proof. It results from (7.45) and (7.46) taking the mathematical expectation of both sides. □

The following properties can be easily checked.

Corollary 7.2.

1. *If the random variables $\{x_n\}_{n\geq 0}$ are uniformly integrable (see (6.32)), or, in particular, if*

$$\boxed{|x_n| \leq C < \infty \quad \text{for all} \quad n \geq 0} \tag{7.49}$$

then the conditions (7.43) and (7.44) are satisfied.

2. *If there exists a constant N such that*

$$\boxed{\mathsf{P}\{\tau_1 \leq N\} = \mathsf{P}\{\tau_2 \leq N\} = 1} \tag{7.50}$$

then the conditions (7.43) and (7.44) are satisfied too. Hence, if, in addition, $\mathsf{P}\{\tau_1 \leq \tau_2\} \leq 1$ and $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale (or submartingale) then

$$\boxed{\mathsf{E}\{x_N\} \underset{(\geq)}{=} \mathsf{E}\left\{x_{\tau_2}\right\} \underset{(\geq)}{=} \mathsf{E}\left\{x_{\tau_1}\right\} \underset{(\geq)}{=} \mathsf{E}\{x_0\}} \tag{7.51}$$

3. *If $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale (or submartingale) and τ is a stopping time then for*

$$\boxed{\tau \wedge n := \min\{\tau, n\}} \tag{7.52}$$

it follows that

$$\boxed{\mathsf{E}\{x_{\tau\wedge n}\} \underset{(\geq)}{=} \mathsf{E}\{x_n\}} \tag{7.53}$$

The following proposition is often used in various applications.

Theorem 7.3. **(Shiryayev, 1984)** *Let $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ be a martingale (or submartingale) and τ be a stopping time such that*

1.

$$\boxed{\mathsf{E}\{\tau\} < \infty} \tag{7.54}$$

2. *for some* $n \geq 0$ *and for some constant* C

$$\boxed{\mathsf{E}\{|x_{n+1} - x_n| \mid \mathcal{F}_n\} \leq C\left(\left\{\tau \overset{a.s.}{\geq} n\right\}\right)} \tag{7.55}$$

Then both

$$\boxed{\mathsf{E}\{|x_\tau|\} < \infty} \tag{7.56}$$

and

$$\boxed{\mathsf{E}\{x_\tau\} \underset{(\geq)}{=} \mathsf{E}\{x_0\}} \tag{7.57}$$

Proof. To state the result it is sufficient to verify that the hypotheses (7.43) and (7.44) of Theorem 7.2 are satisfied with $\tau_2 = \tau$. Define

$$y_0 := |x_0|, \quad y_k := |x_k - x_{k-1}| \text{ for } k \geq 1$$

Then

$$|x_\tau| = \left|\sum_{k=1}^{\tau}(x_k - x_{k-1}) + x_0\right| \leq \sum_{k=1}^{\tau}|x_k - x_{k-1}| + |x_0| = \sum_{k=0}^{\tau} y_k$$

and

$$\mathsf{E}\{|x_\tau|\} \leq \mathsf{E}\left\{\sum_{k=0}^{\tau} y_k\right\} = \int_{\Omega}\left(\sum_{k=0}^{\tau} y_k\right) d\mathsf{P} = \sum_{n=0}^{\infty}\int_{\{\tau=n\}}\left(\sum_{k=0}^{n} y_k\right) d\mathsf{P}$$
$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\int_{\{\tau=n\}} y_k d\mathsf{P} = \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\int_{\{\tau=n\}} y_k d\mathsf{P} = \sum_{k=0}^{\infty}\int_{\{\tau\geq k\}} y_k d\mathsf{P}$$

But

$$\{\tau \geq k\} = \Omega \setminus \{\tau < k\} \in \mathcal{F}_{k-1}$$

and therefore for $k \geq 1$

$$\int_{\{\tau\geq k\}} y_k d\mathsf{P} = \int_{\{\tau\geq k\}} \mathsf{E}\{y_k \mid \mathcal{F}_{k-1}\} d\mathsf{P} \leq C\mathsf{P}\{\tau \geq j\}$$

Hence,

$$\mathsf{E}\{|x_\tau|\} \leq \mathsf{E}\left\{\sum_{k=0}^{\tau} y_k\right\} \leq \mathsf{E}\{|x_0|\} + \mathsf{E}\left\{\sum_{k=1}^{\tau} y_k\right\}$$
$$= \mathsf{E}\{|x_0|\} + C\sum_{k=1}^{\infty}\mathsf{P}\{\tau \geq j\} = \mathsf{E}\{|x_0|\} + C\mathsf{E}\{\tau\} < \infty \tag{7.58}$$

Moreover, if $\tau \geq n$ then $\sum_{k=1}^{n} y_k \leq \sum_{k=1}^{\tau} y_k$ and therefore

$$\begin{aligned}
\int\limits_{\{\tau>n\}} |x_n|\, d\mathsf{P} &\leq \int\limits_{\{\tau>n\}} \left| \sum_{k=1}^{n} (x_k - x_{k-1}) + x_0 \right| d\mathsf{P} \\
&\leq \int\limits_{\{\tau>n\}} \left(\sum_{k=1}^{n} |x_k - x_{k-1}| + |x_0| \right) d\mathsf{P} \\
&= \int\limits_{\{\tau>n\}} \sum_{k=0}^{n} y_k d\mathsf{P} \leq \int\limits_{\{\tau>n\}} \sum_{k=0}^{\tau} y_k d\mathsf{P}
\end{aligned}$$

By (7.58) we obtain that $\mathsf{E}\left\{\sum_{k=0}^{\tau} y_k\right\} < \infty$ and whereas $n \to \infty$ it follows $\{\tau > n\} \downarrow \varnothing$ by the dominated convergence theorem yields

$$\liminf_{n\to\infty} \int\limits_{\{\tau>n\}} |x_n|\, d\mathsf{P} \leq \liminf_{n\to\infty} \int\limits_{\{\tau>n\}} \sum_{k=0}^{\tau} y_k d\mathsf{P} = 0$$

Hence both conditions (7.43) and (7.44) of Theorem 7.2 are fulfilled, which completes the proof. □

7.2.4.4 Stopped martingale

Theorem 7.4. *If $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale (or submartingale) and τ is stopping time then $\{x_{\tau\wedge n}, \mathcal{F}_n\}_{n\geq 0}$ is a martingale (or submartingale) too and called **the stopped martingale (submartingale)**.*

Proof. We have

$$\begin{aligned}
x_{\tau\wedge(n+1)} &= x_\tau \chi(\tau < n) + x_n \chi(\tau \geq n) \\
&= \sum_{k=1}^{n-1} x_k \chi(\tau = k) + x_n \chi(\tau \geq n) \in \mathcal{F}_n
\end{aligned}$$

since each term in the right-hand side of the last inequality is $\mathcal{F}_n$-measurable. Hence,

$$\begin{aligned}
\mathsf{E}\left\{x_{\tau\wedge(n+1)} \mid \mathcal{F}_n\right\} = &\sum_{k=1}^{n} \mathsf{E}\{x_k \chi(\tau = k) \mid \mathcal{F}_n\} \\
&+ \mathsf{E}\{x_{n+1}\chi(\tau \geq n+1) \mid \mathcal{F}_n\} \\
= &\sum_{k=1}^{n} x_k \chi(\tau = k) + \mathsf{E}\{x_{n+1}\chi(\tau > n) \mid \mathcal{F}_n\} \\
= &\sum_{k=1}^{n} x_k \chi(\tau = k) + \chi(\tau > n)\, \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\}
\end{aligned}$$

$$\overset{a.s.}{\underset{(\geq)}{=}} \sum_{k=1}^{n} x_k \chi(\tau = k) + \chi(\tau > n) x_n$$
$$= \sum_{k=1}^{n-1} x_k \chi(\tau = k) + \chi(\tau \geq n) x_n = x_{\tau \wedge n}$$

that completes the proof. □

Corollary 7.3. *In the case when $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a supermartingale and τ is stopping time it follows that $\{x_{\tau\wedge n}, \mathcal{F}_n\}_{n\geq 0}$ is a stopped supermartingale, that is,*

$$\boxed{\mathsf{E}\left\{x_{\tau\wedge(n+1)} \mid \mathcal{F}_n\right\} \overset{a.s.}{\leq} x_{\tau\wedge n}} \tag{7.59}$$

Proof. It is similar to that of Theorem 7.4. □

7.2.4.5 Wald's identities

Theorem 7.5. (Wald, 1947) *Let $\{\xi_n\}_{n\geq 0}$ be independent identically distributed random variables with $\mathsf{E}\{|\xi_n|\} < \infty$, $\mathsf{E}\{\xi_n\} = \mu$ and τ be a stopping time (with respect to $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \ldots, \xi_n)$) with $\mathsf{E}\{\tau\} < \infty$. Then*

$$\boxed{\mathsf{E}\{\xi_1 + \xi_2 + \cdots + \xi_\tau\} = \mu \mathsf{E}\{\tau\}} \tag{7.60}$$

If, additionally, $\mathsf{E}\left\{\xi_n^2\right\} = \sigma^2 < \infty$ then

$$\boxed{\mathsf{E}\left\{[(\xi_1 + \xi_2 + \cdots + \xi_\tau) - \mu \mathsf{E}\{\tau\}]^2\right\} = (\operatorname{var} \xi_1)\, \mathsf{E}\{\tau\}} \tag{7.61}$$

Proof. Evidently $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ with $x_n = (\xi_1 + \xi_2 + \cdots + \xi_\tau) - \mu \mathsf{E}\{\tau\}$ is a martingale such that

$$\mathsf{E}\{|x_{n+1} - x_n| \mid \mathcal{F}_n\} = \mathsf{E}\{|\xi_{n+1} - \mathsf{E}\{\xi_1\}| \mid \mathcal{F}_n\}$$
$$\overset{a.s.}{=} \mathsf{E}\{|\xi_{n+1} - \mathsf{E}\{\xi_1\}|\} \leq 2\mathsf{E}\{|\xi_1|\} < \infty$$

Then by Theorem 7.3 $\mathsf{E}\{x_\tau\} = \mathsf{E}\{x_0\} = 0$ proves (7.60). Similar considerations applied to the martingale $\{y_n, \mathcal{F}_n\}_{n\geq 0}$ with $y_n = x_n^2 - n \operatorname{var} \xi_1$ lead to (7.61). Theorem is proven. □

Corollary 7.4. *If τ is independent of $\{\xi_n\}_{n\geq 0}$ and $\operatorname{var} \tau < \infty$ then*

$$\boxed{\operatorname{var}(\xi_1 + \xi_2 + \cdots + \xi_\tau) = \sigma^2 \mathsf{E}\{\tau\} + \mu^2 \operatorname{var} \tau} \tag{7.62}$$

Proof. The result follows directly from the definition (3.21) of the variance and Theorem 7.5. □

Theorem 7.6. (Wald's Fundamental identity (Wald, 1947)) *Let $\{\xi_n\}_{n\geq 0}$ be independent identically distributed random variables, $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$ be a partial sum $(n \geq 1)$, and $\varphi(t) := \mathsf{E}\left\{e^{t\xi_n}\right\}$, $t \in \mathbb{R}$ and $\varphi(t_0)$ exist for some $t_0 \neq 0$ such that $\varphi(t_0) \geq 1$. If $\tau \geq 1$ is a stopping time (with respect to $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \ldots, \xi_n)$); then*

$$\boxed{\mathsf{E}\left\{\frac{e^{t\xi_n}}{[\varphi(t_0)]^{\tau}}\right\} = 1} \tag{7.63}$$

Proof. Define $y_n := e^{tS_n}[\varphi(t_0)]^{-n}$. It is easy to check that $\{y_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale with $\mathsf{E}\{y_n\} = 1$. Hence on the set $\{\tau \geq n\}$ we have

$$\begin{aligned}
\mathsf{E}\{|y_{n+1} - y_n| \mid \mathcal{F}_n\} &= y_n \mathsf{E}\left\{\left|\frac{e^{t\xi_{n+1}}}{\varphi(t_0)} - 1\right| \mid \mathcal{F}_n\right\} \\
&\overset{a.s}{=} y_n \mathsf{E}\left\{\left|\frac{e^{t\xi_{n+1}}}{\varphi(t_0)} - 1\right|\right\} \leq \text{Const} < \infty
\end{aligned}$$

So, Theorem 7.3 is applicable which leads to (7.63). Theorem is proven. □

Example 7.10. (Shiryayev, 1984) *Let $\{\xi_n\}_{n\geq 0}$ be independent Bernoulli random variables with*

$$\mathsf{P}\{\xi_n = 1\} = p \quad \textit{and} \quad \mathsf{P}\{\xi_n = -1\} = q$$

Let also $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$ and

$$\tau := \inf\{n \geq 1 : S_n = B \text{ or } S_n = A\}$$

where B and $(-A)$ are positive integers. Denote

$$\alpha = \mathsf{P}\{S_n = A\} \quad \textit{and} \quad \beta = \mathsf{P}\{S_n = B\}$$

so that

$$\alpha + \beta = 1 \tag{7.64}$$

It is not difficult to show that $\mathsf{P}\{\tau < \infty\} = 1$ and $\mathsf{E}\{\tau\} < \infty$. Consider two cases.

1. The case $p = q = 1/2$: from (7.61) it follows that

$$\mathsf{E}\{S_n\} = \alpha A + \beta B = \mu \mathsf{E}\{\tau\} = 0 \text{ since } \mu = 0$$

that together with (7.64) implies

$$\alpha = \frac{B}{B + |A|}, \quad \beta = \frac{|A|}{B + |A|}$$

2. *The case* $p \neq q$ $(p > 0)$*: notice that* $\left\{(q/p)^{S_n}, \mathcal{F}_n\right\}_{n\geq 0}$ *with* $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \ldots, \xi_n)$ *is a martingale. Indeed,*

$$\begin{aligned} \mathsf{E}\left\{(q/p)^{S_{n+1}} \mid \mathcal{F}_n\right\} &= \mathsf{E}\left\{(q/p)^{S_n} (q/p)^{\xi_{n+1}} \mid \mathcal{F}_n\right\} \\ &\overset{a.s.}{=} (q/p)^{S_n} \mathsf{E}\left\{(q/p)^{\xi_{n+1}} \mid \mathcal{F}_n\right\} \\ &= (q/p)^{S_n}\left[(q/p)\,p + (q/p)^{-1} q\right] = (q/p)^{S_n} \end{aligned}$$

So,

$$\mathsf{E}\left\{(q/p)^{S_\tau}\right\} = \mathsf{E}\left\{(q/p)^{\xi_1}\right\} = \left[(q/p)\,p + (q/p)^{-1} q\right] = 1$$

and therefore

$$\alpha\,(q/p)^A + \beta\,(q/p)^B = 1$$

that together with (7.64) yields

$$\begin{aligned} \alpha &= \left[(q/p)^B - 1\right]\left[(q/p)^B - (q/p)^{|A|}\right]^{-1} \\ \beta &= \left[1 - (q/p)^{|A|}\right]\left[(q/p)^B - (q/p)^{|A|}\right]^{-1} \end{aligned} \tag{7.65}$$

Finally, by (7.61) one finds

$$\mathsf{E}\{S_n\} = (p-q)\,\mathsf{E}\{\tau\} = 0$$

that leads to the expression

$$\mathsf{E}\{\tau\} = \frac{\mathsf{E}\{S_n\}}{p-q} = \frac{\alpha A + \beta B}{p-q}$$

where α *and* β *are given by (7.65).*

7.3 Main martingale inequalities

Let $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ be a stochastic sequence. Define

$$\boxed{x_n^{\max} := \max_{j=1,\ldots,n} x_j, \quad x_n^* := \max_{j=1,\ldots,n} \left|x_j\right|} \tag{7.66}$$

and

$$\boxed{\|x_n\|_{L_p} := \left(\mathsf{E}\left\{|x_n|^p\right\}\right)^{1/p}} \tag{7.67}$$

Notice that, evidently,

$$\boxed{\|x_n\|_{L_p} \leq \left\|x_n^*\right\|_{L_p}} \tag{7.68}$$

7.3.1 Doob's inequality of the Kolmogorov type

Here we present the inequalities of the Kolmogorov type but derived for martingales and submartingales.

Theorem 7.7. (Doob, 1953)

1. Suppose $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a submartingale. Then for any $\varepsilon > 0$ and all integers $n \geq 0$

$$\boxed{\varepsilon \mathsf{P}\left\{x_n^{\max} > \varepsilon\right\} \leq \int\limits_{\{\omega: x_n^{\max} > \varepsilon\}} x_n d\mathsf{P} \leq \mathsf{E}\left\{x_n^+\right\} \leq \mathsf{E}\{|x_n|\}} \tag{7.69}$$

2. If $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale then for any $\varepsilon > 0$ and all integers $n \geq 0$

$$\boxed{\varepsilon \mathsf{P}\left\{x_n^* > \varepsilon\right\} \leq \int\limits_{\{\omega:\, x_n^{\max} > \varepsilon\}} |x_n|\, d\mathsf{P} \leq \mathsf{E}\{|x_n|\}} \tag{7.70}$$

Proof. Define

$$\tau := \min_k \{k : x_k > \varepsilon\}$$

and

$$\Lambda_n := \left\{\omega : x_n^{\max} > \varepsilon\right\} = \{\omega : x_{\tau\wedge n} > \varepsilon\}$$

By Theorem 7.4 the submartingale property is preserved for $\{x_{\tau\wedge n}, \mathcal{F}_n\}_{n\geq 0}$ and since $\Lambda_n \in \mathcal{F}_n$ we have

$$\begin{aligned}
\varepsilon \mathsf{P}\left\{x_n^{\max} > \varepsilon\right\} &\leq \int\limits_{\{\omega: x_n^{\max} > \varepsilon\}} x_{\tau\wedge n} d\mathsf{P} \leq \int\limits_{\{\omega: x_n^{\max} > \varepsilon\}} x_n d\mathsf{P} \\
&= \int\limits_{\Omega} x_n \chi\left\{\omega : x_n^{\max} > \varepsilon\right\} d\mathsf{P} \leq \int\limits_{\Omega} x_n \chi\left\{\omega : x_n^{\max} \geq 0\right\} d\mathsf{P} \\
&= \int\limits_{\Omega} \max\{x_n; 0\} = \mathsf{E}\left\{x_n^+\right\} \leq \mathsf{E}\{|x_n|\}
\end{aligned}$$

that proves (7.69). The inequality (7.70) follows from (7.69) since

(a) $\{|x_n|, \mathcal{F}_n\}_{n\geq 0}$ is a submartingale because of the relation

$$\mathsf{E}\{|x_{n+1}| \mid \mathcal{F}_n\} \geq \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{=} x_n$$

which implies

$$\mathsf{E}\{|x_{n+1}| \mid \mathcal{F}_n\} = |\mathsf{E}\{|x_{n+1}| \mid \mathcal{F}_n\}| \geq |x_n|$$

(b) $\left\{x_n^*, \mathcal{F}_n\right\}_{n\geq 0}$ is a submartingale because of the relation

$$x_{n+1}^* \geq x_n^*$$

Theorem is proven. □

Corollary 7.5. *Let* $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ *be a* p*-integrable martingale* $(p \geq 1)$. *Then* $\{|x_n|^p, \mathcal{F}_n\}_{n\geq 0}$ *is a submartingale and*

$$\boxed{\mathsf{P}\left\{x_n^* > \varepsilon\right\} \leq \varepsilon^{-p}\mathsf{E}\left\{|x_n|^p\right\}} \tag{7.71}$$

Proof. Since a conditional mathematical expectation is a Lebesgue integral (with respect to a conditional measure) the Jensen inequality (4.19) is also valid. Therefore, in view of the martingale property ($\{|x_n|, \mathcal{F}_n\}_{n\geq 0}$ is a submartingale), it follows that

$$\mathsf{E}\left\{|x_{n+1}|^p \mid \mathcal{F}_n\right\} \geq (\mathsf{E}\left\{|x_{n+1}| \mid \mathcal{F}_n\right\})^p \overset{a.s.}{\geq} |x_n|^p$$

So, $\left\{|x_n|^p, \mathcal{F}_n\right\}_{n\geq 0}$ is a submartingale too. Hence, applying directly (7.70) we get

$$\mathsf{P}\left\{x_n^* > \varepsilon\right\} = \mathsf{P}\left\{\max_{j=1,\ldots,n} \left|x_j\right|^p > \varepsilon^p\right\} \leq \varepsilon^{-p}\mathsf{E}\left\{|x_n|^p\right\}$$

that proves (7.71). □

Corollary 7.6. *If in the previous corollary* $x_n := S_n = \sum_{k=1}^{n} \xi_k$ *where* $\{\xi_n\}_{n\geq 0}$ *is a sequence of independent random variables with* $\mathsf{E}\{\xi_k\} = 0$ *and* $\operatorname{var}\xi_k = \mathsf{E}\left\{\xi_k^2\right\} < \infty$, *then* $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ *is a martingale and the inequality (7.71) becomes*

$$\mathsf{P}\left\{\max_{k=\overline{1,n}} |S_k| > \varepsilon\right\} \leq \varepsilon^{-2}\mathsf{E}\left\{S_n^2\right\} = \varepsilon^{-2}\sum_{k=1}^{n} \operatorname{var}\xi_k$$

that coincides with the ***Kolmogorov inequality*** *(4.38).*

7.3.2 Doob's moment inequalities

The inequalities below relate moments of maxima to moments of the last element in a finite sequence. But first, let us prove the following auxiliary result.

Lemma 7.5. *Let* x *and* y *be non-negative random variables such that*

$$\boxed{\mathsf{P}\{y > v\} \leq v^{-1} \int_{\{\omega: y\geq v\}} x\, d\mathsf{P}, \quad v^p\mathsf{P}\{y > v\} \underset{v\to\infty}{\to} 0 \quad \textit{for} \quad p \geq 1} \tag{7.72}$$

then

$$\boxed{\mathsf{E}\{y^p\} \le \begin{cases} \left(\dfrac{p}{p-1}\right)^p \mathsf{E}\{x^p\} & \text{if } p > 1 \\ \dfrac{e}{e-1}\left[1 + \mathsf{E}\left\{x \log^+(x)\right\}\right] & \text{if } p = 1 \end{cases}} \tag{7.73}$$

Proof. Let q satisfy $p^{-1} + q^{-1} = 1$.

(a) Consider the case $p > 1$. We have

$$\begin{aligned} \mathsf{E}\left\{y^p\right\} &= \int_{v=0}^{\infty} v^p d\mathsf{P}\{y \le v\} = \int_{v=0}^{\infty} v^p d\left(1 - \mathsf{P}\{y > v\}\right) \\ &= -\int_{v=0}^{\infty} v^p d\mathsf{P}\{y > v\} = -v^p \mathsf{P}\{y > v\} \big|_{v=0}^{v=\infty} + \int_{v=0}^{\infty} \mathsf{P}\{y > v\} dv^p \\ &= \int_{v=0}^{\infty} \mathsf{P}\{y > v\} dv^p = p \int_{v=0}^{\infty} v^{p-1} \mathsf{P}\{y > v\} dv \\ &\le p \int_{v=0}^{\infty} v^{p-2} \left(\int_{\{\omega: y \ge v\}} x d\mathsf{P} \right) dv = p \int_{\Omega} x \left(\int_{v=0}^{y} v^{p-2} dv \right) d\mathsf{P} \\ &= \frac{p}{p-1} \mathsf{E}\left\{x y^{p-1}\right\} \end{aligned}$$

Applying then the Hölder inequality (13.73) we get

$$\begin{aligned} \mathsf{E}\left\{y^p\right\} &\le \frac{p}{p-1} \mathsf{E}\left\{x y^{p-1}\right\} \\ &\le \frac{p}{p-1} \|x\|_{L_p} \left\| y^{p-1} \right\|_{L_q} = \frac{p}{p-1} \|x\|_{L_p} \|y\|_{L_p}^{p-1} \end{aligned}$$

(b) If $\|y\|_{L_p} = 0$, obviously, (7.73) holds. If not, the division of the last inequality by $\|y\|_{L_p}^{p-1}$ completes the proof for the case $p > 1$.

(c) Let now $p = 1$. In this case we have

$$\begin{aligned} \mathsf{E}\{y\} &= \int_{v=0}^{\infty} v d\mathsf{P}\{y \le v\} = \int_{v=0}^{\infty} \mathsf{P}\{y > v\} dv \\ &= \int_{v=0}^{1} \mathsf{P}\{y > v\} dv + \int_{v=1}^{\infty} \mathsf{P}\{y > v\} dv \le 1 + \int_{v=1}^{\infty} \mathsf{P}\{y > v\} dv \end{aligned}$$

$$\le 1 + \int_{v=1}^{\infty} \left(v^{-1} \int_{\{\omega: y \ge v\}} x d\mathsf{P} \right) dv = 1 + \int_{\Omega} x \left(\int_{v=1}^{y} v^{-1} dv \right) d\mathsf{P}$$
$$= 1 + \mathsf{E}\left\{ x \log^{+} y \right\}$$

Using the inequality[3]

$$a \log^{+} b \le a \log^{+} a + b/e \tag{7.74}$$

valid for any $a, b > 0$ we obtain (taking $a = x$ and $b = y$)

$$\mathsf{E}\{y\} \le 1 + \mathsf{E}\left\{ x \log^{+} y \right\} \le 1 + \mathsf{E}\left\{ x \log^{+} x \right\} + \mathsf{E}\{y\}/e$$

which after simple rearrangements leads to (7.73). Lemma is proven. □

Now we are ready to formulate the following result.

Theorem 7.8. (Doob's moment maximal inequality)

1. If $\{x_n, \mathcal{F}_n\}_{n \ge 0}$ is a non-negative submartingale then

$$\boxed{\mathsf{E}\left\{ \left(x_n^{\max} \right)^p \right\} \le \begin{cases} \left(\dfrac{p}{p-1} \right)^p \mathsf{E}\left\{ (x_n)^p \right\} & \text{if } p > 1 \\ \dfrac{e}{e-1} \left[1 + \mathsf{E}\left\{ x_n \log^{+}(x_n) \right\} \right] & \text{if } p = 1 \end{cases}} \tag{7.75}$$

2. If $\{x_n, \mathcal{F}_n\}_{n \ge 0}$ is a martingale then

$$\boxed{\mathsf{E}\left\{ \left(x_n^{*} \right)^p \right\} \le \begin{cases} \left(\dfrac{p}{p-1} \right)^p \mathsf{E}\left\{ |x_n|^p \right\} & \text{if } p > 1 \\ \dfrac{e}{e-1} \left[1 + \mathsf{E}\left\{ |x_n| \log^{+}(|x_n|) \right\} \right] & \text{if } p = 1 \end{cases}} \tag{7.76}$$

Proof. Letting $y := x_n^{\max}$ and $x := x_n$ in Lemma 7.5 and using Theorem 7.7 we immediately get (7.75). The inequality (7.76) follows directly from (7.75) if we take into account that $\{|x_n|, \mathcal{F}_n\}_{n \ge 0}$ is a submartingale. □

---

[3] If $a > 0, b \le 1$ the left-hand side of (7.74) is equal to 0, and, hence, there is nothing to prove. For $a > b$ the inequality is also trivial. So, consider the case $1 < a \le b$. Then

$$a \log^{+} b = a \log b \le a \log a + a \log(b/a) = a \log^{+} a + b \frac{\log^{+}(b/a)}{(b/a)} \le a \log^{+} a + b/e$$

since $\dfrac{\log c}{c} \le e^{-1}$ for $c \ge 1$.

7.4 Convergence

One of the most important theorems is the convergence theorem which was formulated and proven by Doob (1953) using the so-called *upcrossing and downcrossing lemma* given below.

7.4.1 'Up-down' crossing lemma

Lemma 7.6. (on 'up-down' crossing (Doob, 1953)) *For any nonnegative real variables a and b ($a < b$) define the random sequence $\{\beta_n\}_{n\geq 0}$ where β_n represents **the number of times** where the process $\{\xi_n\}$ drops below (downcrossing) a or rises above (upcrossing) b during the time n. Then*

$$\boxed{\mathsf{E}\{\beta_n(a,b)\} \leq (b-a)^{-1}\,\mathsf{E}\left\{[a-\xi_n]^+\right\}} \tag{7.77}$$

Proof. Let the random sequence of times $\{\tau_n\}_{n\geq 0}$ correspond to the 'first times' when $\{\xi_n\}$ leaves the interval $[a,b]$, namely (for $k = 1, 2, \ldots$)

$$\tau_1 := \min\left\{n \mid \xi_n < a, \xi_t \geq a \ \forall t = \overline{1, n-1}\right\}$$
$$\tau_{2k} := \min\left\{n \mid \tau_{2k-1}, \xi_n > b, \xi_t \leq b \ \forall t = \overline{\tau_{2k-1}, n-1}\right\}$$
$$\tau_{2k+1} := \min\left\{n \mid \tau_{2k}, \xi_n < a, \xi_n \geq a \ \forall t = \overline{\tau_{2k}, n-1}\right\}$$

Define also the characteristic function

$$\chi_n := \begin{cases} 1, & \tau_{2k-1} < n \leq \tau_{2k} \\ 0, & \tau_{2k} < n \leq \tau_{2k+1} \end{cases}$$

of the event that the random process $\{\beta_n\}_{n\geq 0}$ is inside of the $[a,b]$-interval. Then we have

$$\sum_{t=1}^{n-1} \chi_t\,(\xi_{t+1} - \xi_t) \geq \begin{cases} (b-a)\beta_n(a,b), & \tau_{2k} < n \leq \tau_{2k+1} \\ (b-a)\beta_n(a,b) + \xi_n - \xi_{\tau_{2k-1}}, & \tau_{2k-1} < n \leq \tau_{2k} \end{cases}$$

Notice that $\xi_{2k-1} < a$. So, one has

$$(b-a)\beta_n(a,b) \leq \sum_{t=1}^{n} \chi_t\,(\xi_{t+1} - \xi_t) + \max\{0, a - \xi_n\}$$

Taking into account that the random variable χ_t is $\mathcal{F}_t$-measurable, we derive:

$$\begin{aligned} \mathsf{E}\{\chi_t\,(\xi_{t+1} - \xi_t)\} &= \mathsf{E}\{\mathsf{E}\{\chi_t\,(\xi_{t+1} - \xi_t) \mid \mathcal{F}_t\}\} \\ &= \mathsf{E}\{\chi_t E\,\{(\xi_{t+1} - \xi_t) \mid \mathcal{F}_t\}\} \leq 0 \end{aligned}$$

and, therefore

$$(b-a)\,\mathsf{E}\{\beta_n(a,b)\} \leq \mathsf{E}\left\{[a-\xi_n]^+\right\}$$

that proves (7.77). □

7.4.2 Doob's theorem on sub (super) martingale almost sure convergence

Now due to Doob (1953) the main result on sub (super) martingales convergence can be formulated.

Theorem 7.9. (on sub (super) martingale convergence) *Let* $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ *be a nonnegative sub (super) martingale such that*

$$\boxed{\sup_{n\geq 0} \mathsf{E}\{x_n\} < \infty} \tag{7.78}$$

Then the random sequence $\{x_n\}_{n\geq 0}$ *converges with probability one to a nonnegative integrable random variable, that is, there exists a nonnegative integrable random* x *(defined on the same probability space satisfying* $\mathsf{E}\{x\} < \infty$*) such that*

$$\boxed{x_n \underset{n\to\infty}{\overset{a.s.}{\to}} x} \tag{7.79}$$

Proof. This theorem may be proved by contradiction. Indeed, assume that the limit x does not exist, i.e.,

$$\mathsf{P}\left\{\liminf_{n\to\infty} x_n < \limsup_{n\to\infty} x_n\right\} > 0$$

Hence, there exist numbers a and b such that

$$\mathsf{P}\left\{\liminf_{n\to\infty} \xi_n \leq a < b < \limsup_{n\to\infty} \xi_n\right\} > 0$$

and, as a result, $\mathsf{P}\left\{\lim_{n\to\infty} \beta_n(a,b) = \infty\right\} > 0$. But by the assumption

$$\begin{aligned}\sup_n \mathsf{E}\left\{[a-\xi_n]^+\right\} &= \sup_n \mathsf{E}\{\max\{0, a-\xi_n\}\} \\ &\leq \sup_n \mathsf{E}\{|a-\xi_n|\} \leq \sup_n \mathsf{E}\{|a|+|\xi_n|\} = |a| + \sup_n \mathsf{E}\{|\xi_n|\} < \infty\end{aligned}$$

Finally, in view of Fatou's Lemma 6.4 we obtain the contradiction:

$$\begin{aligned}\infty &> (b-a)^{-1}\sup_n \mathsf{E}\left\{[a-\xi_n]^+\right\} \\ &\geq \liminf_{n\to\infty} \mathsf{E}\{\beta_n(a,b)\} \geq \mathsf{E}\left\{\liminf_{n\to\infty} \beta_n(a,b)\right\} = \infty\end{aligned}$$

Theorem is proven. □

7.4.3 Martingale decomposition and almost sure convergence

Lemma 7.7. (The martingale decomposition) *For any martingale* $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ *such that*

$$\boxed{\begin{aligned}&\sup_n \mathsf{E}\left\{x_n^+\right\} < \infty, \quad x_n^+ := \max\{0; x_n\} \\ &\sup_n \mathsf{E}\left\{x_n^-\right\} < \infty, \quad x_n^- := -\min\{0; x_n\}\end{aligned}} \tag{7.80}$$

there exists two nonnegative submartingales $\left\{M_n^{(i)}, \mathcal{F}_n\right\}_{n\geq 0}$ $(i = 1, 2)$ *such that*

$$\boxed{x_n = M_n^{(1)} - M_n^{(2)}} \tag{7.81}$$

The decomposition (7.81) is not unique.

Proof. This decomposition immediately follows from the presentation $x_n = x_n^+ - x_n^-$ (2.12) since

$$\mathsf{E}\left\{x_n^+ \mid \mathcal{F}_{n-1}\right\} \geq \max\{0; \mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\}\} \overset{a.s.}{=} \max\{0; x_{n-1}\} = x_{n-1}^+$$

and

$$\mathsf{E}\left\{x_n^- \mid \mathcal{F}_{n-1}\right\} = \mathsf{E}\{-\min\{x_n; 0\} \mid \mathcal{F}_{n-1}\}$$
$$\geq -\min\{\mathsf{E}\{\{x_n\} \mid \mathcal{F}_{n-1}\}; 0\} \overset{a.s.}{\geq} -\min\{x_n; 0\} = x_{n-1}^-$$

Non-uniqueness results from the presentation

$$x_n = M_n^{(1)} + a - \left(M_n^{(2)} + a\right)$$

valid for any $a \geq 0$. Lemma is proven. □

Theorem 7.10. (on the martingale convergence) *Let* $\{x_n, \mathcal{F}_n\}_{n\geq 0}$ *be a martingale such that*

$$\boxed{\sup_{n\geq 0} \mathsf{E}\{|x_n|\} < \infty} \tag{7.82}$$

Then the random sequence $\{x_n\}_{n\geq 0}$ *converges with probability one to an integrable random variable, that is, there exists an integrable random x (defined on the same probability space satisfying* $\mathsf{E}\{|x|\} < \infty$*) such that*

$$\boxed{x_n \underset{n\to\infty}{\overset{a.s.}{\to}} x} \tag{7.83}$$

Proof. It follows from (7.81) if we take into account that both nonnegative supermartingales $\left\{M_n^{(i)}, \mathcal{F}_n\right\}_{n\geq 0}$ $(i = 1, 2)$ satisfy (7.78) since

$$|x_n| = x_n^+ + x_n^-$$

and hence,

$$|x_n| \geq x_n^+, \quad |x_n| \geq x_n^-$$

so that, by the assumption (7.82),

$$\sup_{n\geq 0} \mathsf{E}\left\{x_n^+\right\} \leq \sup_{n\geq 0} \mathsf{E}\{|x_n|\} < \infty$$
$$\sup_{n\geq 0} \mathsf{E}\left\{x_n^-\right\} \leq \sup_{n\geq 0} \mathsf{E}\{|x_n|\} < \infty$$

Therefore, by Doob's Theorem 7.9

$$M_n^{(i)} \underset{n\to\infty}{\overset{a.s.}{\to}} M^{*(i)} \ (i=1,2)$$

where both random variables $M^{*(i)}$ $(i=1,2)$ are integrable, and hence,

$$x_n \underset{n\to\infty}{\overset{a.s.}{\to}} x = M^{*(1)} - M^{*(2)}$$

theorem is proven. □

7.4.4 Robbins–Siegmund theorem and its corollaries

Based on Doob's theorem on the supermartingale convergence, the following keystone convergence theorem for nonnegative '*almost supermartingales*', having many applications in stochastic processes theory, can be proven.

7.4.4.1 Robbins–Siegmund theorem and its generalization

Theorem 7.11. (Siegmund and Robbins, 1971) *Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a sequence of σ-algebras and x_n, α_n, β_n and ξ_n be $\mathcal{F}_n$-measurable nonnegative random variables such that for all $n = 1, 2, \ldots$ there exists $\mathsf{E}\{x_{n+1}/\mathcal{F}_n\}$ and the following inequality verified:*

$$\boxed{\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \leq x_n(1+\alpha_n) + \beta_n - \xi_n} \tag{7.84}$$

with probability one. Then, for all $\omega \in \Omega_0$ where

$$\boxed{\Omega_0 := \left\{\omega \in \Omega \,\middle|\, \sum_{n=1}^{\infty} \alpha_n < \infty, \ \sum_{n=1}^{\infty} \beta_n < \infty\right\}} \tag{7.85}$$

the limit

$$\boxed{\lim_{n\to\infty} x_n = x^*(\omega)} \tag{7.86}$$

exists, and the sum

$$\boxed{\sum_{n=1}^{\infty} \xi_n < \infty} \tag{7.87}$$

converges.

Proof.

(a) Consider the following sequences:

$$\widetilde{x}_n := x_n \prod_{t=1}^{n-1} (1+\alpha_t)^{-1}, \quad \widetilde{\beta}_n := \beta_n \prod_{t=1}^{n-1} (1+\alpha_t)^{-1} \tag{7.88}$$

and

$$\widetilde{\xi}_n := \xi_n \prod_{t=1}^{n-1} (1+\alpha_t)^{-1}$$

Then, in view of (7.84) with probability one it follows that

$$\mathsf{E}\{\widetilde{x}_{n+1} \mid \mathcal{F}_n\} = \mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \prod_{t=1}^{n-1} (1+\alpha_t)^{-1} \leq \widetilde{x}_n + \widetilde{\beta}_n - \widetilde{\xi}_n \tag{7.89}$$

Let us introduce also the following random variables:

$$u_n := \widetilde{x}_n - \sum_{t=1}^{n-1} \left(\widetilde{\beta}_t - \widetilde{\xi}_t\right)$$

which by (15.23), obviously, satisfy

$$\mathsf{E}\{u_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\leq} u_n$$

and the random time τ, defined by

$$\tau = \tau(a) := \begin{cases} \inf\left\{ n \,\middle|\, \sum_{t=1}^{n-1} \widetilde{\beta}_t \geq a \right\}, & a = \text{const} > 0 \;\text{ if } \sum_{t=1}^{\infty} \widetilde{\beta}_t \geq a \\ +\infty & \text{if } \sum_{t=1}^{\infty} \widetilde{\beta}_t < a \end{cases}$$

Note that for any $a > 0$ the random variable $\tau(a)$ is a Markov time with respect to σ-algebras flow $\{\mathcal{F}_n\}_{n\geq 0}$ since by the conditions of this theorem, the set

$$(\omega : \tau(a) = n) = \left(\sum_{t=1}^{n-1} \widetilde{\beta}_t < a\right) \bigcap \left(\sum_{t=1}^{n} \widetilde{\beta}_t \geq a\right) \in \mathcal{F}_n$$

The stopping process

$$u_{\tau\wedge n} := u_\tau \chi(\tau < n) + u_n \chi(\tau \geq n)$$

$$\tau \wedge n := \tau\chi(\tau < n) + u\chi(\tau \geq n)$$

obviously has a lower bound:

$$u_{\tau\wedge n} \geq \sum_{t=1}^{(\tau\wedge n)-1} \widetilde{\beta}_t \geq -a$$

Since for each $n = 1, 2, \ldots$ the random variables

$$u_\tau \chi(\tau < n+1) = \sum_{i=1}^{n} u_i \, \chi(\tau = i)$$

and the random variables

$$\chi(\tau \ge n+1) = 1 - \chi(\tau \le n)$$

are $\mathcal{F}_n$-measurable, then (with probability 1) it follows

$$\begin{aligned} \mathsf{E}\left\{u_{\tau\wedge(n+1)} \mid \mathcal{F}_n\right\} &= u_\tau \chi(\tau \le n) + \chi(\tau < n) \mathsf{E}\{u_{n+1} \mid \mathcal{F}_n\} \\ &\overset{a.s.}{\le} u_\tau \chi(\tau \le n) + u_n \chi(\tau > n) = u_{\tau\wedge n} \end{aligned}$$

This implies that $\{u_{\tau\wedge n}, \mathcal{F}_n\}_{n\ge 0}$ is a supermartingale having a lower bound. Hence, by Doob's Theorem 7.9 it converges with probability one, that is,

$$\lim_{n\to\infty} u_{\tau\wedge n} \overset{a.s.}{=} u^*(\omega)$$

(b) Extract the desirable result from this fact. Let us introduce the monotone (on a) family of sets

$$\Omega_a = \Omega_0 \cap \{\omega | \tau(a) = \infty\}$$

Obviously,

$$\Omega_a \uparrow \Omega_0 \quad \text{as} \quad a \uparrow \infty$$

Thus, since $\beta_n \ge 0$ and for any $a > 0$ and for almost all $\omega \in \Omega_a$ the equality

$$u_{\tau\wedge n} = u_{\infty\wedge n} = u_n$$

remains true, we obtain that the limit exists, that is,

$$\lim_{n\to\infty} u_n = u^*(\omega)$$

almost sure over set Ω_0. Note that both left- and right-hand sides of the last identity do not depend on parameter a. Therefore, almost sure over set Ω_0

$$\widetilde{x}_n + \sum_{t=1}^{n-1} \widetilde{\xi}_t \underset{n\to\infty}{\to} u^*(\omega) + \sum_{t=1}^{\infty} \widetilde{\beta}_t$$

Since $S_n = \sum_{t=1}^{n} \widetilde{\xi}_t$ represents a monotone, bounded (for almost all $\omega \in \Omega_0$) from the above sequence, we may conclude that it has a finite limit $S_\infty(\omega)$ over set Ω_0, and consequently, variable $\widetilde{x}_n$ has also finite (for almost all $\omega \in \Omega_0$) limit $\widetilde{x}_\infty(\omega)$. Therefore, by definition (7.88),

$$x_n \underset{n\to\infty}{\to} \widetilde{x}_\infty(\omega) \prod_{t=1}^{\infty} (1 + \alpha_t) = x^*(\omega)$$

and, therefore,

$$\sum_{n=1}^{\infty} \xi_n = \sum_{n=1}^{\infty} \widetilde{\xi}_n \prod_{m=1}^{t} (1+\alpha_m) \le S_\infty \prod_{m=1}^{\infty} (1+\alpha_m) < \infty$$

for almost all $\omega \in \Omega_0$ since, by the relation $1 + z \le e^z$,

$$\prod_{m=1}^{\infty} (1+\alpha_m) < \infty \quad \text{if} \quad \sum_{m=1}^{\infty} \alpha_m < \infty$$

Theorem is proven. □

The next results represents a generalized version of Theorem 7.11 for 'almost super-martingales' involving root-terms.

Theorem 7.12. **(Devyaterikov and Poznyak, 1984)** *Let $\{\mathcal{F}_n\}_{n\ge 0}$ be a sequence of σ-algebras and x_n, α_n, β_n, γ_n and ξ_n be $\mathcal{F}_n$-measurable nonnegative random variables such that for all $n = 1, 2, \ldots$ there exists $\mathsf{E}\{x_{n+1}/\mathcal{F}_n\}$ and the following inequality verified:*

$$\boxed{\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \le x_n(1+\alpha_n) + \beta_n + \gamma_n x_n^r - \xi_n, \quad r \in (0,1)} \tag{7.90}$$

with probability one. Then, for all $\omega \in \Omega_0$ where

$$\boxed{\Omega_0 := \left\{ \omega \in \Omega \,\middle|\, \sum_{n=1}^{\infty} \alpha_n < \infty,\ \sum_{n=1}^{\infty} \beta_n < \infty, \sum_{n=1}^{\infty} \gamma_n \left(\frac{\gamma_n}{\alpha_n}\right)^{r/(1-r)} < \infty \right\}} \tag{7.91}$$

the limit

$$\boxed{\lim_{n\to\infty} x_n = x^*(\omega)}$$

exists, and the sum

$$\boxed{\sum_{n=1}^{\infty} \xi_n < \infty}$$

converges.

Proof. Using the inequality (see Lemma 16.16, formula 16.247 in Poznyak (2008))

$$x^r \le (1-r)\,x_0^r + \frac{r}{x_0^{1-r}} x$$

valid for any $x, x_0 > 0$ letting

$$x := x_n, \quad x_0 := \left(r \frac{\gamma_n}{\alpha_n} \right)^{1/(1-r)}$$

we get

$$
\begin{aligned}
\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} &\le x_n(1+\alpha_n) + \beta_n \\
&\quad + \gamma_n \left[(1-r)\left(r\frac{\gamma_n}{\alpha_n}\right)^{r/(1-r)} + \frac{r}{\left(r\frac{\gamma_n}{\alpha_n}\right)} x_n \right] - \xi_n \\
&= x_n(1+2\alpha_n) + \beta_n + \gamma_n (1-r)\left(r\frac{\gamma_n}{\alpha_n}\right)^{r/(1-r)} - \xi_n
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
&\mathsf{E}\{x_{n+1} \mid \mathcal{F}_n\} \le x_n(1+\bar{\alpha}_n) + \bar{\beta}_n - \xi_n \\
&\bar{\alpha}_n := 2\alpha_n, \ \bar{\beta}_n := \beta_n + (1-r)\, r^{r/(1-r)} \gamma_n \left(\frac{\gamma_n}{\alpha_n}\right)^{r/(1-r)}
\end{aligned}
$$

which after the direct application of the Robbins–Siegmund Theorem 7.11 implies the desired result. □

7.4.4.2 Some corollaries

The following results are the simple corollaries of Theorem 7.11.

Lemma 7.8. *Let $\{\mathcal{F}_n\}_{n\ge 0}$ be a sequence of σ-algebras and η_n, θ_n be $\mathcal{F}_n$-measurable nonnegative random variables such that*

1.

$$
\boxed{\sum_{t=1}^{\infty} \mathsf{E}\{\theta_t\} < \infty} \tag{7.92}
$$

2. for all $n = 1, 2, \ldots$

$$
\boxed{\mathsf{E}\{\eta_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\le} \eta_n + \theta_n} \tag{7.93}
$$

Then

$$
\boxed{\lim_{n\to\infty} \eta_n \overset{a.s.}{=} \eta} \tag{7.94}
$$

Proof. By the assumption 1 it follows that

$$
\sum_{n=1}^{\infty} \theta_n < \infty
$$

Applying then the Robbins–Siegmund Theorem 7.11 for

$$
\alpha_n = 0, \quad \beta_n = \theta_n, \quad \xi_n = 0
$$

we derive the assertion of this lemma. Lemma is proven. □

Lemma 7.9. *Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a sequence of σ-algebras and η_n, θ_n, λ_n, and ν_n be $\mathcal{F}_n$-measurable nonnegative random variables such that*

1.

$$\boxed{\sum_{n=1}^{\infty} \theta_n \overset{a.s.}{<} \infty} \tag{7.95}$$

2.

$$\boxed{\sum_{n=1}^{\infty} \lambda_n \overset{a.s.}{=} \infty, \quad \sum_{n=1}^{\infty} \nu_n \overset{a.s.}{<} \infty} \tag{7.96}$$

3.

$$\boxed{\mathsf{E}(\eta_{n+1} \mid \mathcal{F}_n) \overset{a.s.}{\leq} (1 - \lambda_{n+1} + \nu_{n+1})\eta_n + \theta_n} \tag{7.97}$$

Then

$$\boxed{\lim_{n \to \infty} \eta_n \overset{a.s.}{=} 0} \tag{7.98}$$

Proof. By the assumptions and the Robbins–Siegmund Theorem 7.11 it follows that $\eta_n \underset{n\to\infty}{\overset{a.s.}{\to}} \eta^*$ and $\sum_{n=1}^{\infty} \lambda_{n+1}\eta_n \overset{a.s.}{<} \infty$. As $\sum_{n=1}^{\infty} \lambda_n \overset{a.s.}{=} \infty$, we conclude that there exists a subsequence η_{n_k} which tends to zero with probability 1. Since all subsequences of a convergence sequence have the same limit, it follows that $\eta^* \overset{a.s.}{=} 0$. Lemma is proven. □

Lemma 7.10. *Let $\{v_n\}_{n\geq 0}$ be a sequence of random variables adapted to a sequence $\{\mathcal{F}_n\}_{n\geq 0}$ of σ-algebras $\mathcal{F}_n$, $\{\mathcal{F}_n\} \subseteq \mathcal{F}_{n+1}$ $(n = 0, 1, 2, \ldots)$ such that the random variables $\mathsf{E}(v_n \mid \mathcal{F}_{n-1})$ exist and for some positive monotonically increasing sequence $\{a_n\}_{n\geq 0}$ the following series converges:*

$$\boxed{\sum_{t=1}^{\infty} a_t^{-2} \mathsf{E}\left\{(v_t - \mathsf{E}\{v_t \mid \mathcal{F}_{t-1}\})^2 \mid \mathcal{F}_{t-1})\right\} \overset{a.s.}{<} \infty} \tag{7.99}$$

Then

$$\boxed{\lim_{n\to\infty} \left(\frac{1}{a_n} \sum_{t=1}^{n} v_t - \frac{1}{a_n} \sum_{t=1}^{n} \mathsf{E}\{v_t \mid \mathcal{F}_{t-1}\} \right) \overset{a.s.}{=} 0} \tag{7.100}$$

Proof. Consider the sequence $\{S_n\}_{n\geq 1}$ with the random elements

$$S_n(\omega) := \sum_{t=1}^{n} a_t^{-1} \{v_t - \mathsf{E}(v_t \mid \mathcal{F}_{t-1})\}$$

for which the following relation holds:

$$\mathsf{E}\left\{S_n^2 \mid \mathcal{F}_{n-1})\right\} \stackrel{a.s.}{=} S_{n-1}^2 + a_n^{-2}\mathsf{E}\left\{(v_t - \mathsf{E}\{v_t \mid \mathcal{F}_{t-1}\})^2 \mid \mathcal{F}_{n-1})\right\}$$

By the assumption (7.99) of this lemma and in view of the Robbins–Siegmund Theorem 7.11 we conclude that S_n^2 converges for almost all $\omega \in \Omega$, that is, the limit

$$S_\infty(\omega) := \sum_{t=1}^{\infty} a_t^{-1}\left[v_t - \mathsf{E}(v_t \mid \mathcal{F}_{t-1})\right]$$

exists. But in view of the Kronecker Lemma (see Poznyak (2008)) we conclude that the random sequence with elements given by

$$\frac{1}{a_n}\sum_{t=1}^{n} a_t\left(a_t^{-1}\left[v_t - \mathsf{E}(v_t \mid \mathcal{F}_{t-1})\right]\right) = \frac{1}{a_n}\sum_{t=1}^{n}\left[v_t - \mathsf{E}(v_t \mid \mathcal{F}_{t-1})\right]$$

tends to zero for almost all the random events $\omega \in \Omega$ which implies (7.100). Lemma is proven. □

Corollary 7.7. *For random variables*

$$\boxed{v_n(\omega) = \chi_n(\omega) = \{0; 1\}} \tag{7.101}$$

and

$$\boxed{a_n = n \ (n = 1, 2, \ldots)} \tag{7.102}$$

(and without considering the assumption (7.99)) it follows that

$$\boxed{\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\chi_t - \frac{1}{n}\sum_{t=1}^{n}\mathsf{P}\{\chi_t = 1 \mid \mathcal{F}_{t-1})\}\right) \stackrel{a.s.}{=} 0} \tag{7.103}$$

Proof. It results from the relation $\mathsf{P}\{\chi_t = 1 \mid \mathcal{F}_{t-1}\} = \mathsf{E}\{v_t \mid \mathcal{F}_{t-1}\}$. □

Lemma 7.11. (on the almost-sure convergence rate) *Let $\{v_n\}_{n\geq 0}$ be a sequence of random variables adapted to $\{\mathcal{F}_n\}_{n\geq 0}$ of σ-algebras $\mathcal{F}_n$, $\{\mathcal{F}_n\} \subseteq \mathcal{F}_{n+1}$ $(n = 0, 1, \ldots)$ such that the random variables $\mathsf{E}(v_n \mid \mathcal{F}_{n-1})$ exist and*

$$\boxed{\sum_{t=1}^{\infty} t^{-2}\eta_t \mathsf{E}\left\{(v_t - \mathsf{E}(v_t \mid \mathcal{F}_{t-1}))^2 \mid \mathcal{F}_{t-1})\right\} \stackrel{a.s.}{<} \infty} \tag{7.104}$$

where the deterministic sequence $\{\eta_t\}$ satisfies

$$\boxed{\lim_{n\to\infty}\left(\frac{\eta_n}{\eta_{n-1}}-1\right)n := \lambda < 2} \tag{7.105}$$

Then

$$\boxed{\eta_n S_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{7.106}$$

where

$$\boxed{S_n := \frac{1}{n}\sum_{t=1}^{n}\left(v_t - \mathsf{E}\{v_t \mid \mathcal{F}_{t-1}\}\right)} \tag{7.107}$$

or, in other words,

$$\boxed{S_n \overset{a.s.}{=} o_\omega(1/\sqrt{\eta_n})} \tag{7.108}$$

Proof. One has

$$S_n = S_{n-1}\left(1-\frac{1}{n}\right)+\frac{1}{n}\left(v_n - \mathsf{E}\{v_n/\mathcal{F}_{n-1}\}\right)$$

and therefore for large enough n

$$\begin{aligned}
\mathsf{E}\left\{S_n^2 \mid \mathcal{F}_{n-1})\right\} &\overset{a.s.}{=} S_{n-1}^2\left(1-\frac{1}{n}\right)^2 \\
&\quad + n^{-2}\mathsf{E}\left\{(v_n - \mathsf{E}\{v_n/\mathcal{F}_{n-1}\})^2 \mid \mathcal{F}_{n-1})\right\} \\
&= S_{n-1}^2\left(1-\frac{2+o(1)}{n}\right)^2 \\
&\quad + n^{-2}\mathsf{E}\left\{(v_n - \mathsf{E}(v_n \mid \mathcal{F}_{n-1}))^2 \mid \mathcal{F}_{n-1})\right\}
\end{aligned}$$

Hence, for $W_n := \eta_n S_n^2$, it follows that

$$\begin{aligned}
\mathsf{E}\{W_n \mid \mathcal{F}_{n-1})\} &\overset{a.s.}{=} W_{n-1}\left(1-\frac{2-\lambda+o(1)}{n}\right) \\
&\quad + \eta_n n^{-2}\mathsf{E}\left\{(v_n - \mathsf{E}\{v_n \mid \mathcal{F}_{n-1}\})^2 \mid \mathcal{F}_{n-1})\right\}
\end{aligned}$$

In view of Lemma 7.9 with

$$\lambda_{n+1} := \frac{2-\lambda+o(1)}{n}, \quad \nu_{n+1} := 0$$

$$\theta_n := \eta_n n^{-2}\mathsf{E}\left\{(v_n - \mathsf{E}\{v_n \mid \mathcal{F}_{n-1}\})^2 \mid \mathcal{F}_{n-1})\right\}$$

by the assumptions of this lemma, it follows $W_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ that proves the lemma. □

Corollary 7.8. *If in the assumptions of this lemma*

$$\boxed{n^{-\varkappa}\eta_n = O(1)} \tag{7.109}$$

then

$$\boxed{\lambda = \varkappa} \tag{7.110}$$

and for $\varkappa < 2$

$$\boxed{S_n \overset{a.s.}{=} o_\omega\left(n^{-\frac{\varkappa}{2}}\right)} \tag{7.111}$$

Lemma 7.12. **(Nazin and Poznyak, 1986)** *Let* $\{u_n\}_{n\geq 0}$ *be a sequence of nonnegative random variables* u_n *measurable for all* $n = 1, 2\ldots,$ *with respect to the* σ*-algebra* $\mathcal{F}_n$ *and the following inequality*

$$\boxed{\mathsf{E}\{u_{n+1} \mid \mathcal{F}_n\} \overset{a.s.}{\leq} u_n(1-\alpha_n) + \beta_n} \tag{7.112}$$

holds, where $\{\alpha_n\}$, $\{\beta_n\}$ *are sequences of deterministic variables such that*

$$\boxed{\alpha_n \in (0, 1], \quad \beta_n \geq 0, \quad \sum_{n=1}^{\infty} \alpha_n = \infty} \tag{7.113}$$

and for some nonnegative increasing sequence $\{\nu_n\}_{n\geq 1}$ *the following series converges*

$$\boxed{\sum_{n=1}^{\infty} \beta_n \nu_n < \infty} \tag{7.114}$$

and, the limit

$$\boxed{\lim_{n\to\infty} \frac{\nu_{n+1} - \nu_n}{\alpha_n \nu_n} := \mu < 1} \tag{7.115}$$

exists. Then

$$\boxed{u_n = o_\omega\left(\frac{1}{\nu_n}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{7.116}$$

Proof. Let $\widetilde{u}_n$ be the sequence defined as $\widetilde{u}_n := u_n \nu_n$. Then, based on the assumptions of this lemma, we have

$$\begin{aligned}
\mathsf{E}(\widetilde{u}_{n+1} \mid \mathcal{F}_n) &\overset{a.s.}{\leq} \widetilde{u}_n(1-\alpha_n)\left(\frac{\nu_{n+1}}{\nu_n}\right) + \nu_{n+1}\beta_n \\
&= \widetilde{u}_n(1-\alpha_n)\left(\frac{\nu_{n+1}-\nu_n}{\nu_n} - 1\right) + \nu_{n+1}\beta_n
\end{aligned}$$

and, therefore,

$$\mathsf{E}(\widetilde{u}_{n+1} \mid \mathcal{F}_n) \overset{a.s.}{\leq} \widetilde{u}_n \left[1 - \alpha_n(1-\mu+o(1))\right] + \nu_{n+1}\beta_n$$

Then, from this inequality and Lemma 7.9 we obtain

$$\nu_n \widetilde{u}_n \overset{a.s.}{\to} 0$$

which is equivalent to (7.116). Lemma is proven. □

This page intentionally left blank

8 Limit Theorems as Invariant Laws

Contents

The probability theory has its greatest impact through its limit theorems:

- the *weak law of large numbers* (**LLN**) and its strong version known as the *strong law of large numbers* (**SLLN**),
- the *central limit theorem* (**CLT**),
- the *law of the iterated logarithm* (**LIL**).

All of these limit theorems may be interpreted as some sort of *invariant principles* or *invariant laws* because of the independence of the formulated results (principles) of the distribution of random variables forming considered random sequences.

If $\{x_n\}_{n\geq 1}$ is a stationary sequence of quadratically integrable ($\mathsf{E}\{x_n\} = \mu$, $\mathsf{E}\{x_n^2\} = \sigma^2 > 0$) random (not obligatory independent) variables defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ then denoting

$$\boxed{S_n := \sum_{t=1}^{n} x_t} \tag{8.1}$$

we can formulate these three principle invariant laws as follows:

1. (a) **LLN**:

$$\boxed{n^{-1} S_n \overset{\mathsf{P}}{\underset{n\to\infty}{\to}} \mu} \tag{8.2}$$

   (b) **SLLN**:

$$\boxed{n^{-1} S_n \overset{a.s.}{\underset{n\to\infty}{\to}} \mu} \tag{8.3}$$

2. **CLT**:

$$\boxed{\frac{\sqrt{n}}{\sigma}\left(n^{-1} S_n - \mu\right) \overset{d}{\underset{n\to\infty}{\to}} \mathcal{N}(0,1)} \tag{8.4}$$

   where $\mathcal{N}(0, 1)$ is the standard Gaussian distribution.

3. **LIL**:

$$\boxed{\begin{aligned}&\limsup_{n\to\infty}\frac{\sqrt{n}}{\sigma\sqrt{2\ln\ln n}}\left(n^{-1}S_n-\mu\right)\overset{a.s}{=}1\\&\liminf_{n\to\infty}\frac{\sqrt{n}}{\sigma\sqrt{2\ln\ln n}}\left(n^{-1}S_n-\mu\right)\overset{a.s.}{=}-1\end{aligned}} \tag{8.5}$$

Remark 8.1. *The LNN (8.2) and SLLN (8.3) state the weak and strong convergences, respectively, of the arithmetic average $n^{-1}S_n$ to the arithmetic average of its mathematical expectation, i.e., $n^{-1}\mathsf{E}\{S_n\}=\mu$. This result embodies the idea of probability as a strong limit of relative frequencies (taking $x_n := \chi(A)$ and $\mu := \mathsf{P}\{A\}$, $A\subset\Omega$) and on these grounds may be regarded as the* ***most basic in probability theory****, in fact, the underpinning of the axiomatic theory as a physical realistic subject.*

Remark 8.2. *The convergences (8.2) and (8.3) are not the end of the story: we may hope to say something about the '****rate of convergence****'. The CLT (8.4) and LIL (8.5) may conveniently be interpreted as the rate results about SLLN (8.3). Their usual expressions suppress this basic relationship. So, the CLT (8.4) tells us just the right rate at which to magnify the difference $\left(n^{-1}S_n-\mu\right)$ which is tending to zero, in order to obtain convergence in distribution to a nondegenerate (in fact, Gaussian) law. The delicacy of LIL (8.5) consists in the fact that the norming of $\frac{\sqrt{n}}{\sigma}\left(n^{-1}S_n-\mu\right)$ (as it is in CLT (8.4)) by $\sqrt{2\ln\ln n}$ (as it is in LIL (8.5)) provides a boundary between convergence in probability and convergence with probability one (a.s.) to zero.*

Remark 8.3. *None of the laws (8.2), (8.3), (8.4) and (8.5) depends on the distributions $\mathsf{P}\{x_n\le v\}$ which justifies the name 'invariant laws'.*

The central question discussed in this chapter is: '*Under which conditions do the invariant laws* (8.2)–(8.5) *hold?*' The classical results concerning these laws deal with sums of independent random variables (Petrov, 1975). Here we wish to consider not only independent but also dependent sequences and therefore below we will give different characteristics of stochastic dependency and discuss their interconnection.

8.1 Characteristics of dependence

There exist many notations of dependence. The most important of them are given below.

8.1.1 Main concepts of dependence

8.1.1.1 m-dependence

Definition 8.1. *The random variables $\{x_n\}_{n\ge1}$ defined on the probability space $(\Omega,\mathcal{F},\mathsf{P})$ are said to be* ***m-dependent*** *if x_t and x_k are independent whenever $|t-k|>m$. Independence is treated as 0-dependence.*

Example 8.1. (Peak numbers) *We say that there is a **peak** at x_k if*

$$x_{k-1} < x_k \quad \text{and} \quad x_{k+1} < x_k$$

Define

$$\chi_k := \begin{cases} 1 & \text{if there is a peak at } x_k \\ 0 & \text{if otherwise} \end{cases}$$

Then for independent $x_1, x_2, \ldots$ uniformly distributed on $[0, 1]$, i.e., having the distribution $p_{U(0,1)}(v) = 1$, for $k \geq 2$

$$\begin{aligned} \mathsf{P}\{\chi_k = 1\} &= \mathsf{P}\{\chi(x_{k-1} < x_k)\,\chi(x_{k+1} < x_k)\} \\ &= \int_{v=0}^{1} \mathsf{P}\{\chi(x_{k-1} < v)\,\chi(x_{k+1} < v) \mid x_k = v\}\, p_{U(0,1)}(v)dv \\ &= \int_{v=0}^{1} \mathsf{P}\{\chi(x_{k-1} < v)\,\chi(x_{k+1} < v)\}\, dv \\ &= \int_{v=0}^{1} \mathsf{P}\{\chi(x_{k-1} < v)\}\,\mathsf{P}\{\chi(x_{k+1} < v)\}\, dv \\ &= \int_{v=0}^{1} \mathsf{P}^2\{\chi(x_k < v)\}\, dv \\ &= \int_{v=0}^{1} v^2 dv = 1/3 \end{aligned}$$

We also have

$$\begin{aligned} &\mathsf{P}\{\chi_k = 1, \chi_{k+1} = 1\} = 0 \\ &\mathsf{P}\{\chi_k = 1, \chi_{k+2} = 1\} = 2/15 \\ &\mathsf{P}\{\chi_k = 1, \chi_t = 1\} = 0 \quad \text{for } |k - t| > 2 \ (k, t \geq 2) \end{aligned}$$

So the sequence $\{\chi_k\}_{k\geq 2}$ of the peak-indicators is 2-dependent.

8.1.1.2 Markov dependence

Definition 8.2. *The random variables $\{x_n\}_{n\geq 1}$ defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are said to be **Markov** (or **one-step**) **dependent** if 'the future depends on where one is now, but not on how one got there', or more exactly, if*

$$\boxed{\mathsf{P}\{x_k \leq v \mid x_{k-1}, x_{k-2}, \ldots, x_1\} = \mathsf{P}\{x_k \leq v \mid x_{k-1}\}} \tag{8.6}$$

Example 8.2. *Consider the dynamic plant given by*

$$x_{n+1} = Ax_n + \xi_n, \quad n \geq 0$$

where $\{\xi_n\}_{n\geq 0}$ *is a sequence of independent random variables. Then*

$$\begin{aligned} \mathsf{P}\{x_k \leq v \mid x_{k-1}, x_{k-2}, \ldots, x_1\} &= \mathsf{P}\{Ax_{k-1} + \xi_{k-1} \leq v \mid x_{k-1}, x_{k-2}, \ldots, x_1\} \\ &= \mathsf{P}\{\xi_{k-1} \leq v - Ax_{k-1} \mid x_{k-1}, x_{k-2}, \ldots, x_1\} = \mathsf{P}\{\xi_{k-1} \leq v - Ax_{k-1}\} \end{aligned}$$

So $\{x_n\}_{n\geq 0}$ *is a Markov-dependent sequence.*

8.1.1.3 Martingale dependence

Definition 8.3. *The random variables* $\{x_n\}_{n\geq 1}$ *defined on the probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *are said to be* ***martingale-dependent*** *if*

$$\boxed{\mathsf{E}\{x_k \mid \mathcal{F}_{k-1}\} \overset{a.s.}{=} x_{k-1}} \tag{8.7}$$

or, equivalently, if

$$\boxed{\mathsf{E}\{x_k - x_{k-1} \mid \mathcal{F}_{k-1}\} \overset{a.s.}{=} 0} \tag{8.8}$$

where $\mathcal{F}_{k-1} = \sigma(x_1, x_2, \ldots x_{k-1})$.

Example 8.3. *The sums* S_n *(8.1) of independent random variables* $\{x_n\}_{n\geq 1}$ *are martingale-dependent since*

$$\begin{aligned} \mathsf{E}\{S_k \mid \mathcal{F}_{k-1}\} = {} & \mathsf{E}\{S_{k-1} + x_k \mid \mathcal{F}_{k-1}\} \\ & \overset{a.s.}{=} S_{k-1} + \mathsf{E}\{x_k \mid \mathcal{F}_{k-1}\} \overset{a.s.}{=} S_{k-1} + \mathsf{E}\{x_k\} = S_{k-1} \end{aligned}$$

8.1.1.4 Mixed sequences

Let $\mathcal{H}$ and $\mathcal{G}$ be sub-σ-algebras of $\mathcal{F}$. The following values are some measures of dependence of these σ-algebras:

1. the coefficient of *strong mixing* (Rosenblatt, 1956):

$$\alpha(\mathcal{H}, \mathcal{G}) := \sup_{A\in\mathcal{H}, B\in\mathcal{G}} |\mathsf{P}\{A \cap B\} - \mathsf{P}\{A\}\mathsf{P}\{B\}| \tag{8.9}$$

2. the coefficient of *uniform strong mixing* (Ibragimov, 1962):

$$\begin{aligned} \phi(\mathcal{H}, \mathcal{G}) &:= \sup_{A\in\mathcal{H}, B\in\mathcal{G}, \mathsf{P}\{B\}>0} \mathsf{P}^{-1}\{B\} |\mathsf{P}\{A \cap B\} - \mathsf{P}\{A\}\mathsf{P}\{B\}| \\ &= \sup_{A\in\mathcal{H}, B\in\mathcal{G}} |\mathsf{P}\{A/B\} - \mathsf{P}\{A\}| \end{aligned} \tag{8.10}$$

3. the coefficient of *relative uniform strong mixing* (Blum et al., 1963):

$$\psi(\mathcal{H},\mathcal{G}) := \sup_{A\in\mathcal{H},B\in\mathcal{G},\mathsf{P}\{B\}>0} \mathsf{P}^{-1}\{A\}\,\mathsf{P}^{-1}\{B\}\left|\mathsf{P}\{A\cap B\}-\mathsf{P}\{A\}\,\mathsf{P}\{B\}\right| = \sup_{A\in\mathcal{H},B\in\mathcal{G},\mathsf{P}\{A\}>0,\quad \mathsf{P}\{B\}>0}\left|\frac{\mathsf{P}\{A\cap B\}}{\mathsf{P}\{A\}\,\mathsf{P}\{B\}}-1\right| \tag{8.11}$$

4. the coefficient of *correlative mixing* (Iosifecsu and Theodorescu, 1969):

$$\rho(\mathcal{H},\mathcal{G}) := \sup_{x\in\mathcal{H},y\in\mathcal{G}}\left|\rho_{x,y}\right| \qquad \rho_{x,y} := \frac{\operatorname{cov}(x,x)}{\sqrt{\operatorname{var} x}\sqrt{\operatorname{var} y}} \text{ - the correlation coefficient (3.51)} \tag{8.12}$$

Proposition 8.1. *It is not difficult to show that the following relations hold:*

$$\boxed{\begin{array}{c} \alpha(\mathcal{H},\mathcal{G}) \le 1/4, \quad \psi(\mathcal{H},\mathcal{G}) \le 1, \quad \rho(\mathcal{H},\mathcal{G}) \le 1 \\ 4\alpha(\mathcal{H},\mathcal{G}) \le 2\phi(\mathcal{H},\mathcal{G}) \le \psi(\mathcal{H},\mathcal{G}) \\ 4\alpha(\mathcal{H},\mathcal{G}) \le \rho(\mathcal{H},\mathcal{G}) \le \psi(\mathcal{H},\mathcal{G}),\ \rho(\mathcal{H},\mathcal{G}) \le 2\sqrt{\phi(\mathcal{H},\mathcal{G})} \end{array}} \tag{8.13}$$

To extend these notions to random variables let us define the following σ-algebras:

$$\bigvee_{i=m}^{n}\mathcal{F}_i := \sigma\left(\bigcup_{i=m}^{n}\mathcal{F}_i\right) \tag{8.14}$$

that is, $\bigvee_{i=m}^{n}\mathcal{F}_i$ is the sigma-algebra generated by the union of the sigma-algebras $\mathcal{F}_i$ $(i = m,\dots,n)$.

Introduce the following mixing coefficients for the random sequence $\{x_n\}_{n\ge1}$:

$$\boxed{\begin{array}{l} \alpha_n := \sup\limits_{k\ge1}\alpha\left(\bigvee\limits_{i=1}^{k}\mathcal{F}_i,\ \bigvee\limits_{i=k+n}^{\infty}\mathcal{F}_i\right), \\ \phi_n := \sup\limits_{k\ge1}\phi\left(\bigvee\limits_{i=1}^{k}\mathcal{F}_i,\ \bigvee\limits_{i=k+n}^{\infty}\mathcal{F}_i\right) \\ \psi_n := \sup\limits_{k\ge1}\psi\left(\bigvee\limits_{i=1}^{k}\mathcal{F}_i,\ \bigvee\limits_{i=k+n}^{\infty}\mathcal{F}_i\right) \\ \rho_n := \sup\limits_{k\ge1}\rho\left(\bigvee\limits_{i=1}^{k}\mathcal{F}_i,\ \bigvee\limits_{i=k+n}^{\infty}\mathcal{F}_i\right) \end{array}} \tag{8.15}$$

Definition 8.4. *The sequence of random variables* $\{x_n\}_{n\geq 1}$ *is said to be*

(a) α*-**mixing** or **strong mixing** if*

$$\alpha_n \underset{n\to\infty}{\to} 0$$

(b) ϕ*,* ψ ***and*** ρ***-mixing** if*

$$\phi_n, \psi_n, \rho_n \underset{n\to\infty}{\to} 0$$

respectively.

Remark 8.4. *Obviously, in view of the relations in (8.13),* ϕ*,* ψ *and* ρ*-mixing imply* α*-mixing or strong mixing. This means that proving some invariant law for* ψ*-mixing we, in the same time, obtain the validity of the considered invariant law for all other types of mixing, i.e., for* α*,* ϕ*, and* ρ*-mixing too.*

8.1.1.5 Mixingale-difference sequences

In McLeish (1975) there was introduced the concept of *mixingales* (or asymptotic martingales) which are sufficiently like martingale differences to satisfy a convergence theorem.

Definition 8.5. *Let* $\{x_n\}_{n\geq 1}$ *be a sequence of square-integrable random variables defined on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$*, and let* $\{\mathcal{F}_n\}_{n\geq 1}$ *be an increasing sequence of sub-*σ*-fields of* $\mathcal{F}$*. Then* $\{x_n, \mathcal{F}_n\}$ *is called a **mixingale** (difference) **sequence** if there exist sequences* $\{c_n\}_{n\geq 1}$ *and* $\{\psi_m\}_{m\geq 0}$ *of nonnegative real numbers such that*

1.

$$\boxed{\psi_m \underset{m\to\infty}{\to} 0} \tag{8.16}$$

2.

$$\boxed{q_{n,m} := \sqrt{\mathsf{E}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}|^2\right\}} \leq \psi_m c_n} \tag{8.17}$$

3.

$$\boxed{\sqrt{\mathsf{E}\left\{|x_n - \mathsf{E}\{x_n \mid \mathcal{F}_{n+m}\}|^2\right\}} \leq \psi_{m+1} c_n} \tag{8.18}$$

The following examples give an idea of the scope of mixingales.

Example 8.4. **(A martingale as a mixingale)** *Let* $\{x_n, \mathcal{F}_n\}$ *be a martingale, that is,*

$$\mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\} \overset{a.s.}{=} 0, \quad \mathsf{E}\{x_n \mid \mathcal{F}_{n+m}\} \overset{a.s.}{=} x_n \quad \textit{for} \quad m \geq 0$$

Then $\{x_n, \mathcal{F}_n\}$ *in the same time is a mixingale with* $c_n = \sqrt{\mathsf{E}\left\{x_n^2\right\}}$ *and*

$$\psi_0 = 1, \psi_m = 0 \quad \textit{for} \quad m \geq 1$$

Example 8.5. (Linear process as a mixingale) *Let*

$$x_n = \sum_{t=-\infty}^{\infty} \alpha_{t-n}\xi_t, \qquad \sum_{t=-\infty}^{\infty} \alpha_t^2 < \infty \tag{8.19}$$

be a linear process generated by a sequence $\{\xi_t\}_{t\geq-\infty}$ *of independent zero-mean random variables with the finite variance* σ^2*. The sequence* $\{\alpha_t\}_{t\geq-\infty}$ *is deterministic. Define* $\mathcal{F}_n := \sigma(\ldots, \xi_{n-1}, \xi_n)$*. Then* $\{x_n, \mathcal{F}_n\}$ *is a mixingale with*

$$c_n = \sigma \quad \textit{and} \quad \psi_m^2 = \sum_{i=-\infty}^{-m} \alpha_i^2$$

Indeed,

$$q_{n,m} := \sqrt{\mathsf{E}\left\{\left|\mathsf{E}\left\{x_n \mid \mathcal{F}_{n-m}\right\}\right|^2\right\}} = \sqrt{\mathsf{E}\left\{\left|\sum_{t=-\infty}^{n-m} \alpha_{t-n}\xi_t\right|^2\right\}}$$

$$= \sigma\sqrt{\sum_{t=-\infty}^{n-m} \alpha_{t-n}^2} = \sigma\sqrt{\sum_{i=-\infty}^{-m} \alpha_i^2} = \sigma = c_n\psi_m$$

8.1.1.6 Correlated sequences

Definition 8.6. *The double-index sequence* $\left\{\rho_{n,m}\right\}_{m,n\geq1}$ *is said to be a* ***correlation sequence*** *if* $\rho_{n,m}$ *are defined by*

$$\boxed{\rho_{n,m} := \mathsf{E}\left\{(x_n - \mathsf{E}\left\{x_n\right\})(x_m - \mathsf{E}\left\{x_m\right\})\right\}} \tag{8.20}$$

Usually, $\rho_{n,m}$ *is called a* ***correlation function****.*

Example 8.6. (Martingale-difference) *Let* $\{x_n, \mathcal{F}_n\}_{n\geq0}$ *be a martingale-difference, that is,*

$$\mathsf{E}\left\{x_n\right\} = 0, \quad \mathsf{E}\left\{x_n \mid \mathcal{F}_{n-k}\right\} \overset{a.s.}{=} 0 \quad (k \geq 1)$$

Let, for example, $m < n$*. Then*

$$\rho_{n,m} := \mathsf{E}\left\{(x_n - \mathsf{E}\left\{x_n\right\})(x_m - \mathsf{E}\left\{x_m\right\})\right\} = \rho_{n,m} := \mathsf{E}\left\{x_n x_m\right\}$$
$$= \mathsf{E}\left\{\mathsf{E}\left\{x_n x_m \mid \mathcal{F}_m\right\}\right\} = \mathsf{E}\left\{x_m \mathsf{E}\left\{x_n \mid \mathcal{F}_m\right\}\right\} = \mathsf{E}\left\{x_m \cdot 0\right\} = 0$$

The partial case of a martingale-difference is a sequence of independent random variables. So, for them, $\rho_{n,m} = 0$ *if* $n \neq m$*.*

8.1.2 Some inequalities for the covariance and mixing coefficients

Lemma 8.1. *Let $A \in \mathcal{H}$ and $B \in \mathcal{G}$. Then*

$$\boxed{|\mathsf{P}\{A \cap B\} - \mathsf{P}\{A\}\mathsf{P}\{B\}| \le \phi(\mathcal{H},\mathcal{G})\mathsf{P}\{B\}} \tag{8.21}$$

Proof. Since

$$\begin{aligned}
|\mathsf{P}\{A \cap B\} - \mathsf{P}\{A\}\mathsf{P}\{B\}| &= \left| \int_B [\mathsf{P}\{A/B\} - \mathsf{P}\{A\}]\,\mathsf{P}(d\omega) \right| \\
&\le \int_B |\mathsf{P}\{A/B\} - \mathsf{P}\{A\}|\,\mathsf{P}(d\omega) \\
&\le \int_B \sup_{A\in\mathcal{H},\quad B\in\mathcal{G}} |\mathsf{P}\{A/B\} - \mathsf{P}\{A\}|\,\mathsf{P}(d\omega) \\
&\le \phi(\mathcal{H},\mathcal{G}) \int_B \mathsf{P}(d\omega) = \phi(\mathcal{H},\mathcal{G})\mathsf{P}\{B\}
\end{aligned}$$

the desired inequality (8.21) is proven. □

Theorem 8.1. **(Ibragimov, 1962)** *Suppose that x and y are random variables which are $\mathcal{H}$ and $\mathcal{G}$ measurable, i.e., $x \in \mathcal{H}$ and $y \in \mathcal{G}$, and they have bounded p and q moments, respectively, namely*

$$\boxed{\begin{array}{c} \mathsf{E}\left\{|x|^p\right\} < \infty \quad and \quad \mathsf{E}\left\{|x|^q\right\} \\ p, q > 1, \quad p^{-1} + q^{-1} = 1 \end{array}} \tag{8.22}$$

Then

$$\boxed{|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| \le 2\phi^{1/p}(\mathcal{H},\mathcal{G})\mathsf{E}^{1/p}\left\{|x|^p\right\}\mathsf{E}^{1/q}\left\{|y|^q\right\}} \tag{8.23}$$

The result continue to hold for $p = 1$ and $q = \infty$ where

$$\mathsf{E}^{1/\infty}\left\{|y|^\infty\right\} := ess\sup|y| = \inf\{C \mid \mathsf{P}\{|y| > C\} = 0\} \tag{8.24}$$

namely,

$$\boxed{|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| \le 2\phi(\mathcal{H},\mathcal{G})\mathsf{E}\{|x|\}\,ess\sup|y|} \tag{8.25}$$

Proof. It suffices to consider only the case of simple random variables (since the general case follows from Theorem 2.3 on monotone approximation):

$$x = \sum_i a_i I_{A_i}(\omega), \quad A_{i_1} \bigcap_{i_1 \ne i_2} A_{i_2} = \varnothing, \quad A_i \subset \mathcal{H}$$

$$y = \sum_j b_j I_{B_j}(\omega), \quad B_{j_1} \bigcap_{j_1 \ne j_2} A_{j_2} = \varnothing, \quad B_j \subset \mathcal{G}$$

So, one has

$$
\begin{aligned}
\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\} &= \sum_i \sum_j a_i b_j \mathsf{P}\left\{A_i \cap B_j\right\} - \sum_i \sum_j a_i b_j \mathsf{P}\{A_i\}\mathsf{P}\left\{B_j\right\} \\
&= \sum_i \sum_j a_i b_j \left[\mathsf{P}\left\{A_i \cap B_j\right\} - \mathsf{P}\{A_i\}\mathsf{P}\left\{B_j\right\}\right] \\
&= \sum_i \sum_j a_i \mathsf{P}\{A_i\} b_j \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right] \\
&= \sum_i a_i \mathsf{P}\{A_i\} \sum_j b_j \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right] \\
&= \sum_i a_i \mathsf{P}^{1/p}\{A_i\} \sum_j b_j \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right] \mathsf{P}^{1/q}\{A_i\}
\end{aligned}
$$

By the Hölder inequality (13.73) it follows that

$$
\begin{aligned}
|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| &\le \left(\sum_i \left(|a_i|\,\mathsf{P}^{1/p}\{A_i\}\right)^p\right)^{1/p} \\
&\quad \cdot \left(\sum_i \left[\sum_j |b_j| \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right] \mathsf{P}^{1/q}\{A_i\}\right]^q\right)^{1/q} \\
&\le \left(\sum_i \left(|a_i|^p\,\mathsf{P}\{A_i\}\right)\right)^{1/p} \\
&\quad \cdot \left(\sum_i \mathsf{P}\{A_i\} \left[\sum_j |b_j| \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right]\right]^q\right)^{1/q} \\
&\le \mathsf{E}^{1/p}\left\{|x|^p\right\} \left(\sum_i \mathsf{P}\{A_i\} \cdot \left[\sum_j |b_j| \left|\mathsf{P}\left\{B_j/A_i\right\}\right.\right.\right. \\
&\quad \left.\left.\left. - \mathsf{P}\left\{B_j\right\}\right|^{1/q} \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|^{1/p}\right]^q\right)^{1/q}
\end{aligned}
$$

Applying secondly the Hölder inequality (13.73) we also have

$$
\begin{aligned}
&\sum_j \left(|b_j| \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|^{1/q}\right)\left(\left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|^{1/p}\right) \\
&\quad \le \left(\sum_j |b_j|^q \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|\right)^{1/q} \left(\sum_j \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|\right)^{1/p} \\
&\quad \le \left(\sum_j |b_j|^q \left(\mathsf{P}\left\{B_j/A_i\right\} + \mathsf{P}\left\{B_j\right\}\right)\right)^{1/q} \left(\sum_j \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|\right)^{1/p}
\end{aligned}
$$

that implies

$$\left(\sum_i \mathsf{P}\{A_i\}\left[\sum_j |b_j|\left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right|^{1/q}\left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right|^{1/p}\right]^q\right)^{1/q}$$

$$\leq \left(\sum_i \mathsf{P}\{A_i\}\left[\left(\sum_j |b_j|^q\left(\mathsf{P}\{B_j/A_i\}+\mathsf{P}\{B_j\}\right)\right)^{1/q}\right.\right.$$

$$\left.\left.\cdot\left(\sum_j \left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right|\right)^{1/p}\right]^q\right)^{1/q}$$

$$=\left(\sum_i \mathsf{P}\{A_i\}\left(\sum_j |b_j|^q\left(\mathsf{P}\{B_j/A_i\}+\mathsf{P}\{B_j\}\right)\right)\right.$$

$$\left.\cdot\left(\sum_j \left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right|\right)^{q/p}\right)^{1/q}$$

$$\leq \left[\sum_i \mathsf{P}\{A_i\}\left(\sum_j |b_j|^q\left(\mathsf{P}\{B_j/A_i\}+\mathsf{P}\{B_j\}\right)\right)\gamma^{q/p}\right]^{1/q}$$

$$=\left(\sum_j |b_j|^q\left(\sum_i \mathsf{P}\{A_i\}\mathsf{P}\{B_j/A_i\}+\mathsf{P}\{B_j\}\right)\right)^{1/q}\gamma^{1/p}$$

$$=2^{1/q}\mathsf{E}^{1/q}\left\{|y|^q\right\}\gamma^{1/p}$$

where

$$\gamma := \sup_i \sum_j \left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right|$$

Further, if C_i^+ (respectively, C_i^-) is the union of those B_j for which $\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\geq 0$ (respectively, $\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}<0$) then C_i^+, C_i^- are elements of $\mathcal{G}$, and by Lemma 8.1 it follows that

$$\gamma := \sup_i \sum_j \left|\mathsf{P}\{B_j/A_i\}-\mathsf{P}\{B_j\}\right| = \sup_i \left|\left[\mathsf{P}\{C_i^+/A_i\}-\mathsf{P}\{C_i^+\}\right]\right.$$
$$\left.+\left[\mathsf{P}\{C_i^+/A_i\}-\mathsf{P}\{C_i^+\}\right]\right| \leq 2\phi(\mathcal{H},\mathcal{G})$$

and, therefore,

$$|\mathsf{E}\{xy\}-\mathsf{E}\{x\}\mathsf{E}\{y\}| \leq \mathsf{E}^{1/p}\left\{|x|^p\right\}2^{1/q}\mathsf{E}^{1/q}\left\{|y|^q\right\}\gamma^{1/p}$$
$$\leq 2^{1/q+1/p}\mathsf{E}^{1/p}\left\{|x|^p\right\}\mathsf{E}^{1/q}\left\{|y|^q\right\}\phi(\mathcal{H},\mathcal{G})^{1/p}$$

that gives (8.23). The case $p = 1$ and $q = \infty$ can be considered similar if we take into account that $\left|b_j\right| \overset{a.s.}{\leq} ess\sup|y|$ implies

$$\begin{aligned}
|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| &\leq \left|\sum_i a_i \mathsf{P}\{A_i\} \sum_j b_j \left[\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right]\right| \\
&\leq ess\sup|y| \left|\sum_i |a_i| \mathsf{P}\{A_i\} \sum_j \left|\mathsf{P}\left\{B_j/A_i\right\} - \mathsf{P}\left\{B_j\right\}\right|\right| \\
&\leq 2\phi(\mathcal{H},\mathcal{G})\mathsf{E}\{|x|\}\, ess\sup|y|
\end{aligned}$$

Theorem is proven. □

Corollary 8.1. *If x and y are $\mathcal{H}$ and $\mathcal{G}$ measurable, respectively, i.e., $x \in \mathcal{H}$ and $y \in \mathcal{G}$, and they have second bounded moments, then*

$$\boxed{|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| \leq 2\sqrt{\phi(\mathcal{H},\mathcal{G})}\,(\operatorname{var} x + \operatorname{var} y)} \tag{8.26}$$

Proof. Notice that

$$\begin{aligned}
&\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\} = \mathsf{E}\{(x - \mathsf{E}\{x\})(y - \mathsf{E}\{y\})\} = \mathsf{E}\{\mathring{x}\mathring{y}\} \\
&\mathring{x} := x - \mathsf{E}\{x\}, \quad \mathsf{E}\{\mathring{x}\} = 0, \quad \mathring{y} := y - \mathsf{E}\{y\}, \quad \mathsf{E}\{\mathring{y}\} = 0
\end{aligned}$$

Then (8.25), applied for $\mathring{x}$ and $\mathring{y}$ under $p = q = 2$, gives

$$|\mathsf{E}\{xy\} - \mathsf{E}\{x\}\mathsf{E}\{y\}| = |\mathsf{E}\{\mathring{x}\mathring{y}\}| \leq 2\sqrt{\phi(\mathcal{H},\mathcal{G})}\mathsf{E}^{1/2}\left\{|\mathring{x}|^2\right\}\mathsf{E}^{1/2}\left\{|\mathring{y}|^2\right\}$$

that, after the elementary inequality $2ab \leq a^2 + b^2$ application, implies (8.26). □

Example 8.7. (Uniformly mixing processes) *Let $\{x_n\}_{n\geq 1}$ be a centered process satisfying the uniform mixing condition, namely,*

$$\phi_n := \sup_{k\geq 1}\phi\left(\bigvee_{i=1}^{k}\mathcal{F}_i, \bigvee_{i=k+n}^{\infty}\mathcal{F}_i\right) \underset{n\to\infty}{\to} 0$$

where $\phi(\mathcal{H},\mathcal{G}) := \sup_{A\in\mathcal{H}, B\in\mathcal{G}} |\mathsf{P}\{A/B\} - \mathsf{P}\{A\}|$ as in (8.10). Then, by Ibragimov's formula (8.23) taken for $p = q = 2$, if we put $\mathcal{F}_n := \sigma(x_1 \ldots, x_n)$ it follows that

$$\begin{aligned}
&\mathsf{E}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}|^2\right\} = \mathsf{E}\{\mathsf{E}\{x_n\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\} \mid \mathcal{F}_{n-m}\}\} \\
&\quad = \mathsf{E}\{x_n\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}\} \leq 2\sqrt{\phi_m}\mathsf{E}^{1/2}\left\{x_n^2\right\}\mathsf{E}^{1/2}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}|^2\right\}
\end{aligned}$$

so that

$$\sqrt{\mathsf{E}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\}|^2\right\}} \leq 2\sqrt{\phi_m}\mathsf{E}^{1/2}\left\{x_n^2\right\}$$

or equivalently

$$\boxed{\mathsf{E}\left\{\left|\mathsf{E}\left\{x_n \mid \mathcal{F}_{n-m}\right\}\right|^2\right\} \leq 4\phi_m \mathsf{E}\left\{x_n^2\right\}} \tag{8.27}$$

and hence, $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *is a mixingale with*

$$c_n = 2\mathsf{E}^{1/2}\left\{x_n^2\right\} \quad \text{and} \quad \psi_m = \sqrt{\phi_m}$$

Corollary 8.2. *If* x_i *are* $\mathcal{H}_i$ *measurable with the second bound moments then*

$$\boxed{\begin{array}{c}\left|\operatorname{var}\left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} \operatorname{var} x_i\right| \leq \\ 4\max\left\{\max_{1\leq i<n} \sum_{j=i+1}^{n} \sqrt{\phi(\mathcal{H}_j, \mathcal{H}_j)},\ \max_{1\leq i<n} \sum_{i=1}^{j+1} \sqrt{\phi(\mathcal{H}_j, \mathcal{H}_j)}\right\} \sum_{i=1}^{n} \operatorname{var} x_i\end{array}}$$

Proof. It follows from the previous corollary taking into account that

$$\left|\operatorname{var}\left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} \operatorname{var} x_i\right| \leq 2 \sum_{1\leq i<j\leq n} \left|\mathsf{E}\left\{x_i x_j\right\} - \mathsf{E}\left\{x_i\right\}\mathsf{E}\left\{x_j\right\}\right| \qquad \square$$

8.1.3 Analog of Doob's inequality for mixingales

The key to a mixingale convergence theorem lies in establishing a mixingale analog of Doob's theorem 7.7 of the Kolmogorov type.

Theorem 8.2. **(McLeish, 1975)** *If* $\{x_n, \mathcal{F}_n\}$ *is a mixingale (see Definition 8.5) with some* $\{c_n\}_{n\geq 1}$ *and with*

$$\boxed{\psi_m = O\left(\frac{1}{\sqrt{m}\log^2 m}\right) \underset{m\to\infty}{\to} 0} \tag{8.28}$$

then there exists a constant K *(depending only on* $\{\psi_m\}$*) such that for* $S_i := \sum_{j=1}^{i} x_j$ *it follows that*

$$\boxed{\mathsf{E}\left\{\max_{1\leq i\leq n} S_i\right\} \leq K \sum_{i=1}^{n} c_i^2} \tag{8.29}$$

Proof. By the properties (8.17) and (8.18)

$$\mathsf{E}\{x_n \mid \mathcal{F}_{n-m}\} \underset{m\to\infty}{\overset{a.s.}{\to}} 0$$
$$x_n - \mathsf{E}\{x_n \mid \mathcal{F}_{n+m}\} \underset{m\to\infty}{\overset{a.s.}{\to}} 0$$

so that x_i can be represented as

$$x_i \overset{a.s.}{=} \sum_{m=-\infty}^{\infty} \left[\mathsf{E}\{x_i \mid \mathcal{F}_{i+m}\} - \mathsf{E}\{x_i \mid \mathcal{F}_{i+m-1}\}\right]$$

Define

$$y_{n,k} := \sum_{i=1}^{n} \left[\mathsf{E}\{x_i \mid \mathcal{F}_{i+k}\} - \mathsf{E}\{x_i \mid \mathcal{F}_{i+k-1}\}\right]$$

Then S_n can be represented as

$$\begin{aligned} S_n &:= \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \sum_{m=-\infty}^{\infty} \left[\mathsf{E}\{x_i \mid \mathcal{F}_{i+m}\} - \mathsf{E}\{x_i \mid \mathcal{F}_{i+m-1}\}\right] \\ &= \sum_{m=-\infty}^{\infty} \sum_{i=1}^{n} \left[\mathsf{E}\{x_i \mid \mathcal{F}_{i+m}\} - \mathsf{E}\{x_i \mid \mathcal{F}_{i+m-1}\}\right] = \sum_{m=-\infty}^{\infty} y_{n,m} \end{aligned}$$

Setting

$$a_0 = a_1 = a_{-1} := 1$$
$$a_n = a_{-n} := \frac{1}{n \log^2 n} \quad \text{for } n \geq 2$$

by the Cauchy–Bounyakovski–Schwartz inequality one has

$$\begin{aligned} S_n &= \left(\sum_{m=-\infty}^{\infty} y_{n,m}\right)^2 = \left(\sum_{m=-\infty}^{\infty} \sqrt{a_m}\left[a_m^{-1/2} y_{n,m}\right]\right)^2 \\ &\leq \left(\sum_{m=-\infty}^{\infty} a_m\right)\left(\sum_{m=-\infty}^{\infty} a_m^{-1} y_{n,m}^2\right) \end{aligned}$$

that implies

$$\max_{1\leq i\leq n} S_i \leq \left(\sum_{m=-\infty}^{\infty} a_m\right)\left(\sum_{m=-\infty}^{\infty} a_m^{-1} \max_{1\leq i\leq n} y_{i,m}^2\right)$$

and

$$\mathsf{E}\left\{\max_{1\leq i\leq n} S_i\right\} \leq \left(\sum_{m=-\infty}^{\infty} a_m\right)\left(\sum_{m=-\infty}^{\infty} a_m^{-1} \mathsf{E}\left\{\max_{1\leq i\leq n} y_{i,m}^2\right\}\right)$$

Here notice that $\sum_{m=-\infty}^{\infty} a_m$ converges by Corollary 16.25 in Poznyak (2008). For each k the sequence $\left\{y_{i,k}, \mathcal{F}_{i+k}\right\}$ $\left(i = \overline{1,n}\right)$ is a martingale, and, therefore, by Doob's moment inequality (7.75) applied for $p = 2$ it follows that

$$\mathsf{E}\left\{\max_{1\le i\le n} S_i\right\} \le 4\left(\sum_{m=-\infty}^{\infty} a_m\right)\left(\sum_{m=-\infty}^{\infty} a_m^{-1}\mathsf{E}\left\{y_{n,m}^2\right\}\right) \tag{8.30}$$

Defining now for $m \ge 0$

$$z_{i,m} := x_i - \mathsf{E}\left\{x_i \mid \mathcal{F}_{i+m}\right\}$$

we get

$$\begin{aligned}
\mathsf{E}\left\{y_{n,m}^2\right\} &= \sum_{i=1}^{n}\left[\mathsf{E}\left\{(\mathsf{E}\left\{x_i \mid \mathcal{F}_{i+m}\right\})^2\right\} - \mathsf{E}\left\{(\mathsf{E}\left\{x_i \mid \mathcal{F}_{i+m-1}\right\})^2\right\}\right] \\
&= \sum_{i=1}^{n}\left(\mathsf{E}\left\{z_{i,m-1}^2\right\} - \mathsf{E}\left\{z_{i,m}^2\right\}\right)
\end{aligned}$$

Substituting this identity into (8.30) for $m \ge 1$ we deduce that

$$\begin{aligned}
\mathsf{E}\left\{\max_{1\le i\le n} S_i\right\} &\le 4\left(\sum_{m=-\infty}^{\infty} a_m\right)\sum_{i=1}^{n}\left[a_0^{-1}\mathsf{E}\left\{(\mathsf{E}\left\{x_i \mid \mathcal{F}_i\right\})^2\right\} + a_1^{-1}\mathsf{E}\left\{z_{i,0}^2\right\}\right. \\
&\quad \left. + \sum_{m=1}^{\infty}\left(a_{m+1}^{-1} - a_m^{-1}\right)\mathsf{E}\left\{z_{i,m}^2\right\} + \sum_{m=1}^{\infty}\left(a_m^{-1} - a_{m-1}^{-1}\right)\mathsf{E}\left\{(\mathsf{E}\left\{x_i \mid \mathcal{F}_{i-m}\right\})^2\right\}\right]
\end{aligned}$$

Bounding the right-hand side of this inequality using the mixingale properties (8.17) and (8.18) we finally get

$$\begin{aligned}
\mathsf{E}\left\{\max_{1\le i\le n} S_i\right\} &\le 4\left(\sum_{m=-\infty}^{\infty} a_m\right)\left(\sum_{i=1}^{n} c_i^2\right) \\
&\quad \cdot\left[a_0^{-1}\left(\psi_0^2 + \psi_1^2\right) + 2\sum_{m=1}^{\infty}\left(a_m^{-1} - a_{m-1}^{-1}\right)\psi_m^2\right]
\end{aligned}$$

that completes the proof. □

Corollary 8.3. *The constant K in (8.28) is equal to*

$$\begin{aligned}
K &= 4\left(3 + \sum_{m=2}^{\infty}\frac{1}{m\log^2 m}\right) \\
&\quad \times\left[\left(\psi_0^2 + \psi_1^2\right) + 2\sum_{m=2}^{\infty}\left[m\log^2 m - (m-1)\log^2(m-1)\right]\psi_m^2\right]
\end{aligned}$$

8.2 Law of large numbers

Consider a sequence $\{x_n\}_{n\geq 1}$ of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and having finite first moments (mathematical expectations) $\{m_n\}_{n\geq 1}$, namely,

$$\boxed{\mathsf{E}\{x_n\} = m_n \ (n \geq 1)} \tag{8.31}$$

In the most general form, the *law of large numbers* concerns the conditions providing the asymptotic equivalence between the arithmetic-average of $\{x_n\}_{n\geq 1}$ and its mathematical expectation, that is, this law deals with the conditions guaranteeing that

$$\boxed{\begin{gathered} n^{-1}S_n - n^{-1}\mathsf{E}\{S_n\} = n^{-1}S_n - n^{-1}\sum_{t=1}^{n} m_t \underset{n\to\infty}{\to} 0 \\ S_n := \sum_{t=1}^{n} x_t \end{gathered}} \tag{8.32}$$

in some probabilistic sense. Rewriting (8.32) in the equivalent form, using the centered random variables $\mathring{x}_n := x_n - m_n \ (n \geq 1)$, one has

$$\boxed{n^{-1}\mathring{S}_n \underset{n\to\infty}{\to} 0, \quad \mathring{S}_n := \sum_{t=1}^{n} \mathring{x}_t} \tag{8.33}$$

So, without the loss of generality we will suppose that a sequence $\{x_n\}_{n\geq 1}$ consists of centered random variables and the symbol '$\circ$' upon the variable x_n will be omitted for simplicity.

8.2.1 Weak law of large numbers

As has been mentioned in (8.2) we will be interested in the conditions guaranteeing

$$\boxed{n^{-1}S_n \overset{\mathsf{P}}{\underset{n\to\infty}{\to}} 0} \tag{8.34}$$

The theorem below deals with a *martingale version* of the classical degenerate convergence criterion.

Theorem 8.3. (on weak LLN (Hall and Heyde, 1980)) *Let $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ be a martingale-difference $\left(\mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\} \overset{a.s.}{=} 0, n \geq 1\right)$ and $\{b_n\}_{n\geq 1}$ a sequence of positive constants with*

$$0 < b_n \underset{n\to\infty}{\uparrow} \infty \tag{8.35}$$

Then, writing

$$x_{ni} := x_i \chi\left(|x_i| \leq b_n\right), \quad 1 \leq i \leq n \tag{8.36}$$

we have that (8.34) holds, namely,

$$b_n^{-1} S_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0 \tag{8.37}$$

if as $n \to \infty$

1.

$$\sum_{t=1}^{n} \mathsf{P}\{|x_i| > b_n\} \to 0 \tag{8.38}$$

2.

$$b_n^{-1} \sum_{t=1}^{n} \mathsf{E}\{x_{ni} \mid \mathcal{F}_{n-1}\} \overset{\mathsf{P}}{\to} 0 \tag{8.39}$$

3.

$$b_n^{-2} \sum_{t=1}^{n} \left[\mathsf{E}\left\{ x_{ni}^2 \right\} - \mathsf{E}\left\{ (\mathsf{E}\{x_{ni} \mid \mathcal{F}_{n-1}\})^2 \right\} \right] \to 0 \tag{8.40}$$

Proof. Define

$$S_{nn} := \sum_{i=1}^{n} x_{ni}$$

On account of (8.38)

$$\mathsf{P}\{S_n/b_n \neq S_{nn}/b_n\} \le \sum_{i=1}^{n} \mathsf{P}\{x_i \neq x_{ni}\} = \sum_{t=1}^{n} \mathsf{P}\{|x_i| > b_n\} \to 0$$

So, it is sufficient to prove

$$b_n^{-1} S_{nn} \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0$$

But by (8.39) and the identity

$$b_n^{-1} S_{nn} = b_n^{-1} \sum_{i=1}^{n} \left[x_{ni} - \mathsf{E}\{x_{ni} \mid \mathcal{F}_{n-1}\} \right] + b_n^{-1} \sum_{i=1}^{n} \mathsf{E}\{x_{ni} \mid \mathcal{F}_{n-1}\}$$

it is sufficient to prove

$$b_n^{-1} \sum_{i=1}^{n} \left[x_{ni} - \mathsf{E}\{x_{ni} \mid \mathcal{F}_{n-1}\} \right] \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0$$

By the second Chebyshev inequality (4.10) and in view of the condition (8.40) for any $\varepsilon > 0$ it follows that

$$\begin{aligned}
&\mathsf{P}\left\{\left|b_n^{-1}\sum_{i=1}^{n}\left[x_{ni}-\mathsf{E}\{x_{ni}\mid\mathcal{F}_{n-1}\}\right]\right|>\varepsilon\right\}\\
&\le\frac{1}{\varepsilon^2}\mathsf{E}\left\{\left(b_n^{-1}\sum_{i=1}^{n}\left[x_{ni}-\mathsf{E}\{x_{ni}\mid\mathcal{F}_{n-1}\}\right]\right)^2\right\}\\
&=\frac{1}{\varepsilon^2}b_n^{-2}\sum_{t=1}^{n}\left[\mathsf{E}\left\{x_{ni}^2\right\}-\mathsf{E}\left\{(\mathsf{E}\{x_{ni}\mid\mathcal{F}_{n-1}\})^2\right\}\right]\underset{n\to\infty}{\to}0
\end{aligned}$$

that corresponds to the convergence on probability (6.42). Theorem is proven. □

Remark 8.5. *In the particular case when x_n are independent, the above conditions are also necessary (as well as sufficient) for (8.37) (see Loéve (1977)).*

Remark 8.6. *LLN in the form (8.34) follows from (8.37) if we put $b_n = n$.*

8.2.2 Strong law of large numbers

8.2.2.1 Teöplitz and Kronecker Lemmas as auxiliary instruments

As a prelude of the results of this section let us remember the lemmas of Teöplitz and Kronecker which will be often used below.

Lemma 8.2. (Teöplitz) *Let $\{a_n\}$ $(n = 1, 2, \ldots)$ be a sequence of nonnegative real numbers such that*

$$\boxed{b_n := \sum_{t=1}^{n} a_t \to \infty \quad \text{when} \quad n \to \infty} \tag{8.41}$$

and $\{x_n\}$ $(n = 1, 2, \ldots)$ be a sequence of real numbers which converges to x^*, that is,

$$\boxed{x_n \underset{n\to\infty}{\to} x^*} \tag{8.42}$$

Then

(a) *there exists an integer n_0 such that $b_n > 0$ for all $n \ge n_0$;*

(b)

$$\boxed{\frac{1}{b_n}\sum_{t=1}^{n} a_t x_t \to x^* \quad \text{when} \quad n_0 \le n \to \infty} \tag{8.43}$$

Proof. The claim (a) results from (8.41). To prove (b) let us select $\varepsilon > 0$ and $n_0'(\varepsilon) \ge n_0$ such that for all $n \ge n_0'(\varepsilon)$ we have (in view of (8.42)) $|x_n - x^*| \le \varepsilon$. Then it follows that

$$\left|\frac{1}{b_n}\sum_{t=1}^{n} a_t x_t - x^*\right| = \left|\frac{1}{b_n}\sum_{t=1}^{n} a_t \left(x_t - x^*\right)\right| \le \frac{1}{b_n}\sum_{t=1}^{n} a_t \left|x_t - x^*\right|$$
$$= \frac{1}{b_n}\sum_{t=1}^{n_0'(\varepsilon)-1} a_t \left|x_t - x^*\right| + \frac{1}{b_n}\sum_{n_0'(\varepsilon)}^{n} a_t \left|x_t - x^*\right|$$
$$\le \frac{1}{b_n}\sum_{t=1}^{n_0'(\varepsilon)-1} a_t \left|x_t - x^*\right| + \frac{\varepsilon}{b_n}\sum_{n_0'(\varepsilon)}^{n} a_t \le \frac{\text{const}}{b_n} + \varepsilon \to \varepsilon \quad \text{when } b_n \to \infty$$

Since this is true for any $\varepsilon > 0$ we obtain the proof of the lemma. □

Corollary 8.4. *If* $x_n \underset{n\to\infty}{\to} x^*$ *then*

$$\boxed{\frac{1}{n}\sum_{t=1}^{n} x_t \underset{n\to\infty}{\to} x^*} \tag{8.44}$$

Proof. To prove (8.44) it is sufficient in (8.43) to take $a_n = 1$ for all $n = 1, 2, \ldots$. □

Corollary 8.5. *Let* $\{a_n\}$ $(n = 1, 2, \ldots)$ *be a sequence of nonnegative real numbers such that*

$$\boxed{\sum_{t=1}^{n} a_t \to \infty \quad \textit{when} \quad n \to \infty} \tag{8.45}$$

and for some numerical non-zero sequence $\{b_n\}$ *of real numbers there exists the limit*

$$\boxed{\lim_{n\to\infty} b_n^{-1}\sum_{t=1}^{n} a_t = \alpha} \tag{8.46}$$

Let also $\{x_n\}$ $(n = 1, 2, \ldots)$ *be a sequence of real numbers which converges to* x^**, that is,*

$$\boxed{x_n \underset{n\to\infty}{\to} x^*} \tag{8.47}$$

Then

$$\boxed{\lim_{n\to\infty} b_n^{-1}\sum_{t=1}^{n} a_t x_n = \alpha x^*} \tag{8.48}$$

Proof. Directly applying the Teöplitz Lemma 8.2 we derive

$$b_n^{-1}\sum_{t=1}^{n} a_t x_n = \left[b_n^{-1}\sum_{t=1}^{n} a_t\right]\left[\left(\sum_{t=1}^{n} a_t\right)^{-1}\sum_{t=1}^{n} a_t x_n\right] \to \alpha x^*$$ □

Lemma 8.3. (Kronecker) *Let* $\{a_n\}\,(n = 1, 2, \ldots)$ *be a sequence of nonnegative nondecreasing real numbers such that*

$$\boxed{0 \leq b_n \leq b_n \to \infty \quad when \quad n \to \infty} \tag{8.49}$$

and $\{x_n\}$ $(n = 1, 2, \ldots)$ *be a sequence of real numbers such that the series* $\sum_{t=1}^{n} x_t$ *converges, that is,*

$$\boxed{s_n := \sum_{t=n_0}^{n} x_t \underset{n\to\infty}{\to} s^*, \left|s^*\right| < \infty} \tag{8.50}$$

Then

(a) there exists an integer n_0 *such that* $b_n > 0$ *for all* $n \geq n_0$*;*
(b)

$$\boxed{\frac{1}{b_n}\sum_{t=1}^{n} b_t x_t \to 0 \quad when \quad n_0 \leq n \to \infty} \tag{8.51}$$

Proof. Applying Abel's identity (see Lemma 12.2 in Poznyak (2008)) for the scalar case, namely, using the identity

$$\sum_{t=n_0}^{n} \alpha_t \beta_t = \alpha_n \sum_{t=n_0}^{n} \beta_t - \sum_{t=n_0}^{n} (\alpha_t - \alpha_{t-1}) \sum_{s=n_0}^{t-1} \beta_s$$

we derive

$$\begin{aligned}\frac{1}{b_n}\sum_{t=n_0}^{n} b_t x_t &= \frac{1}{b_n}\left[b_n \sum_{t=n_0}^{n} x_t - \sum_{t=n_0}^{n} (b_t - b_{t-1}) \sum_{s=n_0}^{t-1} x_s\right] \\ &= s_n - \frac{1}{b_n}\sum_{t=n_0}^{n} (b_t - b_{t-1})\, s_{t-1}\end{aligned}$$

Denote $a_t := b_t - b_{t-1}$. Then

$$b_n = \sum_{t=n_0}^{n} a_t + b_{n_0} = \sum_{t=n_0}^{n} a_t \left[1 + b_{n_0} \Big/ \sum_{t=n_0}^{n} a_t\right]$$

and hence, by the Teöplitz Lemma 8.2, we have

$$\begin{aligned}\frac{1}{b_n}\sum_{t=n_0}^{n} b_t x_t &= s_n - \left[1 + b_{n_0} \Big/ \sum_{t=n_0}^{n} a_t\right]^{-1} \left(\sum_{t=n_0}^{n} a_t\right)^{-1} \sum_{t=n_0}^{n} a_t s_{t-1} \\ &\to s^* - s^* = 0\end{aligned}$$

that proves (8.51). □

8.2.2.2 Matrix versions of Teöplitz and Kronecker lemmas

Lemma 8.4. (MT-lemma, Poznyak and Tchikin (1985)) *Let $\{A_{s,n}\}_{1\le s\le n}$ be a family of triangular matrices $A_{s,n} \in \mathbb{R}^{N\times M}$ such that for some matrix norm $\|\cdot\|$*

1.

$$\boxed{\sup_n \left\|A_{s,n} - A\right\| < \infty, \quad 1 \le s \le n} \tag{8.52}$$

2.

$$\boxed{\left\|A_{s,n} - A\right\| \to 0 \quad whenever \quad n \ge s \to \infty} \tag{8.53}$$

and $\{B_n\}_{n\ge1}$ be a sequence of nonsingular matrices $B_n \in \mathbb{R}^{N\times N}$ satisfying

3.

$$\boxed{\|B_n\| \underset{n\to\infty}{\to} 0} \tag{8.54}$$

4.

$$\boxed{\limsup_{n\to\infty} \sum_{s=1}^{n} \left\|B_n\left(B_s^{-1} - B_{s-1}^{-1}\right)\right\| < \infty} \tag{8.55}$$

Then

$$\boxed{\left\|B_n \sum_{s=1}^{n}\left(B_s^{-1} - B_{s-1}^{-1}\right) A_{s,n} - A\right\| \underset{n\to\infty}{\to} 0} \tag{8.56}$$

Proof. By the assumption (8.53) it follows that for any $\varepsilon > 0$ there exists an integer $s(\varepsilon)$ such that for all $n \ge s \ge s(\varepsilon)$ we have $\left\|A_{s,n} - A\right\| < \varepsilon$. Then by the assumptions of the lemma

$$\begin{aligned}
&\limsup_{n\to\infty}\left\|B_n \sum_{t=1}^{n}\left(B_s^{-1} - B_{s-1}^{-1}\right) A_{s,n} - A\right\| \\
&\quad\le \limsup_{n\to\infty}\left\|B_n \sum_{t=s(\varepsilon)+1}^{n}\left(B_s^{-1} - B_{s-1}^{-1}\right)\left(A_{s,n} - A\right)\right\| \\
&\qquad+ \limsup_{n\to\infty}\left\|B_n \sum_{t=1}^{s(\varepsilon)}\left(B_s^{-1} - B_{s-1}^{-1}\right)\left(A_{s,n} - A\right)\right\|
\end{aligned}$$

$$= \limsup_{n\to\infty} \left\| B_n \sum_{t=s(\varepsilon)+1}^{n} \left(B_s^{-1} - B_{s-1}^{-1} \right) \left(A_{s,n} - A \right) \right\|$$

$$\leq \varepsilon \limsup_{n\to\infty} \sum_{s=1}^{n} \left\| B_n \left(B_s^{-1} - B_{s-1}^{-1} \right) \right\| = O(\varepsilon)$$

which, in view of an arbitrary small ε, completes the proof. □

Corollary 8.6. **(Poznyak and Tchikin, 1985)**

1. For a sequence $\{R_n\}_{n\geq 1}$ of squared matrices $R_n \in \mathbb{R}^{N\times N}$ there exists the limit

$$\boxed{\lim_{n\to\infty} \sum_{t=1}^{n} R_t := R} \tag{8.57}$$

2. The sequence $\{B_n\}_{n\geq 1}$ of nonsingular matrices $B_n \in \mathbb{R}^{N\times N}$ satisfies the conditions (8.54) and (8.55).
Then

$$\boxed{B_n \sum_{t=1}^{n} B_t^{-1} R_{n-t} \underset{n\to\infty}{\to} R} \tag{8.58}$$

Proof. Define $S_{k,n} := \sum_{t=n-k}^{n-1} R_t$ and $S_{0,n} := 0$, $B_0^{-1} := 0$. Then, by the Abel's identity (see Lemma 12.2 in Poznyak (2008))

$$\sum_{t=1}^{n} B_t^{-1} R_{n-t} = B_n^{-1} \sum_{t=1}^{n} R_{n-t} - \sum_{t=1}^{n} \left(B_t^{-1} - B_{t-1}^{-1} \right) \sum_{s=1}^{t-1} R_{n-t}$$

$$= B_t^{-1} S_{n,n} - \sum_{t=1}^{n} \left(B_t^{-1} - B_{t-1}^{-1} \right) S_{t-1,n}$$

$$= \sum_{t=1}^{n} \left(B_t^{-1} - B_{t-1}^{-1} \right) \left(S_{n,n} - S_{t-1,n} \right) = \sum_{t=1}^{n} \left(B_t^{-1} - B_{t-1}^{-1} \right) \left(R - T_{n-t} \right)$$

where

$$T_m := \sum_{s=m+1}^{\infty} R_s \underset{m\to\infty}{\to} 0 \quad (m := 0, 1, \ldots)$$

But

$$\limsup_{n\to\infty} \left\| B_n \sum_{t=1}^{n} B_t^{-1} R_{n-t} - R \right\|$$

$$= \limsup_{n\to\infty} \left\| B_n \sum_{t=1}^{n} \left(B_t^{-1} - B_{t-1}^{-1} \right) T_{n-t} \right\| = 0$$

which follows from the lemma above if taken within $A := 0$ and $A_{s,n} := T_{n-s}$. Corollary is proven. □

Lemma 8.5. **(MK-lemma, Poznyak and Tchikin (1985))** *Let $\{B_n\}_{n\geq 1}$ be a sequence of nonsingular matrices $B_n \in \mathbb{R}^{N\times N}$ satisfying (8.54) and (8.55). Then*

(a) for any sequence $\{A_n\}_{n\geq 1}$ of matrix $A_n \in \mathbb{R}^{N\times M}$ such that the limit of $C_n := \sum_{s=1}^{n} A_s$ exists when $n \to \infty$ the following property holds:

$$\boxed{B_n \sum_{s=1}^{n} B_s^{-1} A_s \underset{n\to\infty}{\to} 0} \tag{8.59}$$

(b) for any family $\left\{A_{s,n}\right\}_{1\leq s\leq n}$ of triangular matrices $A_{s,n} \in \mathbb{R}^{N\times M}$ such that

$$\sum_{n=s+1}^{\infty} \left\| A_{s,n} \right\| \underset{n\to\infty}{\to} 0 \tag{8.60}$$

it follows that

$$\boxed{B_n \sum_{s=t+1}^{n} B_s^{-1} A_{t,s} \underset{n>t\to\infty}{\to} 0} \tag{8.61}$$

Proof. (a) By the matrix Abel identity (see Lemma 12.2 in Poznyak (2008)) it follows that

$$B_n \sum_{s=1}^{n} B_s^{-1} A_s = C_n - B_n \sum_{s=1}^{n} \left(B_s^{-1} - B_{s-1}^{-1} \right) C_{s-1}$$

where $C_0 := 0$ and B_0^{-1}. Since $\lim_{n\to\infty} C_n := C_\infty$ exists then all conditions of Lemma 8.4 are fulfilled and therefore by this lemma

$$B_n \sum_{s=1}^{n} \left(B_s^{-1} - B_{s-1}^{-1} \right) C_{s-1} \underset{n\to\infty}{\to} C_\infty$$

that implies (8.59).

(b) Analogously,

$$B_n \sum_{s=t+1}^{n} B_s^{-1} A_{t,s} = C_{t,n} - B_n \sum_{s=t+1}^{n} \left(B_s^{-1} - B_{s-1}^{-1} \right) C_{t,s-1}$$

$$C_{t,n} := \sum_{s=t+1}^{n} A_{t,s} \underset{n>t\to\infty}{\to} 0$$

Hence, for any $\varepsilon > 0$ there exists a number $t_0(\varepsilon)$ such that for all $n > t \geq t_0(\varepsilon)$

$$\left\| C_{t,n} \right\| \leq \varepsilon$$

and then

$$\left\| B_n \sum_{s=t+1}^{n} B_s^{-1} A_{t,s} \right\| \leq \left\| C_{t,n} \right\| + \sum_{s=t+1}^{n} \left\| B_n \left(B_s^{-1} - B_{s-1}^{-1} \right) \right\| \left\| C_{t,s-1} \right\| \leq \varepsilon + \varepsilon \mathsf{Const}$$

that proves the desired result. □

8.2.2.3 SLLN via a series convergence

Lemma 8.6. *To guarantee the SLLN fulfilling, namely,*

$$\boxed{n^{-1} S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{8.62}$$

it is sufficient that the series

$$\boxed{\tilde{S}_n := \sum_{t=1}^{n} \frac{x_t}{t}} \tag{8.63}$$

converges (a.s.).

Proof. It follows directly from the Kronecker Lemma if we take into account the trivial identity

$$n^{-1} S_n = n^{-1} \sum_{t=1}^{n} t \left(\frac{x_t}{t} \right)$$

that proves the lemma. □

Now we embark on SLLN discussion for martingales.

Lemma 8.7. *Let $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ be a martingale-difference, that is,*

$$\mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\} \overset{a.s}{=} 0$$

with $\mathsf{E}\{\sup_n n^{-1} |x_n|\} < \infty$. Then for $\tilde{S}_n := \sum_{t=1}^{n} t^{-1} x_t$

$$\limsup_{n\to\infty} \tilde{S}_n \overset{a.s.}{=} +\infty$$

$$\liminf_{n\to\infty} \tilde{S}_n \overset{a.s.}{=} -\infty$$

on the set $\{\sum_{n=1}^{\infty} n^{-1}x_n$ *diverges*$\}$ *and the limit* $\lim_{n\to\infty} \tilde{S}_n$ *exists and (a.s.) finite on the set* $\{\sum_{n=1}^{\infty} n^{-1}x_n$ *converges*$\}$.

Proof. Let $\tau_a := \min\left\{n : \tilde{S}_n > a\right\}$ with $\tau_a = \infty$ if no such n exists. Random variables $\tau_{a\wedge n}$ $(n \geq 1)$ form a non-decreasing sequence of stopping times and so $\left\{\tilde{S}_{\tau_{a\wedge n}}, \mathcal{F}_n\right\}_{n\geq 1}$ is a martingale. Also we have

$$\tilde{S}^+_{\tau_{a\wedge n}} := \max\left\{0; \tilde{S}_{\tau_{a\wedge n}}\right\} \leq \tilde{S}^+_{\tau_{a\wedge(n-1)}} + n^{-1}x^+_{\tau_{a\wedge(n-1)}} \leq a + \sup_n n^{-1}x^+_{\tau_{a\wedge(n-1)}}$$

and since $\mathsf{E}\{\sup_n n^{-1}|x_n|\} < \infty$ it follows that

$$\mathsf{E}\left\{\left|\tilde{S}_{\tau_{a\wedge n}}\right|\right\} \leq 2\mathsf{E}\left\{\tilde{S}^+_{\tau_{a\wedge n}}\right\}$$

is bounded as $n \to \infty$. Therefore, by the supermartingale convergence Theorem 7.10 $\left\{\tilde{S}_{\tau_{a\wedge n}}\right\}$ converges (a.s.) to a finite limit as $n \to \infty$. □

8.2.2.4 SLLN for mixingale sequences

First, let us prove an auxiliary result which will be used hereafter.

Theorem 8.4. *If* $\{x_n, \mathcal{F}_n\}$ *is a mixingale such that*

$$\boxed{\sum_{n=1}^{\infty} c_n^2 < \infty \quad \text{and} \quad \psi_n = O\left(\frac{1}{\sqrt{n}\log^2 n}\right) \quad \text{as } n \to \infty} \tag{8.64}$$

then the series $S_n = \sum_{t=1}^{n} x_t$ *converges a.s. to a finite limit as* $n \to \infty$.

Proof. **(McLeish, 1975)** By the second Chebyshev inequality (4.10) and by Theorem 8.2 we have that for each $m' > m$

$$P\left\{\max_{m<n\leq m'} |S_n - S_m| > \varepsilon\right\} \leq \varepsilon^{-2}K\sum_{t=m}^{m'} c_t^2$$

so that, taking $m' \to \infty$,

$$P\left\{\max_{m<n} |S_n - S_m| > \varepsilon\right\} \leq \varepsilon^{-2}K\sum_{t=m}^{\infty} c_t^2 \underset{m\to\infty}{\to} 0$$

Therefore, by the condition (6.58), S_n converges (a.s.) to a limit random variable with finite variance. Theorem is proven. □

Now we are ready to formulate the main result concerning SLLN for mixingales.

Theorem 8.5. *Let* $\{x_n, \mathcal{F}_n\}$ *be a mixingale (see Definition 8.5) and* $\{b_n\}$ *be a sequence of positive constants increasing to* ∞*, i.e.,* $0 < b_n \underset{n\to\infty}{\uparrow} \infty$ *such that*

$$\boxed{\sum_{n=1}^{\infty} \frac{c_n^2}{b_n^2} < \infty \quad \textit{and} \quad \psi_n = O\left(\frac{1}{\sqrt{n}\log^2 n}\right) \quad \textit{as } n \to \infty} \tag{8.65}$$

Then

$$\boxed{\frac{1}{b_n} S_n \overset{a.s.}{\underset{n\to\infty}{\to}} 0} \tag{8.66}$$

Proof. By Lemma 8.6 the property (8.66) will be guaranteed if the series $\sum_{t=1}^{n} \frac{x_t}{b_t}$ converges (a.s.). But if $\{x_n, \mathcal{F}_n\}$ is a mixingale, then $\{\tilde{x}_n, \mathcal{F}_n\}$, where $\tilde{x}_n := x_t/b_t$, is also mixingale with

$$\tilde{\psi}_n = \psi_n \quad \text{and} \quad \tilde{c}_n = c_n/b_t$$

So, the desired result (8.66) follows directly from Theorem 8.4. □

8.2.2.5 SLLN for martingale-differences

Theorem 8.6. *Let* $\{x_n, \mathcal{F}_n\}_{n\geq 1}$ *be a martingale-difference, that is,*

$$\boxed{\mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\} \overset{a.s}{=} 0}$$

with $\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\}$ *satisfying*

$$\boxed{\sum_{n=1}^{\infty} b_n^{-2} \mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \overset{a.s.}{<} \infty}$$

for some sequence $\{b_n\}_{n\geq 1}$ *of positive constants monotonically increasing to* ∞*, i.e.,* $0 < b_n \underset{n\to\infty}{\uparrow} \infty$*. Then*

$$\boxed{b_n^{-1} S_n \overset{a.s.}{\underset{n\to\infty}{\to}} 0}$$

Proof. Since

$$s_n := b_n^{-1} S_n = b_n^{-1}\left(b_{n-1} b_{n-1}^{-1} S_{n-1} + x_n\right) = b_n^{-1} b_{n-1} s_{n-1} + b_n^{-1} x_n$$

then the following relation holds:

$$s_n^2 = \left(b_n^{-1}b_{n-1}\right)^2 v_{n-1} + 2b_n^{-2}b_{n-1}s_{n-1}x_n + b_n^{-2}x_n^2$$

and, by the martingale-difference property,

$$\begin{aligned}
\mathsf{E}\left\{s_n^2 \mid \mathcal{F}_{n-1}\right\} &\overset{a.s.}{=} \left(b_n^{-1}b_{n-1}\right)^2 s_{n-1}^2 + b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \\
&= \left(1 - \frac{b_n - b_{n-1}}{b_n}\right)^2 s_{n-1}^2 + b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \\
&= \left(1 - 2\frac{b_n - b_{n-1}}{b_n}\left[1 - \frac{b_n - b_{n-1}}{2b_n}\right]\right) s_{n-1}^2 + b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \\
&= \left(1 - 2\frac{b_n - b_{n-1}}{b_n}\left[\frac{1}{2} + \frac{b_{n-1}}{2b_n}\right]\right) s_{n-1}^2 + b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \\
&\leq \left(1 - \frac{b_n - b_{n-1}}{b_n}\right) s_{n-1}^2 + b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\}
\end{aligned}$$

Taking in Lemma 7.9

$$\eta_n := s_n^2, \quad \nu_n := 0, \quad \lambda_n := \frac{b_n - b_{n-1}}{b_n}, \quad \theta_n := b_n^{-2}\mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\}$$

we obtain $\eta_n = s_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ that proves the desired result. □

Corollary 8.7. (The Kolmogorov sufficient condition) *If $\{x_n\}$ is a sequence of independent random variables with zero-mean and finite variances $\sigma_n^2 < \infty$ $(n = 1, 2, ..)$ satisfying the condition*

$$\boxed{\sum_{n=1}^{\infty} n^{-2}\sigma_n^2 < \infty} \tag{8.67}$$

then

$$\boxed{n^{-1}S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{8.68}$$

Proof. Letting in Theorem 8.6 $b_n = n$ and taking into account that in this case $\mathsf{E}\{x_n^2 \mid \mathcal{F}_{n-1}\} = \sigma_n^2$ we obtain (8.68). □

Theorem 8.7. *Let $r > 1$ and $\{x_n\}$ be a sequence of independent identically distributed absolutely integrable random variables with zero-mean, namely,*

$$\boxed{\mathsf{E}\{x_n\} = 0, \quad m_1 := \mathsf{E}\{|x_n|\} < \infty} \tag{8.69}$$

Then

$$\boxed{n^{-r}S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{8.70}$$

Proof. By the Kronecker lemma 8.3 it is sufficient to prove the convergence (a.s.) of the series $\sum_{t=1}^{n} t^{-r} x_t$ as $n \to \infty$. But it converges if it converges absolutely, i.e., if the series $\sum_{t=1}^{n} t^{-r} |x_t|$ converges (a.s.) as $n \to \infty$. But the series of nonnegative random variables converges almost surely if the series of its mathematical expectations converges, that is, when $\sum_{t=1}^{\infty} t^{-r} \mathsf{E}\{|x_t|\} < \infty$. But

$$\sum_{t=1}^{\infty} t^{-r} \mathsf{E}\{|x_t|\} = \sum_{t=1}^{\infty} t^{-r} m_1 = m_1 \sum_{t=1}^{\infty} t^{-r} < \infty$$

if $r > 1$, which proves the theorem. □

Theorem given below deals with a more delicate case when we are interested in the statement $n^{-\alpha} S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ with $\alpha \in (0, 1)$.

Theorem 8.8. **(Marcinkiewicz and Zygmund, 1937)** *Let $r > 0$ and $\{x_n\}_{n\geq 1}$ be a sequence of independent identically distributed random variables with zero-mean and finite absolute r-moment, namely,*

$$\boxed{\mathsf{E}\{x_n\} = 0, \quad m_r := \left(\mathsf{E}\left\{|x_n|^r\right\}\right)^{1/r} < \infty} \tag{8.71}$$

Then

$$\boxed{n^{-1/r} S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{8.72}$$

Proof. First represent S_n as the sum of two terms:

$$S_n = S_n' + S_n''$$

where

$$S_n' = \sum_{t=1}^{n} \left[x_t - x_t \chi\left(|x_t| \leq t^{1/r+\varepsilon}\right)\right], \quad \varepsilon > 0$$

$$S_n'' = \sum_{t=1}^{n} x_t \chi\left(|x_t| \leq t^{1/r+\varepsilon}\right)$$

So, by this presentation it is sufficient to prove that $n^{-1/r} S_n' \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ and $n^{-1/r} S_n'' \underset{n\to\infty}{\overset{a.s.}{\to}} 0$.

(a) Let us prove that $n^{-1/r} S_n' \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. One has

$$\begin{aligned} v_n' &:= n^{-1/r} \left|S_n'\right| \leq n^{-1/r} \left[(n-1)^{1/r} v_{n-1} + |x_n| \chi\left(|x_n| > n^{1/r+\varepsilon}\right)\right] \\ &= (1 - 1/n)^{1/r} v_{n-1}' + n^{-1/r} |x_n| \chi\left(|x_n| > n^{1/r+\varepsilon}\right) \end{aligned}$$

By the Hölder inequality (13.73) for $p = r$ and $q = r/(r-1)$ and by the Markov inequality (4.8) it follows that

$$\begin{aligned} \mathsf{E}\left\{|x_n|\,\chi\left(|x_n| > n^{1/r+\varepsilon}\right)\right\} &\le \left(\mathsf{E}\left\{|x_n|^r\right\}\right)^{1/r}\left(\mathsf{E}\left\{\chi^q\left(|x_n| > n^{1/r+\varepsilon}\right)\right\}\right)^{1/q} \\ &= m_r \mathsf{P}^{1/q}\left\{|x_n| > n^{1/r}\right\} \le m_r\left(\frac{\mathsf{E}\left\{|x_n|^r\right\}}{\left(n^{1/r+\varepsilon}\right)^r}\right)^{1/q} = m_r\left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{1/q} \end{aligned} \tag{8.73}$$

and therefore

$$\begin{aligned} \mathsf{E}\left\{v_n'\right\} &\le (1-1/n)^{1/r}\,\mathsf{E}\left\{v_{n-1}'\right\} + n^{-1/r} m_r\left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{1/q} \\ &\le \left(1 - \frac{\left[r^{-1} + o(1)\right]}{n}\right)\mathsf{E}\left\{v_{n-1}\right\} + \frac{m_r^{1+r/q}}{n^{1/r+1/q+\varepsilon r/q}} \\ &= \left(1 - \frac{\left[r^{-1} + o(1)\right]}{n}\right)\mathsf{E}\left\{v_{n-1}'\right\} + \frac{m_r^r}{n^{1+\varepsilon r/q}} \end{aligned}$$

Hence, by Lemma 16.14 in Poznyak (2008), it follows that $\mathsf{E}\left\{v_n'\right\} \to 0$ as $n \to \infty$, but for nonnegative random variables v_n' this means that $v_n' \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. So, we have proved the asymptotic closeness of the arithmetic average of a sequence of random variables and its truncated analog for any $r > 0$.

(b) Let us show that $n^{-1/r} S_n'' \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. First, notice that S_n'' can be represented as

$$S_n'' = \sum_{t=1}^{n}\left(y_t - \mathsf{E}\left\{y_t\right\}\right) + \sum_{t=1}^{n}\mathsf{E}\left\{y_t\right\}, \quad y_t := x_t\,\chi\left(|x_t| \le n^{1/r+\varepsilon}\right)$$

where, by the Töeplitz Lemma 8.2,

$$n^{-1/r}\sum_{t=1}^{n}\mathsf{E}\left\{y_t\right\} = n^{-1/r}\sum_{t=1}^{n}\left[n^{1/r} - (n-1)^{1/r}\right]\frac{\mathsf{E}\left\{y_t\right\}}{n^{1/r} - (n-1)^{1/r}} \underset{n\to\infty}{\to} 0$$

since for $r > 1$ it follows that

$$\left|\frac{\mathsf{E}\left\{y_t\right\}}{n^{1/r} - (n-1)^{1/r}}\right| = \frac{\left|\mathsf{E}\left\{y_t\right\}\right|}{n^{1/r}\left(1 - (1-1/n)\right)^{1/r}} = \frac{\left|\mathsf{E}\left\{y_t\right\}\right|}{\left[r^{-1} + o(1)\right] n^{1/r-1}} \underset{n\to\infty}{\to} 0$$

we take into account (8.73):

$$\left|\mathsf{E}\left\{y_t\right\}\right| \le \mathsf{E}\left\{|y_t|\right\} < m_r\left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{1/q}$$

But $v_n'' := n^{-1/r}\sum_{t=1}^{n}\left(y_t - \mathsf{E}\left\{y_t\right\}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ if, by the Kronecker lemma 8.3, the series $\sum_{t=1}^{n} t^{-1/r}\left(y_t - \mathsf{E}\left\{y_t\right\}\right)$ converges that is true when

$$\sum_{n=1}^{\infty} n^{-2/r}\,\mathsf{E}\left\{\left(y_n - \mathsf{E}\left\{y_n\right\}\right)^2\right\} < \infty$$

Estimating the terms of this series we get

$$\left|\mathsf{E}\left\{(y_n - \mathsf{E}\{y_n\})^2\right\}\right| = \left|\mathsf{E}\left\{y_n^2\right\} - \mathsf{E}^2\{y_n\}\right| \le \mathsf{E}\left\{y_n^2\right\} + \mathsf{E}^2\{y_n\}$$

By (8.73),

$$\mathsf{E}^2\{y_n\} \le \mathsf{E}^2\{|y_n|\} < m_r^2 \left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{2/q} \underset{n\to\infty}{\to} 0$$

and

$$\begin{aligned}
\mathsf{E}\left\{y_n^2\right\} &= \mathsf{E}\left\{x_n^2 \chi\left(|x_n| > n^{1/r+\varepsilon}\right)\right\} \le \left(\mathsf{E}\left\{|x_n|^{2(r/2)}\right\}\right)^{2/r} \\
&\quad \cdot \left(\mathsf{E}\left\{\chi^q\left(|x_n| > n^{1/r+\varepsilon}\right)\right\}\right)^{1/q} = m_r^2 \mathsf{P}^{1/q}\left\{|x_n| > n^{1/r}\right\} \\
&\le m_r^2 \left(\frac{\mathsf{E}\left\{|x_n|^r\right\}}{\left(n^{1/r+\varepsilon}\right)^r}\right)^{1/q} = m_r^2 \left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{1/q} = \frac{m_r^{2+r/q}}{n^{1/q+\varepsilon r/q}}
\end{aligned}$$

So, each term in the series $\sum_{n=1}^{\infty} n^{-2/r}\mathsf{E}\left\{(y_n - \mathsf{E}\{y_n\})^2\right\}$ is estimated as

$$\begin{aligned}
\frac{1}{n^{2/r}}\left[\frac{m_r^{2+r/q}}{n^{1/q+\varepsilon r/q}} + m_r^2\left(\frac{m_r^r}{n^{1+\varepsilon r}}\right)^{2/q}\right] &= \frac{\text{const}}{n^{2/r+1/q+\varepsilon r/q}} + \frac{\text{const}}{n^{2/r+2/q+2\varepsilon r/q}} \\
&= \frac{\text{const}\,(1+o(1))}{n^{2/r+1/q+\varepsilon r/q}} = \frac{\text{const}\,(1+o(1))}{n^{1/r+1+\varepsilon r/q}}
\end{aligned}$$

Therefore this series converges for any $r > 0$ that completes the proof. □

8.2.2.6 SLLN for mixing sequences

Theorem 8.9. *Let $\{x_n\}_{n\ge1}$ be an ϕ-mixing sequence (8.10) such that $\mathrm{E}\{x_n\} = 0$ and*

$$\boxed{\sum_{s=1}^{\infty} b_s^{-2}\mathsf{E}\left\{x_s^2\right\} < \infty} \tag{8.74}$$

and

$$\boxed{\sum_{n=1}^{\infty}\sqrt{\mathsf{E}\left\{x_n^2\right\}}b_n^{-1}\sum_{t=1}^{n-1} b_t^{-1}\sqrt{\phi_{n-t}}\sqrt{\mathsf{E}\left\{x_t^2\right\}} < \infty} \tag{8.75}$$

Then

$$\boxed{b_n^{-1}S_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{8.76}$$

Proof. By the Kronecker lemma 8.3 the property (8.76) holds if the series $s_n := \sum_{t=1}^{n} \frac{1}{b_t} x_t$ converges (a.s.) as $n \to \infty$. Then

$$s_n^2 := \left(s_{n-1} + b_n^{-1} |x_n|\right)^2 = s_{n-1}^2 + 2b_n^{-1} s_{n-1} |x_n| + b_n^{-2} x_n^2$$
$$= s_{n-1}^2 + 2b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} x_t x_n + b_n^{-2} x_n^2$$

that (for $\mathcal{F}_n := \sigma(x_1, \ldots, x_n)$) implies

$$\mathsf{E}\left\{s_n^2 \mid \mathcal{F}_{n-1}\right\} = s_{n-1}^2 + 2b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \mathsf{E}\{x_t x_n \mid \mathcal{F}_{n-1}\} + b_n^{-2} \mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\}$$

By the Robbins–Siegmund Theorem 7.11 s_n^2 converges (a.s.) if

$$\sum_{n=1}^{\infty} b_n^{-2} \mathsf{E}\left\{x_n^2 \mid \mathcal{F}_{n-1}\right\} \overset{a.s.}{<} \infty, \quad \sum_{n=1}^{\infty} b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \left|\mathsf{E}\{x_t x_n \mid \mathcal{F}_{n-1}\}\right| \overset{a.s.}{<} \infty$$

that is true if

$$\sum_{n=1}^{\infty} b_n^{-2} \mathsf{E}\left\{x_n^2\right\} < \infty, \sum_{n=1}^{\infty} b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \mathsf{E}\{|x_t \mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\}|\} < \infty$$

The first series converges by the assumption (8.74). To prove the convergence of the second series notice that by the Cauchy–Bounyakovski–Schwartz inequality (4.16) and in view of the mixing property (8.26) we have

$$|\mathsf{E}\{x_t \mathsf{E}\{x_n \mid \mathcal{F}_t\}\}| \le \sqrt{\mathsf{E}\left\{x_t^2\right\}} \sqrt{\mathsf{E}\left\{|\mathsf{E}\{x_n \mid \mathcal{F}_t\}|^2\right\}}$$
$$\le \sqrt{\mathsf{E}\left\{x_t^2\right\}} \sqrt{4\phi_{n-t} \mathsf{E}\left\{x_n^2\right\}} = 2\sqrt{\phi_{n-t}} \sqrt{\mathsf{E}\left\{x_t^2\right\} \mathsf{E}\left\{x_n^2\right\}}$$

that implies

$$b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \mathsf{E}\{|x_t \mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\}|\} \le 2b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \sqrt{\phi_{n-t}} \sqrt{\mathsf{E}\left\{x_t^2\right\} \mathsf{E}\left\{x_n^2\right\}}$$
$$= 2\sqrt{\mathsf{E}\left\{x_n^2\right\}} b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \sqrt{\phi_{n-t}} \sqrt{\mathsf{E}\left\{x_t^2\right\}}$$

and therefore, by the assumption (8.75)

$$\sum_{n=1}^{\infty} b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \mathsf{E}\{|x_t \mathsf{E}\{x_n \mid \mathcal{F}_{n-1}\}|\}$$
$$\le 2 \sum_{n=1}^{\infty} \sqrt{\mathsf{E}\left\{x_n^2\right\}} b_n^{-1} \sum_{t=1}^{n-1} b_t^{-1} \sqrt{\phi_{n-t}} \sqrt{\mathsf{E}\left\{x_t^2\right\}} < \infty$$

that proves the theorem. □

Corollary 8.8. *For $b_n := n$ and for $\mathsf{E}\left\{x_n^2\right\} \leq \sigma^2$ $(n = 1, \ldots)$ the conditions of Theorem 8.9 will be fulfilled if*

$$\boxed{\sum_{n=1}^{\infty} n^{-1} \sum_{t=1}^{n-1} t^{-1} \sqrt{\phi_{n-t}} < \infty} \tag{8.77}$$

which is true, for example, if the sequence $\{\phi_n\}_{n \geq 1}$ of the uniform strong mixing coefficients satisfies

$$\boxed{\sum_{n=1}^{\infty} \sqrt{\phi_n} < \infty} \tag{8.78}$$

Proof. The condition (8.77) directly follows from (8.75) if we substitute there $b_n := n$ and use the estimate $\mathsf{E}\left\{x_n^2\right\} \leq \sigma^2$. To prove that

$$\begin{aligned}
r_n &:= \sum_{t=1}^{n} t^{-1} \sum_{s=1}^{t-1} s^{-1} \sqrt{\phi_{t-s}} \\
&= \sum_{t=1}^{n_0} t^{-1} \sum_{s=1}^{t-1} s^{-1} \sqrt{\phi_{t-s}} + \sum_{t=n_0+1}^{n} t^{-1} \sum_{s=1}^{t-1} s^{-1} \sqrt{\phi_{t-s}} \\
&\leq \text{Const}\,(n_0) + \sum_{t=n_0+1}^{n} t^{-1} \sum_{k=1}^{t-1} (t-k)^{-1} \sqrt{\phi_k} \\
&= \text{Const}\,(n_0) + \sum_{t=n_0+1}^{n} t^{-1} \sum_{k=1}^{n} \chi\,(k \leq t-1)\,(t-k)^{-1} \sqrt{\phi_k} \\
&= \text{Const}\,(n_0) + \sum_{k=1}^{n} \sqrt{\phi_k} \sum_{t=n_0+1}^{n} t^{-1} \chi\,(k \leq t-1)\,(t-k)^{-1} \\
&\leq \text{Const}\,(n_0) + \sum_{k=1}^{n} \sqrt{\phi_k} \beta_k
\end{aligned}$$

where

$$\beta_k := \sum_{t=n_0+1}^{\infty} t^{-1} \chi\,(k \leq t-1)\,(t-k)^{-1} = \sum_{t=(n_0+1)\vee(k+1)}^{\infty} t^{-1}\,(t-k)^{-1}$$

The series in the right-hand side converges for any k and the sequence $\{\beta_k\}_{k \geq 1}$ is monotonically non-increasing. Therefore, by the Abel Test (see Corollary 16.27 in Poznyak (2008)), the result of the theorem follows. □

8.2.2.7 SLLN for correlated sequences

Let all random sequences considered below be defined on the probability space (Ω, F, P). For the given centered quadratic-integrable $\mathbb{R}^n$-valued random process $\{\xi_n\}_{n \geq |}$, given by

$$\xi_n \in \mathbb{R}^n, \quad \mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\left\{\xi_n^\top \xi_n\right\} = \sigma_n^2 < \infty \tag{8.79}$$

introduce the special characteristic, the so-called *'double averaged' correlation function* R_n defined by

$$\boxed{R_n := n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} \rho_{t,s} = n^{-2} \mathsf{E}\left\{S_n^\top S_n\right\}, \quad S_n := \sum_{t=1}^{n} \xi_t} \tag{8.80}$$

where

$$\boxed{\rho_{t,s} := \mathsf{E}\{\xi_t^\top \xi_s\}} \tag{8.81}$$

is the corresponding *correlation function*.

Theorem 8.10. (Poznyak, 1992) [1] *If for the vector process $\{\xi_n\}$ (8.79) the following series converges:*

$$\boxed{\sum_{n=1}^{\infty} \left(\frac{\sigma_n}{n} \sqrt{R_{n-1}} + \frac{1}{n^2} \sigma_n^2 \right) < \infty} \tag{8.82}$$

*then '**the strong law of large numbers**' holds for this process, that is,*

$$\boxed{n^{-1} S_n \overset{a.s.}{\to} 0}$$

Remark 8.7. *If the given process $\{\xi_t\}$ has a bounded variance, that is, $\sigma_n^2 \leq \bar{\sigma}^2 < \infty$ and a 'double averaged' correlation function R_n, decreasing as*

$$R_n = O\left(n^{-\varepsilon}\right) (\varepsilon > 0)$$

then the conditions of this theorem are fulfilled automatically.

Proof. Define $\tilde{S}_n := n^{-1} S_n$ for any $n = 1, 2, \ldots$. Then

$$\left\| \tilde{S}_n \right\|^2 = \left(1 - \frac{1}{n}\right)^2 \left\| \tilde{S}_{n-1} \right\|^2 + v_n \leq \left(1 - \frac{1}{n}\right) \tilde{S}_{n-1}^2 + v_n$$

where

$$v_n := 2\frac{1}{n} \left(1 - \frac{1}{n}\right) \tilde{S}_{n-1}^\top \xi_n + \frac{1}{n^2} \|\xi_n\|^2$$

then the back iterations imply

$$\tilde{S}_n^2 \leq \pi_n^2 \tilde{S}_1 + \pi_n \sum_{t=2}^{n} \pi_t^{-1} v_t$$

[1] The English version of this result can be found in Poznyak (2001).

with

$$\pi_n := \prod_{t=2}^{n} \left(1 - t^{-1}\right)$$

By the Kronecker lemma 8.3, $\tilde{S}_n$ tends to zero if with probability 1 the following sequence

$$r_n^{(1)} := \sum_{t=1}^{n} v_t$$

converges. To fulfill this, it is sufficient to show that under the conditions of this theorem the series

$$r_n^{(2)} := \sum_{t=1}^{n} \frac{1}{t^2} \|\xi_t\|^2, \quad r_n^{(3)} := \sum_{t=1}^{n} \frac{1}{t} |\tilde{S}_{t-1}^{\mathsf{T}} \xi_t|$$

converge with probability 1. But this takes place if

$$\sum_{t=1}^{\infty} \frac{1}{t^2} \sigma_n^2 < \infty, \quad \sum_{t=1}^{\infty} \frac{1}{t} \mathsf{E}\{|\tilde{S}_{t-1}^{\mathsf{T}} \xi_t|\} < \infty$$

By the Cauchy–Bounyakovski–Schwartz inequality (4.16), it follows

$$\sum_{t=1}^{\infty} \frac{1}{t} \mathsf{E}\{|\tilde{S}_{t-1}^{\mathsf{T}} \xi_t|\} \leq \sum_{t=1}^{\infty} \frac{1}{t} \sqrt{\mathsf{E}\{\|\tilde{S}_{t-1}\|^2\} \sigma_n^2}$$

that together with the identity $\mathsf{E}\{\|\tilde{S}_{t-1}\|^2\} = R_{t-1}$ directly leads to the result of this theorem. The theorem is proven. □

Below, two partial cases, most important for the identification and adaptive control applications, are considered in detail.

8.2.2.8 The Cramer–Lidbetter condition

Corollary 8.9. *Assume that the correlation coefficients $\rho_{t,s}$ (8.81) of the given random process (8.79) satisfy the so-called Cramer–Lidbetter condition (Cramer and Lidbetter, 1969), that is,*

$$\boxed{|\rho_{t,s}| \leq K \frac{t^\alpha + s^\alpha}{1 + |t - s|^\beta}} \tag{8.83}$$

where K, α, β are nonnegative constants satisfying

$$\boxed{2\,\alpha < \min\{1, \beta\}} \tag{8.84}$$

Then the strong law of large numbers holds, that is,

$$n^{-1} S_n \stackrel{a.s.}{\to} 0$$

Proof. Since

$$\mathsf{E}\{\|\xi_t\|^2\} = \sigma_t^2 = \rho_{t,t} \le 2Kt^{\alpha}$$

then

$$\begin{aligned} R_n &\le \frac{K}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{t^{\alpha}+s^{\alpha}}{1+|t-s|^{\beta}} \\ &= \frac{2K}{n^2}\left(\sum_{t=1}^{n}t^{\alpha}+\sum_{t=1}^{n}\sum_{s<t}\frac{t^{\alpha}+s^{\alpha}}{1+(t-s)^{\beta}}\right) \le \frac{2K}{n^{1-\alpha}}+2K\,I_n \end{aligned}$$

where

$$\begin{aligned} I_n &:= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s<t}\frac{t^{\alpha}+s^{\alpha}}{1+(t-s)^{\beta}} = I_n' + I_n'' \\ I_n' &:= \frac{1}{n^2}\sum_{t=1}^{n}t^{\alpha}\sum_{s<t}\frac{1}{1+(t-s)^{\beta}} \le \frac{1}{n^2}\int_0^n t^{\alpha}\begin{Bmatrix} \frac{t^{\beta-1}-1}{1-\beta}, & \beta\neq 1 \\ \ln t, & \beta = 1\end{Bmatrix} dt \\ &\le \text{Const}\begin{Bmatrix} n^{\alpha-\beta}, & \beta<1 \\ n^{\alpha+\varepsilon-1}, & \beta=1 \\ n^{\alpha-1}, & \beta>1\end{Bmatrix}, \quad \varepsilon>0 \\ I_n'' &:= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s<t}\frac{s^{\alpha}}{1+(t-s)^{\beta}} \le \frac{1}{n^2}\sum_{t=1}^{n}t^{\alpha}\sum_{s<t}\frac{1}{1+(t-s)^{\beta}} = I_n' \end{aligned}$$

So, finally, the following upper estimate for the 'double averaged' correlation function R_n holds:

$$R_n \le \text{Const}\begin{Bmatrix} n^{\max\{\alpha-1,\alpha-\beta\}}, & \beta<1 \\ n^{\alpha+\varepsilon-1}, & \beta=1 \\ n^{\alpha-1}, & \beta>1\end{Bmatrix}, \quad \varepsilon>0$$

The substitution of the right-hand side of the last inequality in (8.82) implies the desired result. □

8.2.2.9 Dependent processes generated by stable forming filters

Corollary 8.10. *Consider a centered random independent vector process $\{\xi_n\}_{n\ge 1}$ with finite variances σ_n^2 satisfying*

$$\boxed{\sum_{n=1}^{\infty}\frac{1}{n\,(n-1)}\sigma_n\sqrt{\sum_{r=0}^{n-1}\sigma_r^2} < \infty} \tag{8.85}$$

and generating the random vector sequence $\{\zeta_n\}$ according to the following expression:

$$\boxed{\zeta_n = \sum_{t=0}^{n} h_{n,t}\xi_t} \tag{8.86}$$

where the impulse response matrix function $h_{n,t}$ for any $t \leq n$ satisfies

$$\boxed{\left\| h_{n,t} \right\| \leq \hat{h}(n-t), \quad H := \sum_{\tau=0}^{\infty} \hat{h}(\tau) < \infty} \tag{8.87}$$

(such impulse function characterizes a stable, maybe, nonstationary forming filter). Then for the random sequence $\{\zeta_n\}$ the strong law of large numbers holds, that is,

$$\boxed{\frac{1}{n} \sum_{t=1}^{n} \zeta_t \overset{a.s.}{\to} 0}$$

Proof. The inequality

$$R_n \leq \frac{H^2}{n^2} \sum_{r=0}^{n} \sigma_r^2$$

implies

$$\frac{1}{n} \sqrt{R_{n-1} \sigma_n^2} \leq \frac{H}{n(n-1)} \sigma_n \sqrt{\sum_{r=0}^{n-1} \sigma_r^2}$$

that, together with the accepted assumptions, proves this corollary. □

8.3 Central limit theorem

If the LLN states that under certain conditions the arithmetic mean of difference between random variables $\{x_n\}_{n \geq 1}$ and their mathematical expectations converges to zero (in some probabilistic sense), that is,

$$\boxed{n^{-1} \sum_{t=1}^{n} (x_t - \mathsf{E}\{x_t\}) \underset{n \to \infty}{\to} 0} \tag{8.88}$$

the *Central Limit Theorem* (CLT) , which will be discussed in this section, concerns the distributional rate of convergence of the same difference. So, below we will analyze the conditions providing the distributional convergence of the normalized difference of random variables and their mathematical expectations to the standard normal distribution, namely,

$$\boxed{\begin{array}{c} P\left\{ \dfrac{1}{s_n} \displaystyle\sum_{t=1}^{n} (x_t - \mathsf{E}\{x_t\}) \leq x \right\} \underset{n \to \infty}{\to} \dfrac{1}{2\pi} \displaystyle\int_{t=-\infty}^{x} e^{-t^2/2} dt \\ s_n^2 := \displaystyle\sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} \mathsf{E}\left\{ (x_t - \mu_t)^2 \right\}, \quad \mu_t := \mathsf{E}\{x_t\} \end{array}} \tag{8.89}$$

that can be rewritten as

$$\boxed{\frac{1}{s_n}\sum_{t=1}^{n}(x_t - \mathsf{E}\{x_t\}) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1)} \tag{8.90}$$

Notice that if random variables are identically distributed with $\mathsf{E}\{x_n\} = \mu$ and $\sigma_n^2 = \sigma^2$ then

$$s_n = \sigma\sqrt{n}$$

and (8.90) becomes

$$\boxed{\frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n}(x_t - \mu) = \sqrt{n}\left[n^{-1}\sum_{t=1}^{n}\frac{(x_t - \mathsf{E}\{x_t\})}{\sigma}\right] \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1)}$$

which shows that the distributional rate of convergence in LLN (8.88) has the order $\mathcal{N}\left(0, n^{-1}\right)$.

We start with the simplest case of *independent identically distributed* (i.i.d.) random variables, and then will present the general result concerning independent random variables known as Lindeberg–Lévy–Feller theorem (see Lindeberg (1922), Lévy (1937) and Feller (1935)). We will conclude this section by presenting a version of CLM for dependent sequences, which helps a great deal in the analysis of the convergence rate of different procedures in concrete applications.

8.3.1 The i.i.d. case

Theorem 8.11. *Let $\{x_n\}_{n\geq 1}$ be i.i.d. random variables with $\mathsf{E}\{x_n\} = \mu$ and $\sigma_k^2 = \sigma^2 > 0$. Then*

$$\boxed{\frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n}(x_t - \mu) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1)} \tag{8.91}$$

Proof. Using the definition (5.3) of characteristic functions it is suffices to prove that

$$\varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) \underset{n\to\infty}{\to} e^{-t^2/2}$$

uniformly on $t \in \mathbb{R}$. Since $(x_t - \mu)/\sigma$ is zero-mean and has variance equal to 1, without loss of generality we may assume that $\mu = 1$ and $\sigma = 1$. Therefore, taking into account the i.i.d. property of $\{x_n\}_{n\geq 1}$, one has

$$\varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \varphi_{S_n/\sqrt{n}}(t) = \varphi_{S_n}\left(t/\sqrt{n}\right)$$
$$= \left[\varphi_{x_n}\left(t/\sqrt{n}\right)\right]^n = \left[1 - t^2/2n + o\left(t^2/n\right)\right]^n \underset{n\to\infty}{\to} e^{-t^2/2}$$

that completes the proof. □

8.3.2 General case of independent random variables

To present the generalization of the previous theorem to the case of independent but not obligatory identically distributed random variables we will need two fundamental concepts known as the *Lindeberg conditions*, involved in the formulation of the generalized CLT.

8.3.2.1 Lindeberg conditions

Definition 8.7. *The following properties of independent random variables $\{x_n\}_{n\geq 1}$ are called*

1. *the **first Lindeberg condition***

$$\boxed{\mathrm{L}_n^{(1)} := \max_{1\leq k\leq n} \sigma_k^2/s_n^2 \underset{n\to\infty}{\to} 0} \tag{8.92}$$

2. *the **second Lindeberg condition***

$$\boxed{\mathrm{L}_n^{(2)} := s_n^{-2}\sum_{k=1}^{n} \mathsf{E}\left\{|x_k-\mu_k|^2\,\chi\left(|x_k-\mu_k|/s_n>\varepsilon\right)\right\} \underset{n\to\infty}{\to} 0} \tag{8.93}$$

Lemma 8.8.

1. *The condition (8.92) implies*

$$s_n^2 \underset{n\to\infty}{\uparrow} \infty \tag{8.94}$$

2. *The condition (8.93) implies (8.92).*

Proof. Excluding the case when all $\sigma_k = 0$ we conclude that there exists an integer n_0 such that $\sigma_{n_0} > 0$, and therefore by (8.92) we have

$$\frac{\sigma_{n_0}^2}{s_n^2} \leq \mathrm{L}_n^{(1)} \underset{n\to\infty}{\to} 0$$

that proves the assertion 1. To prove 2. it suffices to note that for any $\varepsilon > 0$

$$\mathrm{L}_n^{(1)} \leq s_n^{-2}\max_{1\leq k\leq n} \mathsf{E}\left\{|x_k-\mu_k|^2\,\chi\left(|x_k-\mu_k|\leq\varepsilon s_n\right)\right\} \\ + s_n^{-2}\max_{1\leq k\leq n} \mathsf{E}\left\{|x_k-\mu_k|^2\,\chi\left(|x_k-\mu_k|>\varepsilon s_n\right)\right\} \leq \varepsilon^2 + \mathrm{L}_n^{(2)} \underset{n\to\infty}{\downarrow} \varepsilon^2$$

providing

$$\limsup \mathrm{L}_n^{(1)} \leq \varepsilon^2$$

that proves the theorem. □

Now we are ready to present the legendary result.

8.3.2.2 CLT in the Lindeberg–Lévy–Feller form

Theorem 8.12. (Lindeberg–Lévy–Feller) *If for the sequence $\{x_n\}$ of independent random variables the first Lindeberg condition (8.93) satisfied, then*

$$\boxed{\frac{1}{s_n}\sum_{t=1}^{n}(x_t-\mu_t)\underset{n\to\infty}{\overset{d}{\to}}\mathcal{N}(0,1)} \tag{8.95}$$

To give the proof of this theorem we shall need the auxiliary lemma given below.

Lemma 8.9. (Gut, 2005) *Under the second Lindeberg condition (8.93) the following properties hold:*

1.

$$\boxed{\left|\sum_{k=1}^{n}\left(\ln\varphi_{x_k}(t/s_n)+\left[1-\varphi_{x_k}(t/s_n)\right]\right)\right|\underset{n\to\infty}{\to}0} \tag{8.96}$$

2.

$$\boxed{\left|\sum_{k=1}^{n}\left(\varphi_{x_k}(t/s_n)-\left[1-\frac{\sigma_k^2}{2s_n^2}t^2\right]\right)\right|\underset{n\to\infty}{\to}0} \tag{8.97}$$

Proof.
1. Notice that in view of Lemma 8.8

$$\left|1-\varphi_{x_k}(t/s_n)\right|\le \mathsf{E}\left\{\frac{x_k^2}{2s_n^2}t^2\right\}\le\frac{t^2}{2}\mathrm{L}_n^{(1)}\to 0$$

uniformly on $k=1,2,\ldots,n$ whereas $n\to\infty$. Therefore

$$\begin{aligned}
&\left|\sum_{k=1}^{n}\left(\ln\varphi_{x_k}(t/s_n)+\left[1-\varphi_{x_k}(t/s_n)\right]\right)\right| \\
&\quad\le\sum_{k=1}^{n}\left|\ln\left[1-\left(1-\varphi_{x_k}(t/s_n)\right)\right]+\left[1-\varphi_{x_k}(t/s_n)\right]\right| \\
&\quad\le\sum_{k=1}^{n}\left|1-\varphi_{x_k}(t/s_n)\right|^2\le\max_{1\le k\le n}\left|1-\varphi_{x_k}(t/s_n)\right|\sum_{k=1}^{n}\left|1-\varphi_{x_k}(t/s_n)\right| \\
&\quad\le\frac{t^2}{2}\mathrm{L}_n^{(1)}\sum_{k=1}^{n}\mathsf{E}\left\{\frac{x_k^2}{2s_n^2}t^2\right\}\le\frac{t^4}{4}\mathrm{L}_n^{(1)}\underset{n\to\infty}{\to}0
\end{aligned}$$

2. Applying the estimate

$$\left|\varphi_{x_k}(t/s_n)-\left[1-\frac{\sigma_k^2}{2s_n^2}t^2\right]\right|\le\mathsf{E}\left\{\min\left\{\frac{x_k^2}{2s_n^2}t^2;\ \frac{|x_k|^3}{6s_n^3}t^3\right\}\right\}$$

yields

$$\left|\sum_{k=1}^{n}\left(\varphi_{x_k}(t/s_n)-\left[1-\frac{\sigma_k^2}{2s_n^2}t^2\right]\right)\right| \le \sum_{k=1}^{n}\left|\varphi_{x_k}(t/s_n)-\left[1-\frac{\sigma_k^2}{2s_n^2}t^2\right]\right|$$
$$\le \sum_{k=1}^{n}\mathsf{E}\left\{\min\left\{\frac{x_k^2}{2s_n^2}t^2;\ \frac{|x_k|^3}{6s_n^3}t^3\right\}\right\} \le \sum_{k=1}^{n}\mathsf{E}\left\{\frac{x_k^2}{2s_n^2}t^2\chi(|x_k|>\varepsilon s_n)\right\}$$
$$+\sum_{k=1}^{n}\mathsf{E}\left\{\frac{|x_k|^3}{6s_n^3}t^3\chi(|x_k|\le\varepsilon s_n)\right\} \le \sum_{k=1}^{n}\mathsf{E}\left\{\varepsilon s_n\frac{|x_k|^2}{6s_n^3}t^3\chi(|x_k|\le\varepsilon s_n)\right\}$$
$$+t^2\mathrm{L}_n^{(2)} \le \frac{\varepsilon t^3}{6}+t^2\mathrm{L}_n^{(2)} \underset{n\to\infty}{\to} \frac{\varepsilon t^3}{6}$$

which, due to the arbitrariness of ε, proves (8.97). □

Proof of Theorem 8.12.

1. Assuming $\mu_k = 0$, in view of the independency of $\{x_n\}_{n\ge1}$ and applying Lemma 8.9, one has

$$\varphi_{S_n/s_n}(t) = \varphi_{S_n}(t/s_n) = \prod_{k=1}^{n}\varphi_{x_k}(t/s_n) = \exp\left\{\sum_{k=1}^{n}\ln\varphi_{x_k}(t/s_n)\right\}$$
$$= \exp\left\{-\sum_{k=1}^{n}\left[1-\varphi_{x_k}(t/s_n)\right]\right\}+o(1)$$
$$= \exp\left\{-\sum_{k=1}^{n}\left[1-1+\frac{it}{s_n}0+\frac{(it)^2}{2s_n^2}\sigma_k^2\right]\right\}+o(1)$$
$$= \exp\left\{-\frac{t^2}{2s_n^2}\sum_{k=1}^{n}\sigma_k^2\right\}+o(1) = \exp\left\{-\frac{t^2}{2}\right\}+o(1) \underset{n\to\infty}{\to} \exp\left\{-\frac{t^2}{2}\right\}$$

that completes the proof of the theorem. □

Remark 8.8. *It can be proven (see pp. 337–338 in Gut (2005)) that if (8.92) and (8.95) are satisfied then so is (8.93).*

8.3.2.3 Lyapunov's condition as a relaxing property

The verification of the second Lindeberg condition (8.93) is not a simple task. A slightly stronger, but definitely simpler, condition was given in Lyapunov (1900).

Theorem 8.13. **(Lyapunov, 1900)** *Let $\{x_n\}_{n\ge1}$ are given as before in Theorem 8.12, and, in addition, for some $r > 2$*

$$\mathsf{E}\left\{|x_n|^r\right\} < \infty \quad \textit{for all} \quad k \ge 1$$

and

$$\boxed{\beta_n(r) := \frac{1}{s_n^r}\sum_{k=1}^{n}\mathsf{E}\left\{(x_k-\mu_k)^r\right\} \underset{n\to\infty}{\to} 0} \tag{8.98}$$

Then the second Lindeberg condition (8.93), and therefore, the CLT (8.95) holds.

Proof. For any $\varepsilon > 0$

$$\begin{aligned}
\mathrm{L}_n^{(2)} &:= s_n^{-2} \sum_{k=1}^{n} \mathsf{E}\left\{ |x_k - \mu_k|^2 \chi\left(|x_k - \mu_k| > \varepsilon s_n\right) \right\} \\
&\le s_n^{-2} \sum_{k=1}^{n} \frac{1}{(\varepsilon s_n)^{r-2}} \mathsf{E}\left\{ |x_k - \mu_k|^r \chi\left(|x_k - \mu_k| > \varepsilon s_n\right) \right\} \\
&\le \frac{1}{(\varepsilon)^{r-2} s_n^r} \sum_{k=1}^{n} \frac{1}{(\varepsilon s_n)^{r-2}} \mathsf{E}\left\{ |x_k - \mu_k|^r \right\} = \frac{1}{(\varepsilon)^{r-2}} \beta_n(r) \underset{n\to\infty}{\to} 0
\end{aligned}$$

that proves the desired result. □

8.3.3 *CLT for martingale arrays (or double sequences)*

Consider the following triangular array $\left\{x_{n,j}\right\}_{1\le j\le n}$ of random variables $x_{n,j}$:

$$\begin{array}{llllll}
x_{1,1} & & & & & \\
x_{2,1} & x_{2,2} & & & & \\
x_{3,1} & x_{3,2} & x_{3,3} & & & \\
\cdot & \cdot & \cdot & \cdot & & \\
x_{n,1} & x_{n,2} & \cdot & \cdot & x_{n,n} & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}$$

Below we will consider basically the so-called *martingale arrays* where

$$\mathsf{E}\left\{x_{n,j} \mid \mathcal{F}_{n,j-1}\right\} \overset{a.s.}{=} 0$$

which is usually obtained from ordinary martingales ($\mathsf{E}\left\{\xi_j \mid \mathcal{F}_{j-1}\right\} \overset{a.s.}{=} 0$) by the following construction:

$$\begin{aligned}
&S_n := \sum_{j=1}^{n} \xi_j, \qquad s_n^2 := \sum_{j=1}^{n} \sigma_j^2 \\
&x_{n,j} := \xi_j / s_n, \qquad \mathcal{F}_{n,j} := \mathcal{F}_j \\
&\frac{1}{s_n} S_n = \sum_{j=1}^{n} \xi_j / s_n = \sum_{j=1}^{n} x_{n,j}
\end{aligned}$$

so that the distributions in different rows typically are not the same and the rows are not independent.

To present the vectorial version of CLT for triangular arrays we shall need the following auxiliary technical result.

Lemma 8.10. *Let ξ and η be random vectors from $\mathbb{R}^m$ with finite second moments. Define their characteristic functions*

$$\varphi_\xi(u) := \mathsf{E}\left\{e^{iu^\intercal \xi}\right\}, \quad \varphi_\eta(u) := \mathsf{E}\left\{e^{iu^\intercal \eta}\right\}, \quad u \in \mathbb{R}^m$$

and the corresponding covariation matrices

$$\operatorname{cov}(\xi,\xi) := \mathsf{E}\left\{\xi\xi^{\intercal}\right\}, \quad \operatorname{cov}(\eta,\eta) := \mathsf{E}\left\{\eta\eta^{\intercal}\right\}$$

Then for any $u \in \mathbb{R}^m$ *and any* $\varepsilon > 0$ *such that*

$$\|u\| \geq \varepsilon^{-1}\left(\varepsilon \leq \|u\|^{-1}\right)$$

the following inequalities hold:

$$\begin{aligned}
&\left|\varphi_\xi(u) - \varphi_\eta(u)\right| \leq \left|u^{\intercal}\mathsf{E}\{\xi - \eta\}\right| + \frac{\|u\|^2}{2}\|\operatorname{cov}(\xi,\xi) - \operatorname{cov}(\eta,\eta)\| \\
&\quad + \frac{\varepsilon}{6}\|u\|^3\left[\mathsf{E}\left\{\|\xi\|^2\right\} + \mathsf{E}\left\{\|\eta\|^2\right\}\right] \\
&\quad + \frac{\|u\|^2}{2}\left[\mathsf{E}\left\{\|\xi\|^2\chi(\|\xi\| > \varepsilon)\right\} + \mathsf{E}\left\{\|\eta\|^2\chi(\|\eta\| > \varepsilon)\right\}\right]
\end{aligned} \tag{8.99}$$

Proof. Using the Taylor expansion for the exponent $\exp\{iu^{\intercal}x\}$ up to the term of the third order in the region where $\|x\| \leq \varepsilon$ and up to the term of the second order in the region where $\|x\| > \varepsilon$ we obtain

$$\begin{aligned}
\exp\left\{iu^{\intercal}x\right\} &= \left[1 + iu^{\intercal}x - \frac{(u^{\intercal}x)^2}{2} - i\frac{(u^{\intercal}x)^3}{6}c_1\right]\chi(\|x\| \leq \varepsilon) \\
&\quad + \left[1 + iu^{\intercal}x - \frac{(u^{\intercal}x)^2}{2}c_2\right]\chi(\|x\| > \varepsilon) = 1 + iu^{\intercal}x - \frac{(u^{\intercal}x)^2}{2} \\
&\quad + \frac{1 - c_2}{2}\left(u^{\intercal}x\right)^2\chi(\|x\| > \varepsilon) - i\frac{c_1}{6}\left(u^{\intercal}x\right)^3\chi(\|x\| \leq \varepsilon)
\end{aligned}$$

where $c_i = c_i(u,x) \in (0,1)$ $(i = 1,2)$. Therefore

$$\begin{aligned}
\left|\varphi_\xi(u) - \varphi_\eta(u)\right| &\leq \left|u^{\intercal}\mathsf{E}\{\xi - \eta\}\right| + \frac{1}{2}\left|\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^2\right\} - \mathsf{E}\left\{\left(u^{\intercal}\eta\right)^2\right\}\right| \\
&\quad + \frac{1}{2}\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^2\chi(\|\xi\| > \varepsilon)\right\} + \mathsf{E}\left\{\left(u^{\intercal}\eta\right)^2\chi(\|\eta\| > \varepsilon)\right\} \\
&\quad + \frac{1}{6}\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^2\chi(\|\xi\| \leq \varepsilon)\right\} + \mathsf{E}\left\{\left(u^{\intercal}\eta\right)^2\chi(\|\eta\| \leq \varepsilon)\right\}
\end{aligned}$$

Using then the following evident relations

$$\begin{aligned}
&\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^2\right\} = u^{\intercal}\operatorname{cov}(\xi,\xi)\,u, \quad \mathsf{E}\left\{\left(u^{\intercal}\eta\right)^2\right\} = u^{\intercal}\operatorname{cov}(\eta,\eta)\,u \\
&\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^2\chi(\|\xi\| \leq \varepsilon)\right\} \leq \|u\|^2\,\mathsf{E}\left\{\|\xi\|^2\chi(\|\xi\| > \varepsilon)\right\} \\
&\mathsf{E}\left\{\left(u^{\intercal}\eta\right)^2\chi(\|\eta\| \leq \varepsilon)\right\} \leq \|u\|^2\,\mathsf{E}\left\{\|\eta\|^2\chi(\|\eta\| > \varepsilon)\right\} \\
&\mathsf{E}\left\{\left(u^{\intercal}\xi\right)^3\chi(\|\xi\| \leq \varepsilon)\right\} = \varepsilon\,\|u\|^3\,\mathsf{E}\left\{\|\xi\|^2\right\}
\end{aligned}$$

$$\mathsf{E}\left\{\left(u^{\intercal}\eta\right)^{3}\chi\left(\|\eta\|\leq\varepsilon\right)\right\}=\varepsilon\|u\|^{3}\mathsf{E}\left\{\|\eta\|^{2}\right\}$$

the result follows. □

The next theorem represents the version of CLT for triangular vector-arrays.

Theorem 8.14. **(Sacks, 1958)** *Suppose that for a triangular vector array $\{x_{n,j}\}_{1\leq j\leq n}$ of random vectors $x_{n,j}\in\mathbb{R}^m$ the following properties hold:*

1.

$$\boxed{\mathsf{E}\left\{x_{n,j}\mid\mathcal{F}_{n,j-1}\right\}\overset{a.s.}{=}0} \tag{8.100}$$

2.

$$\boxed{\begin{array}{c}\sup\limits_{n\geq1}\sum\limits_{j=1}^{n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\right\}<\infty,\quad\lim\limits_{n\to\infty}K_n=K\\ K_n:=\sum\limits_{j=1}^{n}K_{n,j},\quad K_{n,j}:=\mathsf{E}\left\{x_{n,j}x_{n,j}^{\intercal}\right\}\end{array}} \tag{8.101}$$

3.

$$\boxed{\begin{array}{c}\lim\limits_{n\to\infty}\sum\limits_{j=1}^{n}\mathsf{E}\left\{\left\|K_{n,j}-L_{n,j}\right\|\right\}=0\\[2ex] L_{n,j}:=\mathsf{E}\left\{x_{n,j}x_{n,j}^{\intercal}\mid\mathcal{F}_{n,j-1}\right\}\\ \mathcal{F}_{n,j}:=\sigma\left(x_{n,1},x_{n,2},\ldots,x_{n,j}\right)\end{array}} \tag{8.102}$$

4. for any $\varepsilon>0$

$$\boxed{\lim_{n\to\infty}\sum_{j=1}^{n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\chi\left(\left|x_{n,j}\right|>\varepsilon\right)\right\}=0} \tag{8.103}$$

Then

$$\boxed{y_n:=\sum_{j=1}^{n}x_{n,j}\underset{n\to\infty}{\overset{d}{\to}}\mathcal{N}(0,K)} \tag{8.104}$$

Proof. Introduce the Gaussian triangular array $\{\eta_{n,j}\}_{1\leq j\leq n}$ which contains Gaussian centered random variables $\eta_{n,j}$ which are independent within this array and of the given array $\{x_{n,j}\}_{1\le j\le n}$ as well. Suppose that, by the construction, the covariation matrices of $\eta_{n,j}$ and $x_{n,j}$ coincide, that is, for all $j = 1, \dots, n$ and all $n = 1, 2, \dots$

$$\mathsf{E}\left\{\eta_{n,j}\eta_{n,j}^{\mathsf{T}}\right\} = K_{n,j}$$

Define

$$\varphi_{y_n}(u) := \mathsf{E}\left\{e^{iu^{\mathsf{T}}y_n}\right\}$$

$$\varphi_{z_n}(u) := \mathsf{E}\left\{e^{iu^{\mathsf{T}}z_n}\right\}, \quad z_n := \sum_{j=1}^{n}\eta_{n,j}$$

First, notice that z_n is also a Gaussian random variable with zero-mean and the covariation matrix K_n which converges to a limit K and, therefore, $\varphi_{z_n}(u)$ converges to the characteristic function of the Gaussian random vector with the covariation matrix K. So, it is sufficient to show that

$$\left|\varphi_{y_n}(u) - \varphi_{z_n}(u)\right| \underset{n\to\infty}{\to} 0$$

for any $u \in \mathbb{R}^m$ ($u \neq 0$). Let us use the following identity:

$$\begin{aligned}\varphi_{y_n}(u) - \varphi_{z_n}(u) &= \sum_{j=1}^{n}\left[\exp\left(iu^{\mathsf{T}}\left(\sum_{s=1}^{j}x_{n,s} + \sum_{s=j+1}^{n}\eta_{n,s}\right)\right)\right.\\&\qquad\left. - \exp\left(iu^{\mathsf{T}}\left(\sum_{s=1}^{j-1}x_{n,s} + \sum_{s=j}^{n}\eta_{n,s}\right)\right)\right]\\&= \sum_{j=1}^{n}\left[\exp\left\{iu^{\mathsf{T}}x_{n,j}\right\} - \exp\left\{iu^{\mathsf{T}}\eta_{n,j}\right\}\right]\exp\left\{iu^{\mathsf{T}}v_{n,j}\right\} \qquad (8.105)\end{aligned}$$

where

$$v_{n,j} := \sum_{s=1}^{j-1}x_{n,s} + \sum_{s=j+1}^{n}\eta_{n,s}$$

Define then the conditional mathematical expectation of $\varphi_{x_{n,k}}(u)$ as

$$\tilde{\varphi}_{x_{n,j}}(u) := \mathsf{E}\left\{e^{iu^{\mathsf{T}}x_{n,j}} \mid \mathcal{F}_{n,j-1}\right\}$$

Then using the identity (8.105) and that $|\exp(iz)| = 1$ for any $z \in \mathbb{R}$, it follows that

$$\begin{aligned}&\left|\varphi_{y_n}(u) - \varphi_{z_n}(u)\right|\\&\quad= \left|\mathsf{E}\left\{\sum_{j=1}^{n}\mathsf{E}\left[\exp\left\{iu^{\mathsf{T}}x_{n,j}\right\} - \exp\left\{iu^{\mathsf{T}}\eta_{n,j}\right\}\right]\exp\left\{iu^{\mathsf{T}}v_{n,j}\right\} \mid \mathcal{F}_{n,j-1}\right\}\right|\end{aligned}$$

$$\left|\mathsf{E}\left\{\sum_{j=1}^{n}\exp\left\{iu^{\intercal}\sum_{s=1}^{j-1}x_{n,s}\right\}\left[\tilde{\varphi}_{x_{n,j}}(u)-\varphi_{\eta_{n,j}}(u)\right]\right\}\right.$$
$$\left.\times\ \mathsf{E}\left\{\sum_{j=1}^{n}\exp\left\{iu^{\intercal}\sum_{s=j+1}^{n}\eta_{n,s}\right\}\right\}\right|\leq\sum_{j=1}^{n}\mathsf{E}\left\{\left|\tilde{\varphi}_{x_{n,j}}(u)-\varphi_{\eta_{n,j}}(u)\right|\right\}$$

Then by Lemma 8.10 we get

$$\left|\tilde{\varphi}_{x_{n,j}}(u)-\varphi_{\eta_{n,j}}(u)\right|\leq\frac{\varepsilon}{6}\|u\|^{3}\left[\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\right\}+\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\right\}\right]$$
$$+\frac{\|u\|^{2}}{2}\left[\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\chi\left(\left\|x_{n,j}\right\|>\varepsilon\right)\right\}+\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\chi\left(\|\eta\|>\varepsilon\right)\right\}\right]$$

and hence,

$$\left|\varphi_{y_n}(u)-\varphi_{z_n}(u)\right|\leq\sum_{j=1}^{n}\mathsf{E}\left\{\left|\tilde{\varphi}_{x_{n,j}}(u)-\varphi_{\eta_{n,j}}(u)\right|\right\}$$
$$\leq\frac{\varepsilon}{6}\|u\|^{3}2\sum_{j=1}^{n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\right\}$$
$$+\frac{\|u\|^{2}}{2}\sum_{j=1}^{n}\left[\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\chi\left(\left\|x_{n,j}\right\|>\varepsilon\right)\right\}+\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\chi\left(\|\eta\|>\varepsilon\right)\right\}\right]$$

Therefore, by the assumptions of the theorem, it is sufficient to demonstrate that for the Gaussian vector processes $\{\eta_{n,j}\}$ we have

$$\sum_{j=1}^{n}\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\chi\left(\left\|\eta_{n,j}\right\|>\varepsilon\right)\right\}\underset{n\to\infty}{\to}0$$

Since $\left\|K_{n,j}\right\|\leq\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\right\}$ then for any $\varepsilon>0$ it follows that

$$\max_{1\leq j\leq n}\left\|K_{n,j}\right\|\leq\max_{1\leq j\leq n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\right\}\leq\varepsilon^{2}+\max_{1\leq j\leq n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\chi\left(\left\|\eta_{n,j}\right\|>\varepsilon\right)\right\}$$
$$\leq\varepsilon^{2}+\sum_{j=1}^{n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^{2}\chi\left(\left\|\eta_{n,j}\right\|>\varepsilon\right)\right\}$$

and, hence, by the property (8.103)

$$\max_{1\leq j\leq n}\left\|K_{n,j}\right\|\underset{n\to\infty}{\to}0$$

Then, by the CBS inequality (4.16) and the Chebyshev inequality (4.10) one has

$$\sum_{j=1}^{n} \mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\chi\left(\left\|\eta_{n,j}\right\|>\varepsilon\right)\right\} \le \sum_{j=1}^{n}\sqrt{\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{4}\right\}P\left\{\left\|\eta_{n,j}\right\|>\varepsilon\right\}}$$

$$\le \varepsilon^{-2}\sum_{j=1}^{n}\sqrt{\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{4}\right\}\mathsf{E}\left\{\left\|\eta_{n,j}\right\|^{2}\right\}} \le \text{Const}\cdot\varepsilon^{-2}\max_{1\le j\le n}\left\|K_{n,j}\right\| \underset{n\to\infty}{\to} 0$$

that completes the proof. □

The next result represents a useful version of Theorem 8.14, which is frequently used in *Stochastic Optimization Theory* (Nevel'son and Khas'minski, 1972).

Theorem 8.15. *If the random sequence* $\{\zeta_n\}_{n\ge1}$ $\left(\zeta_n \in \mathbb{R}^N\right)$ *satisfies the following properties for some sequences of deterministic nonsingular matrices* $\{A_n\}_{n\ge1}$ $\left(A_n \in \mathbb{R}^{N\times N}\right)$

1.

$$\boxed{\begin{array}{c}\mathsf{E}\left\{\xi_{k,j}\xi_{k,j}^{\mathsf{T}} \mid \mathcal{F}_{k,j-1}\right\} \overset{a.s.}{=} \Xi_{k,j}, \quad 1\le j\le k \\ \xi_{k,j} := \mathsf{E}\left\{\zeta_k \mid \mathcal{F}_{k,j}\right\} - \mathsf{E}\left\{\zeta_k \mid \mathcal{F}_{k,j-1}\right\}\end{array}} \tag{8.106}$$

where $\Xi_{k,j} \in \mathbb{R}^{N\times N}$ *is a nonnegative deterministic matrix*

2.

$$\boxed{\frac{1}{n}A_n^{-1}\max_{1\le j\le n}\left(\sum_{k=j+1}^{n} A_k\,\Xi_{k,j}\,A_k^{\mathsf{T}}\right)\left(A_n^{-1}\right)^{\mathsf{T}} \underset{n\to\infty}{\to} 0} \tag{8.107}$$

3. there exists the limit

$$\boxed{V_n := \frac{1}{n}A_n^{-1}\sum_{k=1}^{n} A_k\left(\sum_{j=1}^{k}\Xi_{k,j}\right)A_k^{\mathsf{T}}\left(A_n^{-1}\right)^{\mathsf{T}} \underset{n\to\infty}{\to} K} \tag{8.108}$$

then

$$\boxed{y_n := \frac{1}{\sqrt{n}}A_n^{-1}\sum_{k=1}^{n} A_k\zeta_k \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0,K\right)} \tag{8.109}$$

Proof. Notice that ζ_n for any $n\ge1$ can be represented as

$$\zeta_k = \sum_{j=1}^{k}\xi_{k,j}$$

$$\xi_{k,j} := \mathsf{E}\left\{\zeta_k \mid \mathcal{F}_{k,j}\right\} - \mathsf{E}\left\{\zeta_k \mid \mathcal{F}_{k,j-1}\right\}$$

$$\mathcal{F}_{k,j} := \sigma\left(\zeta_1, \ldots, \zeta_j\right), \quad \mathsf{E}\left\{\zeta_k \mid \mathcal{F}_{k,0}\right\} = \mathsf{E}\left\{\zeta_k\right\} = 0$$

that leads to the following expression for y_n:

$$\begin{aligned} y_n &:= \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=1}^{n} A_k \sum_{j=1}^{k} \xi_{k,j} \\ &= \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=1}^{n} A_k \sum_{j=1}^{n} \chi_{j \le k} \xi_{k,j} = \sum_{j=1}^{n} x_{n,j} \end{aligned}$$

with

$$x_{n,j} := \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=1}^{n} A_k \chi_{j \le k} \xi_{k,j} = \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=j+1}^{n} A_k \xi_{k,j}$$

and to complete the proof one has to check the conditions of Theorem 8.14 for the triangular vector array $\left\{x_{n,j}\right\}_{1 \le j \le n}$ of random vectors $x_{n,j} \in \mathbb{R}^N$.

1. Evidently

$$\mathsf{E}\left\{\xi_{k,j} \mid \mathcal{F}_{k,j-1}\right\} \overset{a.s.}{=} 0, \qquad \mathcal{F}_{k,j-1} \subset \mathcal{F}_{n,j-1} \quad (k = j+1, \ldots, n)$$

and hence,

$$\begin{aligned} \mathsf{E}\left\{x_{n,j} \mid \mathcal{F}_{n,j-1}\right\} &= \mathsf{E}\left\{\frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=j+1}^{n} A_k \xi_{k,j} \mid \mathcal{F}_{n,j-1}\right\} \\ &= \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=j+1}^{n} A_k \mathsf{E}\left\{\xi_{k,j} \mid \mathcal{F}_{n,j-1}\right\} \\ &= \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=j+1}^{n} A_k \mathsf{E}\left\{\mathsf{E}\left\{\xi_{k,j} \mid \mathcal{F}_{k,j-1}\right\} \mid \mathcal{F}_{n,j-1}\right\} \overset{a.s.}{=} 0 \end{aligned}$$

that gives the condition (8.100).

2. For $k > s$ $(s < j)$ one has $\mathsf{E}\left\{\xi_{k,j} \mid \mathcal{F}_{k,s}\right\} \overset{a.s.}{=} 0$ and therefore

$$\begin{aligned} \mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\mathsf{T}}\right\} &= \mathsf{E}\left\{\mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\mathsf{T}} \mid \mathcal{F}_{k,j-1}\right\}\right\} \\ &= \mathsf{E}\left\{\mathsf{E}\left\{\mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\mathsf{T}} \mid \mathcal{F}_{k,s}\right\} \mid \mathcal{F}_{k,j-1}\right\}\right\} \\ &= \mathsf{E}\left\{\mathsf{E}\left\{\mathsf{E}\left\{\xi_{k,j} \mid \mathcal{F}_{k,s}\right\} \xi_{s,j}^{\mathsf{T}} \mid \mathcal{F}_{k,j-1}\right\}\right\} = 0 \end{aligned}$$

By the same reason, for $s > k$ it follows $\mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\mathsf{T}}\right\} = 0$. So,

$$\mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\mathsf{T}}\right\} = \mathsf{E}\left\{\xi_{k,j}\xi_{k,j}^{\mathsf{T}}\right\} \delta_{k,s}$$

and therefore, by the assumption (8.106), it follows that

$$\begin{aligned} L_{n,j} &:= \mathsf{E}\left\{x_{n,j}x_{n,j}^{\intercal} \mid \mathcal{F}_{n,j-1}\right\} \\ &= \frac{1}{n}A_n^{-1}\sum_{k=j+1}^{n}\sum_{s=j+1}^{n}A_k\mathsf{E}\left\{\xi_{k,j}\xi_{s,j}^{\intercal} \mid \mathcal{F}_{n,j-1}\right\}A_s^{\intercal}\left(A_n^{-1}\right)^{\intercal} \\ &= \frac{1}{n}A_n^{-1}\sum_{k=j+1}^{n}A_k\mathsf{E}\left\{\xi_{k,j}\xi_{k,j}^{\intercal} \mid \mathcal{F}_{k,j-1}\right\}A_k^{\intercal}\left(A_n^{-1}\right)^{\intercal} \\ &= \frac{1}{n}A_n^{-1}\sum_{k=j+1}^{n}A_k\,\Xi_{k,j}A_k^{\intercal}\left(A_n^{-1}\right)^{\intercal} \end{aligned}$$

and

$$\begin{aligned} K_{n,j} &:= \mathsf{E}\left\{x_{n,j}x_{n,j}^{\intercal}\right\} \\ &= \mathsf{E}\left\{L_{n,j}\right\} = \frac{1}{n}A_n^{-1}\sum_{k=j+1}^{n}A_k\,\Xi_{k,j}A_k^{\intercal}\left(A_n^{-1}\right)^{\intercal} \overset{a.s.}{=} L_{n,j} \end{aligned}$$

that proves the fulfilling of (8.102).

3. Since by the assumption (8.107)

$$K_{n,j} \le \max_{1\le j\le n} K_{n,j} = \frac{1}{n}A_n^{-1}\max_{1\le j\le n}\left(\sum_{k=j+1}^{n}A_k\,\Xi_{k,j}A_k^{\intercal}\right)\left(A_n^{-1}\right)^{\intercal} \underset{n\to\infty}{\to} 0$$

then, by the CBS inequality (4.16) and the Chebyshev inequality (4.10), it follows that

$$\begin{aligned} &\sum_{j=1}^{n}\mathsf{E}\left\{\left\|x_{n,j}\right\|^2\chi\left(\left\|x_{n,j}\right\| > \varepsilon\right)\right\} \le \sum_{j=1}^{n}\sqrt{\mathsf{E}\left\{\left\|x_{n,j}\right\|^4\right\}P\left\{\left\|x_{n,j}\right\| > \varepsilon\right\}} \\ &\quad \le \varepsilon^{-2}\sum_{j=1}^{n}\sqrt{\mathsf{E}\left\{\left\|x_{n,j}\right\|^4\right\}\mathsf{E}\left\{\left\|x_{n,j}\right\|^2\right\}} \le \text{Const}\cdot\varepsilon^{-2}\max_{1\le j\le n}\left\|K_{n,j}\right\| \underset{n\to\infty}{\to} 0 \end{aligned}$$

that proves the fulfilling of the condition (8.103).

4. The condition (8.101) results from (8.108) since

$$\begin{aligned} \sum_{j=1}^{n}K_{n,j} &= \frac{1}{n}A_n^{-1}\left[\sum_{j=1}^{n}\sum_{k=j+1}^{n}A_k\mathsf{E}\left\{\xi_{kj,}\xi_{k,j}^{\intercal}\right\}A_k^{\intercal}\right]\left(A_n^{-1}\right)^{\intercal} \\ &= \frac{1}{n}A_n^{-1}\sum_{k=1}^{n}A_k\sum_{j=1}^{k}\Xi_{k,j}A_k^{\intercal}\left(A_n^{-1}\right)^{\intercal} \underset{n\to\infty}{\to} K \end{aligned}$$

Finally, the identity $y_n = \sum_{j=1}^{n} x_{n,j}$ leads to (8.109) that completes the proof. □

Remark 8.9. *Obviously K in (8.108) can be also calculated using the covariation matrices $\Theta_{k,s} := \mathsf{E}\left\{\zeta_k \zeta_s^{\mathsf{T}}\right\}$ of the given sequence $\{\zeta_n\}_{n\geq 1}$ as*

$$\boxed{K = \lim_{n\to\infty} \mathsf{E}\left\{y_n y_n^{\mathsf{T}}\right\} = \lim_{n\to\infty} \frac{1}{n} A_n^{-1} \sum_{k=1}^{n} \sum_{s=1}^{n} A_k \Theta_{k,s} A_s^{\mathsf{T}} \left(A_n^{-1}\right)^{\mathsf{T}}} \tag{8.110}$$

Define

$$\boxed{R_k := \sum_{j=1}^{k} \Xi_{k,j}} \tag{8.111}$$

Then

$$V_n := \frac{1}{n} A_n^{-1} \left(\sum_{k=1}^{n} A_k R_k A_k^{\mathsf{T}}\right) \left(A_n^{-1}\right)^{\mathsf{T}}$$

which can be represented in the following recurrent form

$$\boxed{V_n = \left(1 - \frac{1}{n}\right) A_n^{-1} A_{n-1} V_{n-1} \left(A_n^{-1} A_{n-1}\right)^{\mathsf{T}} + \frac{1}{n} R_n} \tag{8.112}$$

The condition (8.108) means exactly that there exists the limit $K = \lim_{n\to\infty} V_n$. The next lemma states the conditions to the matrix sequence $\{A_n\}_{n\geq 1}$ which guarantee the existence of this limit.

Lemma 8.11. **(Poznyak and Tchikin, 1985)** *Let*

1. *there exist the limit*

$$\boxed{A := \lim_{n\to\infty} n \left(A_n^{-1} A_{n-1} - I\right)} \tag{8.113}$$

such that the matrix $\left(A - \frac{1}{2} I\right)$ is Hurwitz;

2. *there exist the limit*

$$\boxed{R := \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} R_t = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{t} \Xi_{t,s} > 0} \tag{8.114}$$

Then $\{V_n\}_{n\geq 1}$ converges to K, that is,

$$\boxed{V_n \underset{n\to\infty}{\rightarrow} K = K^{\mathsf{T}} > 0} \tag{8.115}$$

where K satisfies the following algebraic matrix Lyapunov equation

$$\boxed{\left(A - \frac{1}{2}I\right) K + K^{\intercal} \left(A - \frac{1}{2}I\right)^{\intercal} = -R} \tag{8.116}$$

Proof. By the assumption (8.113) it follows that

$$\begin{aligned}
&V_n \\
&= \left(1 - \frac{1}{n}\right) \left(\frac{n\left[A_n^{-1}A_{n-1} - I\right]}{n} + I\right) V_{n-1} \left(\frac{n\left[A_n^{-1}A_{n-1} - I\right]}{n} + I\right)^{\intercal} + \frac{1}{n} R_n \\
&= \left(1 - \frac{1}{n}\right) \left(\frac{A}{n} + I\left(1 + o\left(n^{-1}\right)\right)\right) V_{n-1} \left(\frac{A^{\intercal}}{n} + I\left(1 + o\left(n^{-1}\right)\right)\right) + \frac{1}{n} R_n \\
&= V_{n-1} + \frac{1}{n}\left[\left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right) V_{n-1}\right. \\
&\quad \left. + V_{n-1}\left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right)^{\intercal}\right] + \frac{1}{n} R_n
\end{aligned}$$

For the symmetric matrix $\Delta_n := V_n - K$ we have

$$\begin{aligned}
\Delta_n = \Delta_{n-1} + \frac{1}{n}&\left[\left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right) \Delta_{n-1}\right. \\
&\left. + \Delta_{n-1}\left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right)^{\intercal}\right] + \frac{1}{n}\left[R_n - R\left(1 + o\left(n^{-1}\right)\right)\right]
\end{aligned}$$

Using the operator $\mathcal{D}_n : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}$ acting as

$$\begin{aligned}
&\mathcal{D}_n X := \operatorname{col}^{-1}\left(D_n \operatorname{col} X\right), \quad X \in \mathbb{R}^{M \times N} \\
&D_n := I \otimes \left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right) + \left(A - \frac{\left(1 + o\left(n^{-1}\right)\right)}{2} I\right) \otimes I
\end{aligned}$$

$\operatorname{col} X := \left(x_{1,1}, \ldots, x_{1,N}, \ldots, x_{M,1}, \ldots, x_{M,N}\right)^{\intercal}$ - the spreading operator

$\otimes$ is the Kronecker product (see Chapter 8 in Poznyak (2008))

the last recurrence can be rewritten as

$$\Delta_n = \left(I + \frac{1}{n}\mathcal{D}_n\right) \Delta_{n-1} + \frac{1}{n}\left[R_n - R\left(1 + o\left(n^{-1}\right)\right)\right] \tag{8.117}$$

Since

$$\mathcal{D}_n \to \mathcal{D}_\infty := I \otimes \left(A - \frac{1}{2}I\right) + \left(A - \frac{1}{2}I\right) \otimes I$$

and by the assumption (8.113)

$$\operatorname{Re}\lambda_i\left(\mathcal{D}_\infty\right) \le \lambda_0 < 0$$

it follows that for any $\alpha > 0$ there exists the matrix $Q = Q^\intercal > 0$ satisfying the Lyapunov matrix equation

$$\mathcal{D}_\infty^\intercal Q + Q\mathcal{D}_\infty = -\alpha I$$

Define in the matrix space $\mathbb{R}^{N\times N}$ the following scalar product

$$\langle X, Y\rangle := (\operatorname{col}X)^\intercal Q \operatorname{col}X$$

and the norm

$$\|X\|_Q^2 := \langle X, X\rangle$$

Then for large enough n and some $0 < \alpha_0 < \alpha$ we get

$$\begin{aligned}
\left\|\left(I + \frac{1}{n}\mathcal{D}_n\right)X\right\|_Q^2 &= \left\langle\left(I + \frac{1}{n}\mathcal{D}_n\right)X, \left(I + \frac{1}{n}\mathcal{D}_n\right)X\right\rangle \\
&= \|X\|_Q^2 + \frac{1}{n}\left[\langle\mathcal{D}_n X, X\rangle + \langle X, \mathcal{D}_n X\rangle\right] + \frac{1}{n^2}\|\mathcal{D}_n X\|_Q^2 \\
&= \|X\|_Q^2 + \frac{1}{n}\left[(\operatorname{col}X)^\intercal\left(\mathcal{D}_n^\intercal Q + Q\mathcal{D}_n\right)\operatorname{col}X\right] + \frac{1}{n^2}\|\mathcal{D}_n X\|_Q^2 \\
&= \|X\|_Q^2 - \frac{\alpha + o(1)}{n}\operatorname{Tr}\left(X^\intercal X\right) + \frac{1}{n^2}\|\mathcal{D}_n X\|_Q^2 \\
&\le \|X\|_Q^2\left(1 - \frac{\alpha + o(1)}{n}\right) \le \|X\|_Q^2\left(1 - \frac{\alpha_0}{n}\right)
\end{aligned} \tag{8.118}$$

Using the inequality (8.118) show now that the process $\{\Delta_n\}_{n\ge1}$ may be '*dominated*' by the following auxiliary process:

$$\tilde{\Delta}_n = \left(1 - \frac{1}{n}\right)\tilde{\Delta}_{n-1} + \frac{1}{n}\left[R_n - R\left(1 + o\left(n^{-1}\right)\right)\right], \quad \tilde{\Delta}_0 = 0 \tag{8.119}$$

Indeed,

$$\begin{aligned}
\left\|\Delta_n - \tilde{\Delta}_n\right\|_Q &= \left\|\Delta_{n-1} - \tilde{\Delta}_{n-1} + \frac{1}{n}\mathcal{D}_n\Delta_{n-1} - \frac{1}{n}\tilde{\Delta}_{n-1}\right\|_Q \\
&= \left\|\left(I + \frac{1}{n}\mathcal{D}_n\right)\left(\Delta_{n-1} - \tilde{\Delta}_{n-1}\right) + \frac{1}{n}\left(\mathcal{D}_n + I\right)\tilde{\Delta}_{n-1}\right\|_Q \\
&\le \left\|\left(I + \frac{1}{n}\mathcal{D}_n\right)\left(\Delta_{n-1} - \tilde{\Delta}_{n-1}\right)\right\|_Q + \left\|\frac{1}{n}\left(\mathcal{D}_n + I\right)\tilde{\Delta}_{n-1}\right\|_Q \\
&\le \left\|\left(I + \frac{1}{n}\mathcal{D}_n\right)\left(\Delta_{n-1} - \tilde{\Delta}_{n-1}\right)\right\|_Q + \left\|\frac{1}{n}\left(\mathcal{D}_n + I\right)\tilde{\Delta}_{n-1}\right\|_Q
\end{aligned}$$

$$\leq \left(1-\frac{\alpha_0}{n}\right)^{1/2} \left\| \left(\Delta_{n-1}-\tilde{\Delta}_{n-1}\right) \right\|_Q + \frac{\text{Const}}{n} \left\| \tilde{\Delta}_{n-1} \right\|_Q$$
$$\leq \left(1-\frac{\alpha_0}{2n}\right) \left\| \left(\Delta_{n-1}-\tilde{\Delta}_{n-1}\right) \right\|_Q + \frac{\text{Const}}{n} \left\| \tilde{\Delta}_{n-1} \right\|_Q$$

which, in view of Lemma 16.14 in Poznyak (2008), we obtain

$$\limsup_{n\to\infty} \left\| \Delta_n - \tilde{\Delta}_n \right\|_Q \leq \text{Const} \limsup_{n\to\infty} \left\| \tilde{\Delta}_n \right\|_Q$$

This exactly means that if $\tilde{\Delta}_n \underset{n\to\infty}{\to} 0$, then also $\Delta_n \underset{n\to\infty}{\to} 0$ which is referred to as 'dominating'. But by the assumption (8.114)

$$\tilde{\Delta}_n = \left(1-\frac{1}{n}\right)\tilde{\Delta}_{n-1} + \frac{1}{n}\left[R_n - R\left(1+o\left(n^{-1}\right)\right)\right]$$
$$= \frac{1}{n}\sum_{t=1}^{n}\left[R_n - R\left(1+o\left(n^{-1}\right)\right)\right] = \frac{1}{n}\sum_{t=1}^{n} R_n - R\left(1+o\left(1\right)\right) \underset{n\to\infty}{\to} 0$$

that proves the desired result. □

8.4 Logarithmic iterative law

8.4.1 Brief survey

The *logarithmic iterative law* (LIL) for sum of independent random variables grew from early efforts of Hausdorff (1913) and Hardy and Littlewood (1914) to determine the rate of convergence in the *Law of Large Numbers* (LLN) for normal sequences. The similar law for Bernoulli variables was obtained by Khintchine (1924), and for more general sequences (independent, but not obligatory identically distributed) by Kolmogorov (1929).

In fact, the LIL's principal importance in applications is still as a rate of convergence. Consider a sequence $\{x_t\}_{t\geq 1}$ of independent identically distributed (i.i.d.) random variables with zero mean and unit variance, and let

$$\boxed{S_n := \sum_{t=1}^{n} x_t} \tag{8.120}$$

By Theorem 8.11 it follows that

$$n^{-1/2} S_n \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1)$$

although it is well known that the sequence $\left\{n^{-1/2}S_n\right\}_{n\geq 1}$ has as its set of (a.s.) limit points the whole real line. In Hartman and Wintner (1941) there was presented the version of LIL describing this behavior in much more detail, asserting that

$$\boxed{\begin{array}{c} \limsup\limits_{n\to\infty} \dfrac{1}{\sqrt{2n\ln\ln n}} S_n \overset{a.s}{=} 1 \\ \liminf\limits_{n\to\infty} \dfrac{1}{\sqrt{2n\ln\ln n}} S_n \overset{a.s.}{=} -1 \end{array}}$$

In Strassen (1964) this result was extended to obtain a rate of convergence considering the set K as a set of absolutely continuous functions $x \in C[0,1]$ with $x(0) = 0$ and whose derivatives $\dot{x}$ were such that

$$\int_{t=0}^{1} \dot{x}^2(t)\,dt \le 1 \tag{8.121}$$

Then there was shown that K is compact and the sequence

$$\left\{ \frac{\xi_n(t)}{\sqrt{2n \ln \ln n}} \right\}$$

with

$$\xi_n(t) := n^{-1/2}\{S_i + (nt - i)\,x_i\}, \quad t \in [0,1]$$

is relatively compact with a.s. limit set K. Later on, in Stout (1970b) there was obtained the martingale analog of Kolmogorov's LIL and in Stout (1970a) the analog of the Hartman–Wintner law. A further extension was provided by Hall and Heyde (1976). The comprehensive survey on different versions of LIL can be found in Bingham (1986).

Here we will present the same approach as in Hall and Heyde (1980) applying the so-called *Skorokhod representation*, which provides sharp results without the technicalities involved in estimating some tail probabilities.

8.4.2 Main result on the relative compactness

Let $\{x_t\}_{t\ge1}$ be a sequence of martingale-differences so that $\{S_n, \mathcal{F}_n\}$ with S_n given by (8.120) is a zero-mean, squared-integrable martingale and suppose that the σ-field $\mathcal{F}_n$ is generated by $\{S_1, \ldots, S_n\}$. Let $\xi_n(t)$ be the random function of $C[0,1]$ defined by

$$\boxed{\begin{aligned} \xi_n(t) &:= U_n^{-1}\left[S_i + x_{i+1}^{-2}\left(tU_n^2 - U_i^2\right)x_{i+1}\right] \\ U_i^2 &:= \sum_{t=1}^{i} x_t^2,\ U_i^2 \le tU_n^2 < U_{i+1}^2, \quad i \le n+1 \end{aligned}} \tag{8.122}$$

which is obtained by linear interpolating between the points

$$(0,0), \left(U_n^{-2}U_1^2, U_n^{-1}S_1\right), \ldots, \left(1, U_n^{-1}S_n\right)$$

Define also

$$\boxed{\zeta_n(t) := (2\ln\ln U_n)^{-1/2}\,\xi_n(t)} \tag{8.123}$$

supposing that $x_1^2 \overset{a.s.}{>} 0$ and adopting the convention that $\ln\ln x = 1$ if $0 < x \le e^e$. Analogously to Strassen (1964), we would expect that under certain conditions the sequence $\{\zeta_n(t)\}_{n\ge1}$ would be relatively compact with limit set K (8.121).

Remark 8.10. *$\{U_n\}_{n\geq 1}$ is just one possible norming sequences. Another one would be the sequence $\{V_n\}_{n\geq 1}$ with*

$$\boxed{V_n^2 := \sum_{i=1}^{n} \mathsf{E}\left\{x_i^2 \mid \mathcal{F}_{i-1}\right\}} \tag{8.124}$$

Below we shall accordingly formulate LIL for a general norming sequence $\{W_n\}_{n\geq 1}$ where W_n is a non-decreasing positive random variable such that

$$0 < W_1 \leq W_2 \leq \cdots$$

defining the continuous [0, 1] function by

$$\boxed{\begin{gathered}\zeta_n(t) := \left[\phi\left(W_n^2\right)\right]^{-1}\left[S_i + \left(W_{i+1}^2 - W_i^2\right)^{-1}\left(tW_n^2 - W_i^2\right)x_{i+1}\right]\\ \phi(t) := (2t\ln\ln t)^{1/2}\end{gathered}} \tag{8.125}$$

Theorem 8.16. **(Hall and Heyde, 1980)** *Let $\{Z_n\}_{n\geq 1}$ be a sequence of non-negative random variables and suppose that Z_n, as well as W_n, is $\mathcal{F}_{n-1}$-measurable. If*

1.

$$\left[\phi\left(W_n^2\right)\right]^{-1}\sum_{i=1}^{n}\left[x_i\chi\left(|x_i| > Z_i\right) - \mathsf{E}\left\{x_i\chi\left(|x_i| > Z_i\right) \mid \mathcal{F}_{i-1}\right\} \underset{n\to\infty}{\overset{a.s.}{\to}}\right]0 \tag{8.126}$$

2.

$$\begin{aligned}W_n^{-2}\sum_{i=1}^{n}&\left[\mathsf{E}\left\{x_i^2\chi\left(|x_i| \leq Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right.\\ &\left.- \ \mathsf{E}^2\left\{x_i\chi\left(|x_i| \leq Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right] \underset{n\to\infty}{\overset{a.s.}{\to}} 1\end{aligned} \tag{8.127}$$

3.

$$\sum_{i=1}^{\infty} W_i^{-4}\mathsf{E}\left\{x_i^4\chi\left(|x_i| \leq Z_i\right) \mid \mathcal{F}_{i-1}\right\} \overset{a.s.}{<} \infty \tag{8.128}$$

4.

$$W_{n+1}^{-1}W_n \underset{n\to\infty}{\overset{a.s.}{\to}} 1, \qquad W_n \underset{n\to\infty}{\overset{a.s.}{\to}} \infty \tag{8.129}$$

then with probability 1 the sequence $\{\zeta_n(t)\}_{n\geq 1}$ is relatively compact in $C[0,1]$ and its set of a.s. limit points coincides with K (8.121).

Proof. 1. In Strassen (1964) (see also Theorem 9.5 in this book) the approach was based on a limit law for Brownian motion which states that with probability 1 the sequence $\{\alpha_u(t)\}$ with

$$\begin{aligned}&\alpha_u(t) := (2u \ln\ln u)^{-1/2} W(ut), \quad u > e, t \in [0, 1] \\ &W(t), \quad t \geq 0 \text{ is standard Brownian motion}^2\end{aligned} \tag{8.130}$$

is relatively compact and the set of its a.s. limit points coincides with K (8.121). Below we present the auxiliary result needed for the proof of this theorem which also follows the Strassen approach via a limit theorem for Brownian motion.

Lemma 8.12. *Let $W(t), t \geq 0$ be standard Brownian motion. Let also $\{T_n\}_{n\geq1}$ and $\{W_n\}_{n\geq1}$ be non-decreasing sequences of positive random variables. Set*

$$S^*(t) := W(T_n) \quad \text{and} \quad x_n^* := S_n^* - S_{n-1}^*$$

and let $\zeta_n^*(t)$ be the random element of $C[0, 1]$ defined by

$$\begin{aligned}\zeta_n^*(t) &:= \left[\phi\left(W_n^2\right)\right]^{-1}\left[S_i^* + \left(W_{i+1}^2 - W_i^2\right)^{-1}\left(tW_n^2 - W_i^2\right)x_{i+1}^*\right] \\ &\qquad \text{for } W_i^2 \leq tW_n^2 < W_{i+1}^2, \quad i \leq n-1\end{aligned} \tag{8.131}$$

If

$$T_n \underset{n\to\infty}{\overset{a.s.}{\to}} \infty, \quad T_{n+1}^{-1}T_n \underset{n\to\infty}{\overset{a.s.}{\to}} \infty, \quad T_n^{-1}W_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} \infty \tag{8.132}$$

then with probability 1 the sequence $\left\{\zeta_n^*(t)\right\}_{n\geq1}$ is relatively compact and the set of its a.s. limit points coincides with K (8.121).

Proof of Lemma 8.12. For $t \in [0, \infty)$ define $\beta(t)$ by

$$\beta(t) := S_p^* + \left(W_{p+1}^2 - W_p^2\right)^{-1}\left(t - W_p^2\right)x_{p+1}^*$$

where

$$p = p(t) := \max\left\{i \geq 1 \mid W_i^2 \leq t\right\}$$

Then

$$\zeta_n^*(t) := \left[\phi\left(W_n^2\right)\right]^{-1}\beta\left(W_n^2 t\right)$$

[2]Standard Brownian motion $W(t)$ is a random process with independent increments which are normally distributed with $\mathcal{N}(0, 1)$ and $W(0) \overset{a.s.}{=} 0$.

By Strassen (1964) with (8.130) it is sufficient to prove that under the conditions (8.132) it follows that

$$\lim_{t\to\infty} [\phi(t)]^{-1} |\beta(t) - W(t)| \overset{a.s.}{=} 0 \tag{8.133}$$

Suppose that $\left|1 - t^{-1}s\right| \le \varepsilon < 1/2$ and let $u = t(1+\varepsilon)$. Then for large enough t

$$\begin{aligned}
&[\phi(t)]^{-1} |W(s) - W(t)| \le 2[\phi(u)]^{-1} |W(s) - W(t)| \\
&\quad = 2\left|\alpha_u\left(u^{-1}s\right) - \alpha_u\left(u^{-1}t\right)\right| \le 4 \sup_{1-2\varepsilon\le r\le 1} |\alpha_u(r) - \alpha_u(1)|
\end{aligned}$$

where α_u is as in (8.130). Let K_ε denote the set of functions distant less than ε from K. By Strassen (1964) it follows that if λ is sufficiently large, then

$$\mathsf{P}\left\{\alpha_{t(1+\varepsilon)} \in K_\varepsilon \text{ for all } t > \lambda\right\} \ge 1 - \varepsilon$$

and if $\alpha_u \in K_\varepsilon$, then for some $x \in K$

$$\sup_{1-2\varepsilon\le r\le 1} |\alpha_u(r) - \alpha_u(1)| \le 2\varepsilon + \sup_{1-2\varepsilon\le r\le 1} |x(r) - x(1)|$$

But

$$\begin{aligned}
|x(r) - x(1)| &= \left|\int_r^1 \dot{x}(t)\,dt\right| \\
&\le \left[\left(\int_r^1 1^2 dt\right)\left(\int_r^1 \dot{x}^2(t)\,dt\right)\right]^{1/2} \le (1-r)^{1/2}
\end{aligned}$$

and hence,

$$\sup_{1-2\varepsilon\le r\le 1} |\alpha_u(r) - \alpha_u(1)| \le 2\varepsilon + (2\varepsilon)^{1/2}$$

if $\alpha_u \in K_\varepsilon$. Since

$$|\beta(t) - W(t)| \le \max\left\{\left|W\left(T_{p(t)}\right) - W(t)\right|, \left|W\left(T_{p(t)+1}\right) - W(t)\right|\right\}$$

then combining the results above one can deduce that if λ is sufficiently large,

$$\begin{aligned}
&\mathsf{P}\left\{[\phi(t)]^{-1} |\beta(t) - W(t)| > 4\left(2\varepsilon + (2\varepsilon)^{1/2}\right) \text{ for some } t > \lambda\right\} \\
&\quad \ge 1 - \varepsilon - \mathsf{P}\left\{\left|1 - t^{-1}T_{p(t)}\right| > \varepsilon \text{ for some } t > \lambda\right\} \\
&\qquad - \mathsf{P}\left\{\left|1 - t^{-1}T_{p(t)+1}\right| > \varepsilon \text{ for some } t > \lambda\right\}
\end{aligned}$$

Therefore the desired result (8.133) will follow if we show that

$$t^{-1}T_{p(t)} \underset{t\to\infty}{\overset{a.s.}{\to}} 1 \quad \text{and} \quad t^{-1}T_{p(t)+1} \underset{t\to\infty}{\overset{a.s.}{\to}} 1$$

The conditions (8.132) imply that $W_{n+1}^{-2} W_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} 1$, and hence

$$1 \geq t^{-1} W_{p(t)}^2 \geq W_{p(t)+1}^{-2} W_{p(t)}^2 \underset{t\to\infty}{\overset{a.s.}{\to}} 1$$

so that $t^{-1} W_{p(t)}^2 \underset{t\to\infty}{\overset{a.s.}{\to}} 1$. Similarly, $t^{-1} W_{p(t)+1}^2 \underset{t\to\infty}{\overset{a.s.}{\to}} 1$. These properties, combined with (8.132), give the required result. □

To establish Theorem 8.16 we first define

$$\tilde{x}_i := x_i \chi\left(c_i < |x_i| \leq Z_i\right) + \frac{1}{2} x_i \chi\left(|x_i| \leq c_i\right) + \frac{1}{2} \operatorname{sign}(x_i)\, c_i \left(1 + Z_i |x_i|^{-1}\right) \chi\left(|x_i| > Z_i\right)$$

where $\{c_i\}_{i\geq 1}$ is a monotone sequence of positive constants such that $c_i \underset{i\to\infty}{\to} 0$ so fast that

$$\sum_{i=1}^{\infty} c_i < \infty, \quad \sum_{i=1}^{\infty} c_i Z_i W_i^{-2} \overset{a.s.}{<} \infty$$

If $z_i(\omega) < c_i$, let $\chi\left(c_i < |x_i| \leq Z_i\right) \equiv 0$. Set also

$$x_i^* := \tilde{x}_i - \mathsf{E}\left\{\tilde{x}_i \mid \mathcal{F}_{i-1}\right\}$$

It is easy to check that $(\tilde{x}_1, \ldots, \tilde{x}_n)$ and hence $\left(x_1^*, \ldots, x_n^*\right)$ also generate the σ-field $\mathcal{F}_n$. Define

$$S_n^* := \sum_{i=1}^{n} x_i^*$$

$$V_n^* := \sum_{i=1}^{n} \mathsf{E}\left\{\left(x_i^*\right)^2 \mid \mathcal{F}_{i-1}\right\}$$

and $\zeta_n^*(t)$ as in (8.131). Then from the definition of x_i^* it follows that

$$\left| x_i - x_i^* - \left[x_i \chi\left(|x_i| > Z_i\right) - \mathsf{E}\left\{x_i \chi\left(|x_i| > Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right]\right| \leq 3c_i$$

and so

$$\begin{aligned}
\sup_{t\in[0,1]} \left|\zeta_n(t) - \zeta_n^*(t)\right| &\leq \left[\phi\left(W_n^2\right)\right]^{-1} \sup_{1\leq j\leq n} \left|\sum_{i=1}^{j} \left(x_i - x_i^*\right)\right| \\
&\leq \left[\phi\left(W_n^2\right)\right]^{-1} \sup_{1\leq j\leq n} \left|\sum_{i=1}^{j} \left[x_i \chi\left(|x_i| > Z_i\right) - \mathsf{E}\left\{x_i \chi\left(|x_i| > Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right]\right| \\
&+ \left[\phi\left(W_n^2\right)\right]^{-1} \sum_{i=1}^{n} 3c_i \underset{n\to\infty}{\overset{a.s.}{\to}} 1
\end{aligned} \tag{8.134}$$

in view of assumption (8.126).

Next we introduce the so-called *Skorokhod representation*. By extending the original probability space if necessary, one may suppose that there exists a Brownian motion $W(t)$ and a sequence $\{T_n\}_{n\geq 1}$ of nonnegative random variables defined on this probability space such that $S_n^* = W(T_n)$ almost surely for all $n \geq 1$. Let

$$\tau_n := T_n - T_{n-1}, \quad n \geq 1, T_0 := 0$$

if $\mathcal{G}_n$ is the σ-field generated by $(x_1, \ldots, x_n)$ and $W(u)$ for $u < T_n$, then τ_n is $\mathcal{G}_n$-measurable, so that

$$\mathsf{E}\{\tau_n \mid \mathcal{G}_{n-1}\} = \mathsf{E}\left\{\left(x_n^*\right)^2 \mid \mathcal{G}_{n-1}\right\} \overset{a.s.}{=} \mathsf{E}\left\{\left(x_n^*\right)^2 \mid \mathcal{F}_{n-1}\right\}$$

and for some constant L

$$\mathsf{E}\left\{\tau_n^2 \mid \mathcal{G}_{n-1}\right\} \leq L\mathsf{E}\left\{\left(x_n^*\right)^4 \mid \mathcal{G}_{n-1}\right\} \overset{a.s.}{=} L\mathsf{E}\left\{\left(x_n^*\right)^4 \mid \mathcal{F}_{n-1}\right\}$$

In view of the assumptions (8.129), (8.134) and Lemma 8.12 it suffices to prove that

$$W_n^{-2} T_n \underset{n\to\infty}{\overset{a.s.}{\to}} 1 \tag{8.135}$$

To this end, first, we will show that

$$T_n - \left(V_n^*\right)^2 = o(W_n) \tag{8.136}$$

Taking into account that

$$\begin{aligned}
\mathsf{E}\left\{\left(x_i^*\right)^4 \mid \mathcal{F}_{i-1}\right\} &= \mathsf{E}\left\{(\tilde{x}_i)^4 \mid \mathcal{F}_{i-1}\right\} - 4\mathsf{E}\left\{(\tilde{x}_i)^3 \mid \mathcal{F}_{i-1}\right\}\mathsf{E}\{\tilde{x}_i \mid \mathcal{F}_{i-1}\} \\
&\quad + 6\mathsf{E}\left\{(\tilde{x}_i)^2 \mid \mathcal{F}_{i-1}\right\}\left[\mathsf{E}\{\tilde{x}_i \mid \mathcal{F}_{i-1}\}\right]^2 - 3\left[\mathsf{E}\{\tilde{x}_i \mid \mathcal{F}_{i-1}\}\right]^4 \\
&\leq 11\left[\mathsf{E}\{\tilde{x}_i \mid \mathcal{F}_{i-1}\}\right]^4 \leq 11\left[\mathsf{E}\{\tilde{x}_i \chi(|x_i| \leq Z_i) \mid \mathcal{F}_{i-1}\}\right]^4 + 11c_i^4
\end{aligned}$$

and in view of (8.128) it follows that

$$\sum_{i=1}^{\infty} W_i^{-4}\mathsf{E}\left\{\left(x_i^*\right)^4 \mid \mathcal{F}_{i-1}\right\} \overset{a.s.}{<} \infty$$

and therefore, by the strong law of large numbers we get

$$\sum_{i=1}^{n}\left[\tau_i - \mathsf{E}\{\tau_i \mid \mathcal{G}_{i-1}\}\right] \overset{a.s.}{=} o\left(W_n^2\right)$$

which implies (8.136). Next we have

$$\left|\mathsf{E}\left\{(\tilde{x}_i)^2 \mid \mathcal{F}_{i-1}\right\} - \mathsf{E}\left\{x_i^2 \chi(|x_i| \leq Z_i) \mid \mathcal{F}_{i-1}\right\}\right| \leq 2c_i^2 \tag{8.137}$$

$$\mathsf{E}\left\{\left(x_i^*\right)^2 \mid \mathcal{F}_{i-1}\right\} = \mathsf{E}\left\{(\tilde{x}_i)^2 \mid \mathcal{F}_{i-1}\right\} - \left[\mathsf{E}\{\tilde{x}_i \mid \mathcal{F}_{i-1}\}\right]^2 \tag{8.138}$$

$$\left|\mathsf{E}\left\{\tilde{x}_i \mid \mathcal{F}_{i-1}\right\} - \mathsf{E}\left\{x_i \chi\left(|x_i| \leq Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right| \leq 2c_i \tag{8.139}$$

The last relation (8.139) implies

$$\sum_{i=1}^{n}\left[\mathsf{E}\left\{(\tilde{x}_i)^2 \mid \mathcal{F}_{i-1}\right\} - \left[\mathsf{E}\left\{\tilde{x}_i \mid \mathcal{F}_{i-1}\right\}\right]^2\right] \overset{a.s.}{=} o\left(W_n^2\right) \tag{8.140}$$

since

$$\sum_{i=1}^{n} c_i \left|\mathsf{E}\left\{x_i \chi\left(|x_i| \leq Z_i\right) \mid \mathcal{F}_{i-1}\right\}\right| \overset{a.s.}{=} o\left(W_n^2\right)$$

by virtue of the Kronecker's lemma and in view of the fact that $\sum_{i=1}^{\infty} c_i Z_i W_i^{-2} \overset{a.s.}{<} \infty$. The conditions (8.127), (8.137), (8.138) and (8.140) now imply that

$$\left(V_n^*\right)^2 - W_n^2 \overset{a.s.}{=} o\left(W_n^2\right)$$

Combined with (8.136) this fact establishes (8.135) and completes the proof. □

Corollary 8.11. *If $\{S_n, \mathcal{F}_n\}$ is a martingale with uniformly bounded differences, i.e. $|x_n| \leq C$ for all $n \geq 1$, then Theorem 8.16 remains valid with $W_n = V_n$ (8.124) on the set $\{V_n \to \infty\}$.*

Proof. Let $Z_n := C + 1$ and

$$W_n^2 = V_n^2 := \sum_{i=1}^{n} \mathsf{E}\left\{x_i^2 \mid \mathcal{F}_{i-1}\right\}$$

The condition (8.126) and (8.127) hold trivially while (8.129) holds on the set $\{V_n \to \infty\}$. The series in (8.128) reduces to

$$\sum_{i=1}^{\infty} V_i^{-4}\mathsf{E}\left\{x_i^4 \mid \mathcal{F}_{i-1}\right\} \leq C^2 \sum_{i=1}^{\infty} V_i^{-4}\mathsf{E}\left\{x_i^2 \mid \mathcal{F}_{i-1}\right\}$$

which is finite on the set $\{V_n \to \infty\}$. This results from the following consideration: for any sequence $\{a_n\}_{n\geq 1}$ of nonnegative numbers $a_n > 0$ such that $a_n/b_n \to 0$ as $\to \infty$ with $b_n := \sum_{i=1}^{n} a_i$ and $b_0 = 0$ it follows that

$$\begin{aligned}\sum_{i=1}^{n} b_i^{-2} a_i &= \sum_{i=1}^{n} b_i^{-2}\left(b_i - b_{i-1}\right) = \sum_{i=1}^{n} b_i\left(b_i^{-2} - b_{i+1}^{-2}\right) + b_n^{-1} \\ &\leq 2\sum_{i=1}^{n}\left(b_i^{-1} - b_{i+1}^{-1}\right) + b_n^{-1} = 2b_1^{-1} + b_n^{-1}\end{aligned}$$

that completes the proof. □

Corollary 8.12. *If*

1.

$$\boxed{s_n^{-2} U_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} \eta^2} \tag{8.141}$$

where

$$s_n^2 := \mathsf{E}\left\{V_n^2\right\}$$

2. for all $\varepsilon > 0$

$$\boxed{\sum_{i=1}^{\infty} s_i^{-1} \mathsf{E}\left\{|x_i| \chi\left(|x_i| > \varepsilon s_i\right)\right\} < \infty} \tag{8.142}$$

3. for some $\delta > 0$

$$\boxed{\sum_{i=1}^{\infty} s_i^{-4} \mathsf{E}\left\{x_i^4 \chi\left(|x_i| \le \delta s_i\right)\right\} < \infty} \tag{8.143}$$

then Theorem 8.16 holds with $W_n = U_n$*, namely, the sequence*

$$\boxed{\begin{gathered} \zeta_n(t) := \left[\phi\left(U_n^2\right)\right]^{-1}\left[S_i + \left(U_{i+1}^2 - U_i^2\right)^{-1}\left(tU_n^2 - U_i^2\right)x_{i+1}\right] \\ \phi(t) := (2t \ln\ln t)^{1/2}, \quad S_i := \sum_{t=1}^{i} x_t \\ U_i^2 := \sum_{t=1}^{i} x_t^2,\, U_i^2 \le tU_n^2 < U_{i+1}^2, \quad i \le n+1 \end{gathered}} \tag{8.144}$$

is relatively compact in $C\,[0,1]$ *and the set of a.s. limit points coincides with* K *(8.121).*

Proof. First we shall establish the validity of this theorem for $W_n = U_{n-1}$ (not for $W_n = U_n$). Let $Z_j := \delta s_j$ in Theorem 8.16. By (8.141) it suffices to verify the conditions (8.126)–(8.129) of this theorem with W_n replaced by s_n (taking into account that the limit in (8.127) should be η^2 rather than 1). The condition (8.142) implies

$$\sum_{i=1}^{\infty} s_i^{-1} |x_i| \chi\left(|x_i| > \varepsilon s_i\right) \overset{a.s.}{<} \infty$$

that, by Kronecker's Lemma 8.3, leads to

$$s_n^{-1} \max_{i\le n} |x_i| \chi\left(|x_i| > \varepsilon s_i\right) \le s_n^{-1} \sum_{i=1}^{n} s_i^{-1} |x_i| \chi\left(|x_i| > \varepsilon s_i\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

Then for any $\varepsilon > 0$ one has

$$s_n^{-1}\max_{i\le n}|x_i| \le \varepsilon + s_n^{-1}\max_{i\le n}|x_i|\,\chi\left(|x_i| > \varepsilon s_i\right) \underset{n\to\infty}{\overset{a.s.}{\to}} \varepsilon$$

and hence

$$s_n^{-1}\max_{i\le n} x_i^2 \underset{n\to\infty}{\overset{a.s.}{\to}} 0 \tag{8.145}$$

Combined with (8.141) this implies

$$1 - s_{n+1}^{-2}s_n^2 = U_n^{-2}s_n^2\left(s_{n+1}^{-2}x_{n+1}^2 + s_n^{-2}U_n^2 - s_{n+1}^{-2}U_{n+1}^2\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

the condition (8.126) follows from (8.141)–(8.142) and Kronecker's Lemma 8.3, while (8.128) results from (8.141) and (8.143). To prove the validity of (8.127) note that

$$\begin{aligned}
&s_n^{-2}\sum_{i=1}^{n}\left(\mathsf{E}\left\{x_i^2\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\} - \left[\mathsf{E}\left\{x_i\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\}\right]^2\right) \\
&\quad = s_n^{-2}U_n^{-2} - s_n^{-2}\sum_{i=1}^{n} x_i^2\chi\left(|x_i| > \delta s_i\right) - s_n^{-2}\sum_{i=1}^{n}\left[\mathsf{E}\left\{x_i\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\}\right]^2 \\
&\qquad + s_n^{-2}\sum_{i=1}^{n}\left(\mathsf{E}\left\{x_i^2\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\} - x_i^2\chi\left(|x_i| \le \delta s_i\right)\right)
\end{aligned}$$

The first term in the right-hand side converges (a.s.) to η^2. The second one is dominated by

$$s_n^{-1}\max_{i\le n} x_i^2\left[s_n^{-1}\sum_{i=1}^{n}|x_i|\,\chi\left(|x_i| > \delta s_i\right)\right] \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

and the absolute values of the third one are dominated by

$$\begin{aligned}
&s_n^{-2}\sum_{i=1}^{n}\left|\mathsf{E}\left\{x_i\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\}\mathsf{E}\left\{x_i\chi\left(|x_i| > \delta s_i\right) \mid \mathcal{F}_{i-1}\right\}\right| \\
&\quad \le \delta s_n^{-1}\sum_{i=1}^{n}\mathsf{E}\left\{|x_i|\,\chi\left(|x_i| > \delta s_i\right) \mid \mathcal{F}_{i-1}\right\} \underset{n\to\infty}{\overset{a.s.}{\to}} 0
\end{aligned}$$

(by (8.142) and Kronecker's Lemma 8.3). So, to establish (8.127) it suffices to show that the series

$$\sum_{i=1}^{\infty} s_i^{-2}\left(\mathsf{E}\left\{x_i^2\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\} - x_i^2\chi\left(|x_i| \le \delta s_i\right)\right)$$

converges (a.s.). Since the terms in this series are martingale-differences, it suffices to establish the corresponding mean square series convergence which, in turn, is dominated as

$$\begin{aligned}
&\sum_{i=1}^{\infty} s_i^{-4}\mathsf{E}\left\{\left[\mathsf{E}\left\{x_i^2\chi\left(|x_i| \le \delta s_i\right) \mid \mathcal{F}_{i-1}\right\} - x_i^2\chi\left(|x_i| \le \delta s_i\right)\right]^2\right\} \\
&\quad \le \sum_{i=1}^{\infty} s_i^{-4}\mathsf{E}\left\{x_i^4\chi\left(|x_i| \le \delta s_i\right)\right\} < \infty
\end{aligned}$$

that prove (8.127) for the case $W_n = U_{n-1}$. The condition (8.145) ensures that the results remains true when $W_n = U_n$. Corollary is proven. □

8.4.3 The classical form of LIL

Here we will show that the functional form of the LIL formulated in Theorem 8.16 implies the classical form.

Theorem 8.17. (the classical form of the LIL, Stout (1970a)) *If Theorem 8.16 is valid then*

$$\boxed{\begin{array}{c}\limsup_{n\to\infty}\left[\phi\left(U_n^2\right)\right]^{-1} S_n \overset{a.s.}{=} 1 \\ \liminf_{n\to\infty}\left[\phi\left(U_n^2\right)\right]^{-1} S_n \overset{a.s.}{=} -1\end{array}} \tag{8.146}$$

Proof. First note that for any $x(\cdot) \in K$, by the Cauchy–Schwartz inequality, it follows that

$$x^2(t) = \left[\int_{s=0}^{t} \dot{x}(s)\,ds\right]^2 \le \left[\int_{s=0}^{t} 1^2 ds\right]\left[\int_{s=0}^{t} \dot{x}^2(s)\,ds\right] \le t$$

So, $|x(t)| \le \sqrt{t}$, and hence,

$$\sup_{t\in[0,1]} |x(t)| \le 1$$

that implies

$$\limsup_{n\to\infty} \sup_{t\in[0,1]} |\zeta_n(t)| \overset{a.s.}{\le} 1$$

Letting then $t = 1$ we get

$$\limsup_{n\to\infty}\left[\phi\left(U_n^2\right)\right]^{-1} |S_n| \overset{a.s.}{\le} 1$$

If we show that

$$\limsup_{n\to\infty}\left[\phi\left(U_n^2\right)\right]^{-1} |S_n| \overset{a.s.}{\ge} 1 \tag{8.147}$$

then the LIL will follow by symmetry. Let now $x(t) \equiv t$ for all $t \in [0, 1]$. Then $x \in K$ and so for all $\omega \in \Omega$ there is a subsequent $n_k = n_k(\omega)$ such that

$$\zeta_{n_k}(t)(\omega) \to x(t)$$

the convergence being in the uniform metric on [0, 1]. In particularly, since $x(1) = 1$ it follows that

$$\zeta_{n_k}(1)(\omega) \to x(1) = 1$$

that is,

$$\left[\phi\left(U_{n_k}^2\right)\right]^{-1} S_{n_k} \underset{k\to\infty}{\overset{a.s.}{\to}} 1$$

which implies (8.147). □

PART III

Continuous Time Processes

This page intentionally left blank

9 Basic Properties of Continuous Time Processes

Contents

The purpose of this chapter is to introduce and review the basic definitions and facts concerning continuous time stochastic processes that are important in understanding random effects in different engineering applications such as filtering, identification, signal processing, stochastic control and so on.

9.1 Main definitions

9.1.1 *Sample path or trajectory*

Definition 9.1. *Let* $(\Omega, \mathcal{F}, \mathsf{P})$ *be a probability space, and* $\mathcal{T}$ *be an arbitrary set. A family*

$$\{x(t, \omega)\}, \quad \omega \in \Omega, t \in \mathcal{T}$$

of random N*-vectors* $x(t, \omega)$ *is called a* ***stochastic process*** *with the index (or parameter) set* $\mathcal{T}$ *and the state space* $\mathbb{R}^N$*. Hereafter we will associate* $\mathcal{T}$ *with the time interval so that*

$$\mathcal{T} := \left[t_0, t_f\right]$$

where t_0 *is admitted to be* $(-\infty)$ *and* t_f *may be* $(+\infty)$*. In any case* $\mathcal{T} \subseteq \mathbb{R}$.

Since stochastic processes are functions of two variables, the usual notation suppresses the probability space variable, namely,

$$x(t) = x(t, \cdot)$$

denotes a random n-vector defined for each fixed $t \in \mathcal{T}$ on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$.

Definition 9.2. *For each fixed* $\omega \in \Omega$ *the* N*-vector valued function* $x(\cdot, \omega)$*, defined on* $\mathcal{T}$*, is called a* ***sample path***, ***trajectory*** *or* ***realization*** *of the process* $\{x(t, \omega)\}$.

The theory of stochastic continuous time stochastic processes relates these inherent random and sample path structures.

9.1.2 Finite-joint distributions

The individual random vectors (or, even, variables) of the given process are usually dependent. Therefore understanding the stochastic process requires knowing the uncountably many (since $\mathcal{T}$ is an interval) joint distributions of these random vectors. The collection of such joint distributions constitutes the probability law of the process.

Definition 9.3. *The functional family*

$$F_{t_1,\ldots,t_n} : \mathbb{R}^N \times \mathbb{R}^N \cdots \times \mathbb{R}^N \to [0, 1] \mid t_i \in \mathcal{T} \quad (i = 1, \ldots, n), t_i \neq t_j \ (i \neq j)$$

*is called the family of n-**finite joint** (finite-dimensional) **distributions** if $F_{t_1,\ldots,t_n}$ is defined by*

$$\begin{aligned} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) &= F_{x(t_1),\ldots,x(t_n)}(x_1, \ldots, x_n) \\ &:= \mathsf{P}\{x(t_1) \leq x_1, \ldots, x(t_n) \leq x_n\}, \quad x_i \in \mathbb{R}^N \ (i = 1, \ldots, n) \end{aligned}$$

Evidently, the n-finite joint (finite-dimensional) distribution $F_{t_1,\ldots,t_n}$ satisfies the following properties.

Claim 9.1. *For any $F_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$ the following properties hold:*

1. *it is individually nondecreasing and continuous from the right, that is, for any $i = 1, \ldots, n$ the relation $x_i < y_i$ implies*

   $$F_{t_1,\ldots,t_n}(x_1, \ldots, x_i, \ldots, x_n) \leq F_{t_1,\ldots,t_n}(x_1, \ldots, y_i, \ldots, x_n)$$

   and

   $$\lim_{y_i \to x_i + 0} F_{t_1,\ldots,t_n}(x_1, \ldots, y_i, \ldots, x_n) = F_{t_1,\ldots,t_n}(x_1, \ldots, x_i, \ldots, x_n)$$

2. $$\lim_{x_i \to -\infty, i=1,\ldots,n} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = 0$$

   and

   $$\lim_{x_i \to +\infty, i=1,\ldots,n} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = 1$$

3. *the **symmetry property**, that is,*

   $$F_{t_{i_1},\ldots,t_{i_n}}\left(x_{i_1}, \ldots, x_{i_n}\right) = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$$

   where $\{i_1, \ldots, i_n\}$ is any permutation of the integers $\{1, \ldots, n\}$
4. *the **compatibility property**, that is,*

   $$\lim_{x_i \to +\infty, m+1 \leq i \leq n} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = F_{t_1,\ldots,t_m}(x_1, \ldots, x_m)$$

One may ask, conversely, when a given collection of distribution functions uniquely (in some probabilistic sense) determines a stochastic process. The following existence or consistency of Kolmogorov's theorem (see also Theorem 1.6 in this book) answers this question.

Theorem 9.1. **(Kolmogorov, 1933)** *For any collection $\mathcal{K}$ of distributions*

$$F_{t_1,\ldots,t_n} : \mathbb{R}^N \times \mathbb{R}^N \cdots \times \mathbb{R}^N \to [0, 1]$$
$$t_i \in \left[t_0, t_f\right] \quad (i = 1, \ldots, n), t_i \neq t_j \;\; (i \neq j)$$

satisfying the properties in Claim 9.1 there exists a stochastic process

$$\{x(t, \omega)\}, \quad \omega \in \Omega, t \in \left[t_0, t_f\right]$$

whose family of finite-dimensional distributions coincides with this collection $\mathcal{K}$. Furthermore, the process is ***unique in the probabilistic sense****, namely, maybe there exists another stochastic process*

$$\{y(t, \omega)\}, \quad \omega \in \Omega, t \in \left[t_0, t_f\right]$$

but obligatory such that for each $t \in \left[t_0, t_f\right]$

$$\mathsf{P}\{x(t, \omega) \neq y(t, \omega)\} = 0 \tag{9.1}$$

Each process $\{y(t, \omega)\}$ satisfying (9.1) is referred to as a P-*equivalent version* of the process $\{x(t, \omega)\}$.

9.2 Second-order processes

The *equivalent version concept* provides the opportunity to introduce such important notions as the 'continuity' of a stochastic process meaning the existence of a *continuous equivalent version* of the given random process which, as it is well-known, has absolutely non-smooth behavior.

9.2.1 Quadratic-mean continuity

Definition 9.4. *A stochastic process $\{x(t, \omega)\}, \omega \in \Omega, t \in \left[t_0, t_f\right]$, given for each $t \in \left[t_0, t_f\right]$ on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, is said to be a* ***second-order*** *(**s.o.**) process if*

$$\boxed{\mathsf{E}\left\{\|x(t, \omega)\|^2\right\} < \infty \quad \text{for all} \quad t \in \left[t_0, t_f\right]} \tag{9.2}$$

Definition 9.5. *An s.o. stochastic process $\{x(t, \omega)\}$ is called* ***mean square continuous*** *at $t \in \left[t_0, t_f\right]$ if*

$$\boxed{\mathsf{E}\left\{\|x(t+h, \omega) - x(t, \omega)\|^2\right\} \underset{h \to 0}{\to} 0} \tag{9.3}$$

and mean square continuous on $\mathcal{T} = \left[t_0, t_f\right]$ if (9.3) holds respectively for all $t \in \mathcal{T}$.

The next result (given without the proof) represents the sufficient conditions when a stochastic process has an equivalent continuous version.

Theorem 9.2. **(Loéve, 1977)** *Suppose $\{x(t,\omega)\}$ is an s.o. stochastic process such that there exist positive constants β, C and h_0 such that*

$$\mathsf{E}\left\{\|x(t+h,\omega)-x(t,\omega)\|^2\right\} \leq C|h|^{1+\beta}$$

whenever $t, t+h \in \mathcal{T} = [t_0, t_f]$ and $|h| \leq h_0$. Then there exists an s.o. stochastic process $\{y(t,\omega)\}$ (with the same index set $\mathcal{T}$) which is mean square continuous on $\mathcal{T}$ and is stochastically equivalent to $\{x(t,\omega)\}$.

9.2.2 Separable stochastic processes

As it follows from the previous considerations a stochastic process given on a time interval constitutes an uncountable family of random variables, leading to the fact that *equivalent processes need not have the same sample paths*. This situation also implies a more general problem: *the sets associated with any uncountable subsets of random vectors (or variables)* ***may not be measurable****, that is, for a random process $\{x(t,\omega)\}$, $\omega \in \Omega, t \in [t_0, t_f]$, given for each $t \in [t_0, t_f]$ on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, the set*

$$\left\{\omega : x(t,\omega) \leq c \text{ at all } t \in [t_0, t_f], c \in \mathbb{R}\right\} = \bigcap_{t\in[t_0,t_f]} \{\omega : x(t,\omega) \leq c\} \tag{9.4}$$

may not be $\mathcal{F}$-measurable, and hence, could not be assigned a probability!

From applications and practical points of view it seems to be very important to be able to associate a probability to the event as (9.4). The concept of *separability* introduced by Doob (1953) provides a solution to this dilemma.

Definition 9.6. *Suppose there exists a countable dense subset $\mathcal{T}_{count} \subset \mathcal{T} = [t_0, t_f]$ such that for every open subinterval $\mathcal{T}_{open} \subset \mathcal{T}$ and any closed subinterval $\mathcal{X}_{closed} \subset \mathbb{R}$ the following property holds:*

$$\begin{aligned} &\left\{\omega : x(t,\omega) \in \mathcal{X}_{closed} \text{ at all } t \in \mathcal{T}_{open} \cap \mathcal{T}\right\} \\ &\quad = \left\{\omega : x(t,\omega) \in \mathcal{X}_{closed} \text{ at all } t \in \mathcal{T}_{open} \cap \mathcal{T}_{count}\right\} \end{aligned} \tag{9.5}$$

Then such stochastic process $\{x(t,\omega)\}$, $\omega \in \Omega, t \in \mathcal{T}$, given for each $t \in \mathcal{T}$ on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, is said to be ***separable***.

Remark 9.1. *If $\{x(t,\omega)\}$ is separable, taking $\mathcal{X}_{closed} := [-\infty, c]$ and $\mathcal{T}_{open} = (-\infty, \infty)$ in (9.5) the sets (9.4) can be represented as*

$$\left\{\omega : x(t,\omega) \leq c \text{ at all } t \in [t_0, t_f], c \in \mathbb{R}\right\} = \bigcap_{t\in\mathcal{T}_{count}} \{\omega : x(t,\omega) \leq c\}$$

Obviously, this set is $\mathcal{F}$-measurable since there are only countable many sets in the intersection.

It is not so difficult to see that separable equivalent processes have the same sample paths with probability 1. The considerations above imply the following statement.

Claim 9.2. *Any (scalar) stochastic process* $\{x(t,\omega)\}$ *has an equivalent extended real-valued process* $\{y(t,\omega)\}$ *which is separable.*

Another difficulty arises when we deal with a vector stochastic process, namely, even though a process is measurable on each variable separately, it may not be product-measurable, which prevents the application of Fubini's theorem on changing of the order of integration. The following statement relates the properties of mean-square continuity, separability and product-measurability.

Claim 9.3. *If a vector stochastic process* $\{x(t,\omega)\}$ *is mean-square continuous then it is product-measurable and, in addition, it is separable with respect to any countable dense subset* $\mathcal{S} \subset \mathcal{T}$.

9.2.3 Criterion of mean-square continuity

The next lemma relates the mean-square continuity property with some properties of the corresponding auto-covariance matrix function.

Lemma 9.1.

1. *An s.o. vector stochastic process* $\{x(t,\omega)\}, t \in [t_0, t_f]$ *is mean-square continuous if and only if the* ***auto-covariance matrix***

$$\boxed{\rho(t,s) := \mathsf{E}\left\{x(t,\omega)\, x^{\intercal}(s,\omega)\right\}} \tag{9.6}$$

is continuous at the diagonal point (t,t).

2. *If* $\rho(t,t)$ *is a continuous matrix on* $[t_0, t_f]$ *then* $\rho(t,s)$ *is continuous at any point* $(t,s) \in [t_0, t_f] \times [t_0, t_f]$, *i.e.,*

$$\|\rho(t+h, s+g) - \rho(t,s)\| \underset{h,g\to 0}{\to} 0$$

where the matrix norm is defined by

$$\|A\| := \lambda_{\max}^{1/2}\left(A^{\intercal}A\right)$$

Proof. First, notice that

$$\begin{aligned} &\mathsf{E}\left\{[x(t+h,\omega) - x(t,\omega)]\,[x(t+h,\omega) - x(t,\omega)]^{\intercal}\right\} \\ &\quad = \rho(t+h,t+h) - \rho(t+h,t) - \rho(t,t+h) + \rho(t,t) \end{aligned} \tag{9.7}$$

1. *Sufficiency*. Suppose that an s.o. vector stochastic process $\{x(t,\omega)\}, t \in [t_0, t_f]$ has a continuous $\rho(t,t)$. Shows that

$$\mathsf{E}\left\{\|x(t+h,\omega) - x(t,\omega)\|^2\right\} \underset{h\to 0}{\to} 0$$

Then by (9.7)

$$0 \leq \operatorname{tr}\{[\rho(t+h,t+h)-\rho(t+h,t)]-[\rho(t,t+h)-\rho(t,t)]\} \underset{h\to 0}{\to} 0$$

$$\operatorname{tr}\{[\rho(t+h,t+h)-\rho(t,t)+\rho(t,t)-\rho(t+h,t)]-[\rho(t,t+h) \\ -\rho(t,t)]\} \underset{h\to 0}{\to} 0$$

$$\lim_{h\to 0}\operatorname{tr}\left\{[\rho(t+h,t+h)-\rho(t,t)]\} = 2\lim_{h\to 0}\operatorname{tr}\{\rho(t,t+h)-\rho(t,t)\right\} = 0$$

that exactly means the continuity of $\rho(t,t)$ at any $t \in [t_0, t_f]$.
Necessity. It follows from the Cauchy–Schwarz inequality and the relation

$$\begin{aligned}\|\rho(t+h,t+h)-\rho(t,s)\| &= \Big\| \mathsf{E}\left\{x(t+h,\omega)[x(t+h,\omega)-x(t,\omega)]^{\mathsf{T}}\right\} \\ &\quad - \mathsf{E}\left\{[x(t,\omega)-x(t+h,\omega)]x^{\mathsf{T}}(t,\omega)\right\}\Big\| \\ &\leq \sqrt{\mathsf{E}\left\{\|x(t+h,\omega)\|^2\right\}}\sqrt{\mathsf{E}\left\{\|x(t+h,\omega)-x(t,\omega)\|^2\right\}} \\ &\quad + \sqrt{\mathsf{E}\left\{\|x(t,\omega)\|^2\right\}}\sqrt{\mathsf{E}\left\{\|x(t+h,\omega)-x(t,\omega)\|^2\right\}} \underset{h\to 0}{\to} 0\end{aligned}$$

2. It is sufficient to notice that

$$\begin{aligned}\|\rho(t+h,s+v)-\rho(t,s)\| &= \Big\| \mathsf{E}\left\{x(t+h,\omega)[x(s+v,\omega)-x(s,\omega)]^{\mathsf{T}}\right\} \\ &\quad - \mathsf{E}\left\{[x(t,\omega)-x(t+h,\omega)]x^{\mathsf{T}}(s,\omega)\right\}\Big\| \\ &\leq \sqrt{\mathsf{E}\left\{\|x(t+h,\omega)\|^2\right\}}\sqrt{\mathsf{E}\left\{\|x(s+v,\omega)-x(s,\omega)\|^2\right\}} \\ &\quad + \sqrt{\mathsf{E}\left\{\|x(s,\omega)\|^2\right\}}\sqrt{\mathsf{E}\left\{\|x(t+h,\omega)-x(t,\omega)\|^2\right\}} \\ &= \sqrt{\operatorname{tr}\rho(t+h,t+h)}\sqrt{\operatorname{tr}\rho(s+v,s)} + \sqrt{\operatorname{tr}\rho(s,s)}\sqrt{\operatorname{tr}\rho(t+h,t)} \underset{v,h\to 0}{\to} 0\end{aligned}$$

Lemma is proven. □

9.3 Processes with orthogonal and independent increments

9.3.1 Processes with orthogonal increments

Definition 9.7. *A second order (s.o.) vector stochastic process*

$$\{x(t,\omega)\}, \quad t \in [t_0, t_f]$$

is called ***a process with orthogonal increments*** *(o.i.) if for any non-overlapping open intervals* $(s,t) \in [t_0,t_f]$ *and* $(s',t') \in [t_0,t_f]$ *such that*

$$\boxed{(s,t)\cap(s',t') = \varnothing} \tag{9.8}$$

the following property holds:

$$\boxed{\mathsf{E}\left\{[x(t,\omega)-x(s,\omega)]\left[x\left(t',\omega\right)-x\left(s',\omega\right)\right]^{\mathsf{T}}\right\}=0} \tag{9.9}$$

Below, for simplicity we will consider scalar random processes.

Lemma 9.2. *Let* $\{x(t,\omega)\}, t\in\left[t_0,t_f\right]$ *be an o.i.-scalar random process with*

$$\boxed{x(t_0,\omega)\overset{a.s.}{=}0} \tag{9.10}$$

and

$$\boxed{\sigma^2(t):=\rho(t,t)=\mathsf{E}\left\{x^2(t,\omega)\right\}} \tag{9.11}$$

Then

1. $\sigma^2(t)$ *is a non-decreasing function of* t
2. for any $s,t\in\left[t_0,t_f\right]$

$$\boxed{\rho(t,s)=\mathsf{E}\{x(t,\omega)\,x(s,\omega)\}=\sigma^2(t\wedge s)} \tag{9.12}$$

where

$$\boxed{t\wedge s:=\min\{t,s\}=\frac{1}{2}(t+s-|t-s|)} \tag{9.13}$$

Proof. Take $t_0\le s<t\le t_f$. Then by (9.10)

$$x(t,\omega)=[x(t,\omega)-x(s,\omega)]+[x(s,\omega)-x(t_0,\omega)]$$

and by (9.9)

$$\sigma^2(t)=\mathsf{E}\left\{[x(t,\omega)-x(s,\omega)]^2\right\}+\mathsf{E}\left\{x^2(s,\omega)\right\}\ge\sigma^2(s)$$

that proves 1. Also

$$\begin{aligned}\rho(t,s)&=\mathsf{E}\{x(s,\omega)[x(t,\omega)-x(s,\omega)]\}+\mathsf{E}\left\{x^2(s,\omega)\right\}\\&=\mathsf{E}\{[x(s,\omega)-x(t_0,\omega)][x(t,\omega)-x(s,\omega)]\}+\mathsf{E}\left\{x^2(s,\omega)\right\}=\sigma^2(s)\end{aligned}$$

Analogously, if $t_0\le t<s\le t_f$ then

$$\rho(t,s)=\sigma^2(t)$$

Lemma is proven. □

9.3.2 Processes with stationary orthogonal increments

Next, we consider a specific subclass of *second order processes with orthogonal increments* (s.o.o.i.) which will attract our attention during the rest of the book.

Definition 9.8. *An s.o.o.i. vector random process*

$$\boxed{\{x(t,\omega)\}, t \in [t_0, t_f], x(t_0,\omega) \overset{a.s.}{=} 0} \tag{9.14}$$

is said to have ***stationary*** *(in the wide sense)* ***increments*** *if the mathematical expectation and auto-covariance matrix of its increments depend only on the time distance $|t-s|$, that is, if for any $s, t \in [t_0, t_f]$ and any h such that $t+h, s+h \in [t_0, t_f]$ it follows that*

$$\boxed{\begin{aligned} &\mathsf{E}\{x(t,\omega) - x(s,\omega)\} = \mathsf{E}\{x(t+h,\omega) - x(s+h,\omega)\} \\ &\rho(t,s) := \mathsf{E}\left\{[x(t,\omega) - x(s,\omega)][x(t,\omega) - x(s,\omega)]^{\mathsf{T}}\right\} \\ &\quad = \mathsf{E}\left\{[x(t+h,\omega) - x(s+h,\omega)][x(t+h,\omega) - x(s+h,\omega)]^{\mathsf{T}}\right\} \\ &\quad = \rho(t-s) \end{aligned}} \tag{9.15}$$

Lemma 9.3. *If $\{x(t,\omega)\}, t \in [t_0, t_f], x(t_0,\omega) \overset{a.s.}{=} 0$ is an s.o.o.i. scalar random process with* ***stationary increments*** *then for any $s, t, r, (t+r) \in [t_0, t_f]$*

(a)

$$\boxed{\sigma^2(t) = \sigma^2(t-s) + \sigma^2(s)} \tag{9.16}$$

(b)

$$\boxed{\sigma^2(t+r) = \sigma^2(t) + \sigma^2(r)} \tag{9.17}$$

(c)

$$\boxed{\sigma^2(t) = \sigma^2 t, \quad \sigma^2 := \sigma^2(1)} \tag{9.18}$$

Proof. The property (a) results from the relation

$$\begin{aligned} \sigma^2(t) &= \mathsf{E}\left\{[x(t,\omega) - x(s,\omega)]^2\right\} + \mathsf{E}\left\{x^2(s,\omega)\right\} \\ &= \rho(t,s) + \sigma^2(s) = \rho(t-s) + \sigma^2(s) \end{aligned}$$

The property (b) follows from (a) if we put there $t \to t+r$ and $s \to r$. In turn, (c) results from (b) if we into account that for $r = 1$

$$\sigma^2(t+1) = \sigma^2(t) + \sigma^2(1) = \cdots = (t+1)\sigma^2(1)$$

Lemma is proven. □

Lemma 9.4. (Criterion of increment stationarity) *An s.o.o.i. scalar random process*

$$\{x(t,\omega)\}, \quad t \in [t_0, t_f], x(t_0,\omega) \overset{a.s.}{=} 0$$

is a process with ***stationary increments*** *if and only if*

$$\boxed{\rho(t,s) = \sigma^2 \cdot (t \wedge s)} \tag{9.19}$$

for some nonnegative constant σ^2.

Proof. Necessity. It follows from the property (9.12) in Lemma 9.2 and the property (9.18) in Lemma 9.3.
Sufficiency. It results from the following relation valid for any $s \le t \le s' \le t'$:

$$\begin{aligned}
&\mathsf{E}\left\{\left[x\left(t',\omega\right) - x\left(s',\omega\right)\right][x(t,\omega) - x(s,\omega)]\right\} = \mathsf{E}\left\{x\left(t',\omega\right)x(t,\omega)\right\} \\
&\qquad - \mathsf{E}\left\{x\left(s',\omega\right)x(t,\omega)\right\} - \mathsf{E}\left\{x\left(t',\omega\right)x(s,\omega)\right\} + \mathsf{E}\left\{x\left(s',\omega\right)x(s,\omega)\right\} \\
&\quad = \sigma^2 \cdot \left[\left(t' \wedge t\right) - \left(s' \wedge t\right) - \left(t' \wedge s\right) + \left(s' \wedge s\right)\right] \\
&\quad = \sigma^2 \cdot [t - t - s + s] = 0
\end{aligned}$$

Lemma is proven. □

9.3.3 Processes with independent increments

Definition 9.9. *A scalar random process*

$$\{x(t,\omega)\}, \quad t \in [t_0, t_f], \quad x(t_0,\omega) \overset{a.s.}{=} 0$$

is said to be a process with ***independent increments*** *if for any integer* $n = 0, 1, ..$ *and all* $0 = t_0 < t_1 < \cdots < t_n$ *the following random variables*

$$x(t_1,\omega) - x(t_0,\omega), x(t_2,\omega) - x(t_1,\omega), \ldots, x(t_n,\omega) - x(t_{n-1},\omega)$$

represent an independent collection.

The theorem below gives the criterion (the necessary and sufficient conditions) for a random process to be a process with independent increments.

Theorem 9.3. *A scalar random process* $\{x(t,\omega)\}$, $t \in [t_0, t_f]$, $x(t_0,\omega) \overset{a.s.}{=} 0$ *is a process with independent increments* ***if and only if*** *the following property for the corresponding characteristic functions*

$$\boxed{\varphi_{x_t}(v) := \int_{\mathbb{R}^n} e^{i(v,x_t)} dF_{x_t}(v) = \mathsf{E}\{e^{i(v,x_t)}\}} \tag{9.20}$$

holds:

$$\boxed{\varphi_{x_t - x_s}(v) = \varphi_{x_t - x_u}(v)\,\varphi_{x_u - x_s}(v)} \tag{9.21}$$

for any $0 \le s < u < t < \infty$ *and any* $v \in \mathbb{R}$.

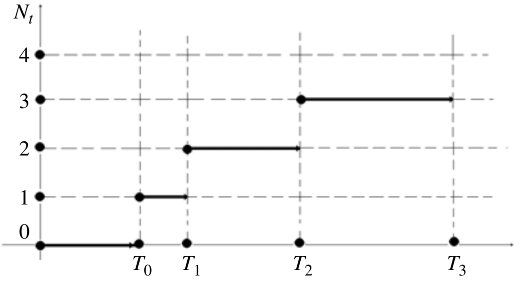

Fig. 9.1. Poisson process.

Proof. (main ideas).
(a) *Necessity*. The property (9.21) obviously results from the fact (see Lemma 5.2) that the characteristic function of a sum of independent random variables is equal to the product of corresponding characteristic functions.
(b) *Sufficiency*. Assume that (9.21) holds for some random variables $\xi(t, s, \omega)$, namely,

$$\varphi_{\xi_{t,s}}(v) = \varphi_{\xi_{t,u}}(v)\, \varphi_{\xi_{u,s}}(v)$$

Define by recurrence

$$x\left(t_j, \omega\right) = x\left(t_{j-1}, \omega\right) + \xi\left(t_j, t_{j-1}, \omega\right)$$

taking $x(t_0, \omega) \overset{a.s.}{=} 0$. Then $x(t_n, \omega)$ evidently may be defined as

$$x(t_n, \omega) = \sum_{j=1}^{n} \left[x\left(t_j, \omega\right) - x\left(t_{j-1}, \omega\right)\right] = \sum_{j=1}^{n} \xi\left(t_j, t_{j-1}, \omega\right)$$

that completes the proof. □

9.3.4 Poisson process

Let $\{s_i\}_{i \geq 0}$ be a sequence of independent identically distributed (i.i.d.) random variables whose distributions for all i are exponential and given by

$$\boxed{\mathsf{P}\{s_i > t\} = e^{-\lambda t}, \quad \lambda > 0, t \geq 0} \tag{9.22}$$

Let

$$\boxed{T_n := \sum_{i=0}^{n} s_i} \tag{9.23}$$

Definition 9.10. *The continuous-time process $\{N_t\}_{t \geq 0}$ defined by*

$$\boxed{N_t = n \quad if \quad T_{n-1} \leq t < T_n} \tag{9.24}$$

*is called the **Poisson process** with the rate λ (see Fig. 9.1).*

The next lemma states the form of the distribution for the Poisson process.

Lemma 9.5. (on the Poisson distribution) *For the Poisson process with the rate* λ*,*

$$\boxed{p_n(t) := \mathsf{P}\{N_t = n\} = \frac{(\lambda t)^n}{n!}e^{-\lambda t}} \tag{9.25}$$

Proof. The direct calculation gives

$$p_0(t) = \mathsf{P}\{N_t = 0\} = \mathsf{P}\{t < s_0\} = e^{-\lambda t}$$

and by (1.68)

$$\begin{aligned}
p_1(t) &= \mathsf{P}\{N_t = 1\} = \mathsf{P}\{s_0 \le t < s_0 + s_1\} \\
&= \int_{v=0}^{\infty} \mathsf{P}\{s_0 \le t < s_0 + s_1 \mid s_0 = v\}\, d\mathsf{P}\{s_0 \le v\} \\
&= \int_{v=0}^{\infty} \mathsf{P}\{0 \le t - v < s_1\}\, d\mathsf{P}\{s_0 \le v\} \\
&= \int_{v=0}^{t} e^{-\lambda(t-v)} d\left(1 - e^{-\lambda v}\right) = \int_{v=0}^{t} e^{-\lambda(t-v)} \lambda e^{-\lambda v} dv \\
&= \lambda \int_{v=0}^{t} e^{-\lambda t} dv = \lambda t e^{-\lambda t}
\end{aligned}$$

Then let us use the induction method assuming that (9.25) is valid for some n. We show that it remains to be true for $(n+1)$. First, notice that

$$\begin{aligned}
p_{n+1}(t) &= \mathsf{P}\{N_t = n+1\} = \mathsf{P}\{T_n \le t < T_{n+1}\} \\
&= \int_{v=0}^{\infty} \mathsf{P}\{T_n \le t < T_n + s_{n+1} \mid T_n = v\}\, d\mathsf{P}\{T_n \le v\} \\
&= \int_{v=0}^{t} \mathsf{P}\{0 \le t - v < s_{n+1} \mid T_n = v\}\, d\mathsf{P}\{T_n \le v\} \\
&= \int_{v=0}^{t} \mathsf{P}\{0 \le t - v < s_{n+1}\}\, d\mathsf{P}\{T_n \le v\} \\
&= \int_{v=0}^{t} e^{-\lambda(t-v)} \frac{d\mathsf{P}\{T_n \le v\}}{dv} dv
\end{aligned} \tag{9.26}$$

But, by the Bayes formula (1.69) and in view of (9.23) and by independency of s_n,

$$\begin{aligned}
\mathsf{P}\{T_n \le v\} &= \int_{w=0}^{v} \mathsf{P}\{T_{n-1} + s_n \le v \mid T_{n-1} = w\}\, d\mathsf{P}\{T_{n-1} \le w\} \\
&= \int_{w=0}^{v} \mathsf{P}\{s_n \le v - w \mid T_{n-1} = w\}\, d\mathsf{P}\{T_{n-1} \le w\} \\
&= \int_{w=0}^{v} \mathsf{P}\{s_n \le v - w\}\, d\mathsf{P}\{T_{n-1} \le w\}
\end{aligned}$$

that, by the assumption that (9.25) holds, implies

$$\begin{aligned}
&\frac{d}{dv}\mathsf{P}\{T_n \le v\} = \underbrace{\mathsf{P}\{s_n \le 0\}}_{0}\frac{d}{dv}\mathsf{P}\{T_{n-1} \le 0\} \\
&+ \int_{w=0}^{v} \frac{d}{dv}\left[1 - e^{-\lambda(v-w)}\right] d\mathsf{P}\{T_{n-1} \le w\} = \lambda p_n(v) = \lambda \frac{(\lambda v)^n}{n!} e^{-\lambda v}
\end{aligned} \tag{9.27}$$

Substituting (9.27) into (9.26) leads to the following relation:

$$\begin{aligned}
p_{n+1}(t) &= \int_{v=0}^{t} e^{-\lambda(t-v)} \frac{d\mathsf{P}\{T_n \le v\}}{dv} dv = \int_{v=0}^{t} e^{-\lambda(t-v)} \lambda \frac{(\lambda v)^n}{n!} e^{-\lambda v} dv \\
&= \frac{\lambda^{n+1}}{n!} e^{-\lambda t} \int_{v=0}^{t} v^n dv = \frac{\lambda^{n+1}}{n!} e^{-\lambda t} \frac{t^{n+1}}{n+1} = \frac{(\lambda t)^{n+1}}{(n+1)!} e^{-\lambda t}
\end{aligned}$$

that completes the proof. □

Lemma 9.6. *The Poisson process $\{N_t\}_{t \ge 0}$ (9.24) is an s.o. process such that*

$$\boxed{\begin{aligned}
\mathsf{E}\{N_t\} &= e^{-\lambda t} \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} = \lambda t \\
\mathsf{E}\left\{N_t^2\right\} &= e^{-\lambda t} \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} = \lambda t (\lambda t + 1)
\end{aligned}} \tag{9.28}$$

Proof. Show that the Poisson process (9.24) is the second order process (s.o.), i.e., it has a bounded second moment for any fixed t. It follows from the following calculation:

$$\mathsf{E}\left\{N_t^2\right\} = \sum_{n=0}^{\infty} n^2 \mathsf{P}\{N_t = n\} = \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} e^{-\lambda t} = e^{-\lambda t} \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!}$$

$$e^{\lambda t} = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!}$$

$$\frac{d}{dt} e^{\lambda t} = \lambda e^{\lambda t} = \lambda \sum_{n=0}^{\infty} n \frac{(\lambda t)^{n-1}}{n!} = \frac{1}{t} \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!}$$

that leads to the first formula in (9.28). Then we have

$$\frac{d^2}{dt^2} e^{\lambda t} = \lambda^2 e^{\lambda t} = \lambda^2 \sum_{n=0}^{\infty} n\,(n-1)\,\frac{(\lambda t)^{n-2}}{n!}$$
$$= \frac{1}{t^2} \sum_{n=0}^{\infty} n\,(n-1)\,\frac{(\lambda t)^n}{n!} = \frac{1}{t^2} \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} - \frac{1}{t^2} \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!}$$

implying

$$\sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} = (\lambda t)\, e^{\lambda t}$$
$$\sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^{n-1}}{n!} = (\lambda t)^2\, e^{\lambda t} + \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!}$$
$$= (\lambda t)\,[(\lambda t) + 1]\, e^{\lambda t}$$

and, as the result,

$$\mathsf{E}\left\{N_t^2\right\} = e^{-\lambda t} \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} = \lambda t\,(\lambda t + 1)$$ □

Lemma 9.7. *The Poisson process* $\{N_t\}_{t\geq 0}$ *(9.24) is an s.o. process with* ***independent increments***.

Proof. (a) To prove this lemma it is sufficient to show that

$$\mathsf{P}\left\{N_{t_1} - N_{t_0} = k_1; N_{t_2} - N_{t_1} = k_2; \cdots; N_{t_n} - N_{t_{n-1}} = k_n\right\}$$
$$= \prod_{j=1}^{n} q_{k_j}\left(\lambda\left(t_j - t_{j-1}\right)\right) \tag{9.29}$$

for all $n \geq 2$, $0 = t_0 < t_1 < \cdots < t_n$ and $k_1, k_2, \ldots, k_n > 0$ where

$$q_k(\mu) = \begin{cases} \dfrac{\mu^k}{k!} e^{-\mu} & \text{for } \mu > 0 \\ 0 & \text{for } \mu < 0 \\ \delta_{k,0} & \text{for } \mu = 0 \end{cases} \quad (k = 0, 1, \ldots)$$

Indeed, assuming the validity of (9.29), one has

$$\mathsf{P}\left\{N_{t_2} - N_{t_1} = k_2\right\} = \sum_{k_1=0}^{\infty} \mathsf{P}\left\{N_{t_2} - N_{t_1} = k_2 \mid N_{t_1} = k_1\right\} \mathsf{P}\left\{N_{t_1} = k_1\right\}$$

$$= \sum_{k_1=0}^{\infty} \mathsf{P}\left\{N_{t_2} - N_{t_1} = k_2\right\} \mathsf{P}\left\{N_{t_1} = k_1\right\}$$
$$= q_{k_2}\left(\lambda\left(t_2 - t_1\right)\right) \sum_{k_1=0}^{\infty} q_{k_1}\left(\lambda\left(t_1\right)\right) = q_{k_2}\left(\lambda\left(t_2 - t_1\right)\right)$$

So, it is sufficient to check the property (9.29). Denote the event in the right-hand side of (9.29) by A, that is,

$$A := \left\{\omega : N_{t_1} - N_{t_0} = k_1; N_{t_2} - N_{t_1} = k_2; \cdots; N_{t_n} - N_{t_{n-1}} = k_n\right\}$$

(b) If $k_1 = k_2 = \ldots = k_n = 0$ then

$$\mathsf{P}\{A\} = \mathsf{P}\{s_1 > t_n\} = e^{-\lambda t_n} = e^{-\lambda t_1} e^{-\lambda(t_2 - t_1)} \cdots e^{-\lambda\left(t_n - t_{n-1}\right)}$$

and the property (9.29) holds. Next, let us consider an induction. Let for some $m \geq 2$ we have

$$k_1 = k_2 = \ldots = k_n = 0, \quad k_m \geq 1, k_j \geq 0 \text{ for } m < j \leq n$$

Then

$$A_0 := \left\{\omega : t_{m-1} < s_1 \leq t_m, T_{k_m} \leq t_m, T_{k_m+1} > t_m, \right.$$
$$\left. \ldots, T_{k_m+\cdots+k_n} \leq t_n, T_{k_m+\cdots+k_n} > t_n\right\}$$

and

$$\mathsf{P}\{A_0\} = \mathsf{E}\{\chi(\omega \in A_0)\} = \mathsf{E}\{\mathsf{E}\{\chi(\omega \in A_0) \mid s_1\}\} \tag{9.30}$$

Therefore, we have

$$\mathsf{E}\{\chi(\omega \in A_0) \mid s_1 = z\}$$
$$= \chi(t_{m-1} < z \leq t_m) \cdot \mathsf{P}\left\{\omega : T_{k_m - 1} \leq t_m - z, T_{k_m} > t_m - z, \ldots, \right.$$
$$\left. T_{k_m+\cdots+k_n-1} \leq t_n - z, T_{k_m+\cdots+k_n-1} > t_n - z\right\}$$
$$= \chi(t_{m-1} < z \leq t_m)$$
$$\cdot \mathsf{P}\left\{\omega : N_{t_m - z} = k_m - 1, N_{t_m+1-z} - N_{t_m - z} = k_{m+1}, \right.$$
$$\left. \ldots, N_{t_n - z} - N_{t_m - 1 - z} = k_{m+1}\right\}$$
$$= \chi(t_{m-1} < z \leq t_m)\, q_{k_m - 1}\left(\lambda\left(t_m - z\right)\right) \cdot \prod_{j=m+1}^{n} q_{k_j}\left(\lambda\left[\left(t_j - z\right) - \left(t_{j-1} - z\right)\right]\right)$$

and, by (9.30), it follows that

$$\mathsf{P}\{A\} = \mathsf{E}\left\{\chi(t_{m-1} < s_1 \leq t_m)\, q_{k_m - 1}\left(\lambda\left(t_m - s_1\right)\right)\right.$$
$$\left. \cdot \prod_{j=m+1}^{n} q_{k_j}\left(\lambda\left[\left(t_j - s_1\right) - \left(t_{j-1} - s_1\right)\right]\right)\right\}$$

$$= \int_{z=t_m-1}^{t_m} \lambda e^{-\lambda z} \frac{[\lambda (t_m - z)]^{k_m-1}}{(k_m-1)!} e^{-\lambda(t_m-z)} dz \prod_{j=m+1}^{n} q_{k_j} \left(\lambda \left[t_j - t_{j-1}\right]\right)$$

$$= \lambda \frac{e^{-\lambda t_m}}{(k_m-1)!} \int_{z=t_m-1}^{t_m} [\lambda (t_m - z)]^{k_m-1} dz \prod_{j=m+1}^{n} q_{k_j} \left(\lambda \left[t_j - t_{j-1}\right]\right)$$

$$= e^{-\lambda t_1} e^{-\lambda(t_2-t_1)} \cdots e^{-\lambda(t_m-t_{m-1})} \frac{\left[\lambda (t_m - t_{m-1})\right]^{k_m}}{k_m!} \prod_{j=m+1}^{n} q_{k_j} \left(\lambda \left[t_j - t_{j-1}\right]\right)$$

So, now the property (9.29) is proven for the considered case.
(c) Let now $k_1 \geq 1$. Analogously, for

$$\begin{aligned} A_{k_1} := \{\omega : 0 \leq z \leq t_1, z + s_{k_1+1} > t_1, z + s_{k_1+1} + \cdots s_{k_1+k_2} \leq t_2, \\ z + s_{k_1+1} + \cdots s_{k_1+k_2+1} > t_2, \ldots, z + s_{k_1+1} + \cdots s_{k_1+k_2+\cdots k_n} \leq t_n, \\ z + s_{k_1+1} + \cdots s_{k_1+k_2+\cdots k_n+1} > t_n\} \end{aligned}$$

it follows that

$$\begin{aligned} &\mathsf{E}\left\{\chi\left(\omega \in A_{k_1}\right) \mid T_{k_1} = z\right\} \\ &\quad = \chi\left(0 \leq z \leq t_1\right) \cdot \mathsf{P}\left\{\omega : \ s_1 > t_1 - z, \ldots, T_{k_2} \leq t_2 - z, \ldots, T_{k_2+1} > t_2 - z,\right. \\ &\qquad \left. \ldots T_{k_2+\cdots+k_n} \leq t_n - z, T_{k_m+\cdots+k_n+1} > t_n - z\right\} = \chi\left(0 \leq z \leq t_1\right) \\ &\qquad \cdot \mathsf{P}\left\{\omega : \ N_{t_1-z} = 0, N_{t_2-z} = k_2, \ldots, N_{t_n-z} = k_2 + \cdots + k_n\right\} \\ &\quad = \chi\left(0 \leq z \leq t_1\right) e^{-\lambda(t_1-z)} \prod_{j=2}^{n} q_{k_j} \left(\lambda \left[t_j - t_{j-1}\right]\right) \end{aligned}$$

So, now we have

$$\begin{aligned} \mathsf{P}\left\{A_{k_1}\right\} &= \mathsf{E}\left\{\chi\left(\omega \in A_{k_1}\right)\right\} = \mathsf{E}\left\{\mathsf{E}\left\{\chi\left(\omega \in A_{k_1}\right) \mid s_1\right\}\right\} \\ &= \mathsf{E}\left\{\chi\left(0 \leq s_1 \leq t_1\right) e^{-\lambda(t_1-s_1)}\right\} \\ &= \int_{z=0}^{t_1} \lambda \frac{(\lambda z)^{k_1-1}}{(k_1-1)!} e^{-\lambda z} e^{-\lambda(t_1-z)} dz = e^{-\lambda t_1} \int_{z=0}^{t_1} \lambda \frac{(\lambda z)^{k_1-1}}{(k_1-1)!} dz \\ &= \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda t_1} \end{aligned}$$

that completes the proof. □

Theorem 9.4. (on the asymptotic behavior) *For the Poisson continuous-time process $\{N_t\}_{t\geq 0}$ (9.24) with the rate λ the following property holds:*

$$\boxed{\frac{1}{t} N_t \underset{t\to\infty}{\overset{a.s.}{\to}} \lambda} \tag{9.31}$$

Proof. **(the main scheme)** By the property (9.28) to prove (9.28) it is equivalent to prove

$$\frac{1}{t}\overset{\circ}{N}_t \underset{t\to\infty}{\overset{a.s.}{\to}} 0, \quad \overset{\circ}{N}_t := N_t - \mathsf{E}\{N_t\}$$

To do that it is necessary and sufficient to show that (see (6.57))

$$\mathsf{P}\left\{\sup_{k\geq n}\frac{1}{t_k}\left|\overset{\circ}{N}_{t_k}\right| \geq \varepsilon\right\} \underset{n\to\infty}{\to} 0$$

for any subsequence $\{t_k\}_{k\geq 1}$ taking into account that $\{N_t\}_{t\geq 0}$ can be represented as

$$N_t = \max\left\{m : T_m := \sum_{i=0}^{m} s_i \leq t\right\}$$

□

9.3.5 Wiener process or Brownian motion

Definition 9.11.

1. ***Wiener process*** *or* ***Brownian motion*** *(BM) is a zero-mean s.o. scalar process*

$$\{W_t(\omega)\}_{t\geq 0}, \quad W_0(\omega) \overset{a.s.}{=} 0$$

with ***stationary normal independent increments****.*

2. *If, additionally,*

$$\mathsf{E}\left\{W_1^2(\omega)\right\} = 1 \tag{9.32}$$

then $\{W_t(\omega)\}_{t\geq 0}$ *is a* ***standard Brownian motion*** *(SMB).*

Proposition 9.1.

1. *As it follows from Lemma 9.4, for an SBM*

$$\boxed{\rho(t,s) = \mathsf{E}\{W_t(\omega)\, W_s(\omega)\} = \sigma^2 \cdot (t \wedge s) = t \wedge s} \tag{9.33}$$

since

$$\sigma^2 = \rho(1,1) = 1$$

It is also true that (even $\sigma^2 = \rho(1,1) \neq 1$*)*

$$\sigma_t^2 = \rho(t,t) = \sigma^2 t \tag{9.34}$$

So, BM

$$\left\{\overset{\circ}{W}_t(\omega)\right\}_{t\geq 0}, \quad \overset{\circ}{W}_t(\omega) := \frac{1}{\sigma}W_t(\omega) \tag{9.35}$$

is always SBM.

2. *From the normality of BM's increments one has*

$$\boxed{\frac{d}{dx}\mathsf{P}\{W_t(\omega) \le x\} = \frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{x^2}{2\sigma^2 t}\right\}} \tag{9.36}$$

Remark 9.2. *Being a process with independent increments, a Wiener process or a Brownian motion* ***(BM) is also a process with orthogonal increments****.*

Proof. The independency of increments, taken for $(s,t) \cap (s',t') = \varnothing$, implies

$$\begin{aligned}
&\mathsf{E}\left\{[W_t(\omega) - W_s(\omega)][W_{t'}(\omega) - W_{s'}(\omega)]^\mathsf{T}\right\} = 0 \\
&\mathsf{E}\{[W_t(\omega) - W_s(\omega)]\}\,\mathsf{E}\left\{[W_{t'}(\omega) - W_{s'}(\omega)]^\mathsf{T}\right\} \\
&\quad = [\mathsf{E}\{W_t(\omega)\} - \mathsf{E}\{W_s(\omega)\}][\mathsf{E}\{W_{t'}(\omega)\} - \mathsf{E}\{W_{s'}(\omega)\}]^\mathsf{T} = 0
\end{aligned}$$

which proves this remark. □

9.3.6 An invariance principle and LIL for Brownian motion

Let $\{W_t\}_{t\ge 0}$ be a standard Brownian motion (SBM) in $\mathbb{R}^k$. Below we will omit the argument ω. Define

$$\boxed{\zeta_n(t) := (2n\ln\ln n)^{-1/2}\,W_{nt}} \tag{9.37}$$

for $t \in [0,1]$ and $n \ge 3$. Let $C^k[0,1]$ be the Banach space of continuous maps from $[0,1]$ to $\mathbb{R}^k$ endowed with the supremum norm $\|\cdot\|$, using the usual Euclidian norm in $\mathbb{R}^k$. Then $\zeta_n(t)$ is a random variable with values in $C^k[0,1]$.

Definition 9.12. *Define the set K as the set of absolutely continuous function $x(\cdot) \in C^k[0,1]$ such that $x(0) = 0$ and*

$$\boxed{\int_{t=0}^{1} \|\dot{x}(t)\|^2\,dt \le 1} \tag{9.38}$$

(here $\dot{x}(t)$ may be determined almost everywhere with respect to the Lebesgue measure).

Remark 9.3. *K is a* ***norm-compact*** *subset of C. In fact, for any $a \le b$*

$$\begin{aligned}
\|x(b) - x(a)\| &= \left\|\int_{t=a}^{b} \dot{x}(t)\,dt\right\| \\
&\le \left(\int_{t=a}^{b} 1^2 dt \int_{t=a}^{b} \|\dot{x}(t)\|^2\,dt\right)^{1/2} \le \sqrt{b-a}
\end{aligned}$$

so that K is relatively norm-compact.

Remark 9.4. *Considering $x(t)$ as the motion of a mass point with the mass 2 from time 0 to time 1, then K consists of those motions for which the mean kinetic energy is less than (or equal to) 1.*

Theorem 9.5. **(Strassen, 1964)** *With probability 1 the sequence $\{\zeta_n(\cdot)\}_{n\geq 3}$ is relatively norm-compact and the set of its norms-limit points coincides with K.*

Proof.

1. Let $\varepsilon > 0$, K_ε be the set of all points in $C^k[0,1]$ which have a distance less than ε from K. Then for any positive integer m and $r > 1$ it follows that

$$\mathsf{P}\{\zeta_n(\cdot) \notin K_\varepsilon\} \leq (I) + (II)$$
$$(I) := \mathsf{P}\left\{2m\sum_{i=1}^{2m}\left(\zeta_n\left(\frac{i}{2m}\right) - \zeta_n\left(\frac{i-1}{2m}\right)\right)^2 > r^2\right\}$$
$$(II) := \mathsf{P}\left\{\left[2m\sum_{i=1}^{2m}\left(\zeta_n\left(\frac{i}{2m}\right) - \zeta_n\left(\frac{i-1}{2m}\right)\right)^2 \leq r^2\right] \vee [\zeta_n(\cdot) \notin K_\varepsilon]\right\}$$

But for large enough n

$$(I) = \mathsf{P}\left\{\chi^2_{2mk} > 2r^2 \ln\ln n\right\}$$
$$= \frac{1}{\Gamma(mk)} \int_{2r^2 \ln\ln n}^{\infty} t^{mk-1}e^{-t}dt \simeq \frac{\left(2r^2 \ln\ln n\right)^{mk-1} e^{-r^2 \ln\ln n}}{\Gamma(mk)}$$

(recall that k is the dimension of the Brownian motion W). Let also η_n be the random vector with values in $C^k[0,1]$ obtained by linearly interpolating the points $\zeta_n\left(\frac{i}{2m}\right)$ at $\frac{i}{2m}(i = 1, \ldots, 2m)$. Then $2m\sum_{i=1}^{2m}\left(\zeta_n\left(\frac{i}{2m}\right) - \zeta_n\left(\frac{i-1}{2m}\right)\right)^2 \leq r^2$ just means that $\frac{1}{r}\eta_n \in K$. So, one has

$$(II) = \mathsf{P}\left\{\frac{1}{r}\eta_n \in K \vee [\zeta_n(\cdot) \notin K_\varepsilon]\right\}$$
$$\leq \mathsf{P}\left\{\frac{1}{r}\eta_n \in K \vee \left[\left\|\frac{1}{r}\eta_n - \zeta_n(\cdot)\right\| \geq \varepsilon\right]\right\}$$

Define

$$T = \begin{cases} \min\left\{t \in [0,1] : \left\|\frac{1}{r}\eta_n - \zeta_n(\cdot)\right\| \geq \varepsilon\right\} & \text{if} \quad \begin{array}{l}\text{this set is}\\ \text{nonempty}\end{array} \\ 2 & \text{if} \quad \text{otherwise}\end{cases}$$

and let F be its distribution function, so that

$$(II) \leq \int_{t=0}^{1} \mathsf{P}\left\{\frac{1}{r}\eta_n \in K \,\middle|\, T = t\right\} dF(t)$$

$$= \int_{t=0}^{1} \mathsf{P}\left\{\left[\frac{1}{r}\eta_n \in K\right] \vee \left[\left\|\frac{1}{r}\eta_n - \zeta_n(\cdot)\right\| = \varepsilon\right] \middle| T = t\right\} dF(t)$$

If $i(t)$ is the smallest i with $\frac{i}{2m} \geq t$, the statement $\frac{1}{r}\eta_n \in K$ implies

$$\left\|\eta_n\left(\frac{i(t)}{2m}\right) - \eta_n(t)\right\| \leq r \int_{t}^{i(t)/2m} \left\|\frac{1}{r}\dot{\eta}_n(s)\right\| ds$$
$$\leq r \int_{t}^{i(t)/2m} \left\|\frac{1}{r}\dot{\eta}_n(s)\right\|^2 ds \frac{1}{2m} \leq \frac{r}{\sqrt{2m}}$$

Two statements

$$\frac{1}{r}\eta_n \in K, \quad \left\|\frac{1}{r}\eta_n - \zeta_n(\cdot)\right\| = \varepsilon$$

hold together therefore, and because of the relation $\eta_n\left(\frac{i(t)}{2m}\right) = \zeta_n\left(\frac{i(t)}{2m}\right)$, imply

$$\left\|\zeta_n\left(\frac{i(t)}{2m}\right) - \zeta_n(t)\right\| \geq \|\eta_n(t) - \zeta_n(t)\| - \left\|\eta_n\left(\frac{i(t)}{2m}\right) - \eta_n(t)\right\|$$
$$\geq r\varepsilon - (r-1) - \frac{r}{\sqrt{2m}} \geq \varepsilon/2$$

if r is chosen close enough to 1 and m is sufficiently large. Then

$$(II) \leq \int_{t=0}^{1} \mathsf{P}\left\{\left\|\zeta_n\left(\frac{i(t)}{2m}\right) - \zeta_n(t)\right\| \geq \varepsilon/2 \mid T = t\right\} dF(t)$$
$$\leq \mathsf{P}\left\{\left\|\zeta_n\left(\frac{1}{2m}\right)\right\| \geq \varepsilon/2\right\} \int_{t=0}^{1} dF(t) \leq \mathsf{P}\left\{\left\|\zeta\left(\frac{n}{2m}\right)\right\| \geq \varepsilon/2\sqrt{2n \ln\ln n}\right\}$$
$$\simeq \frac{1}{\Gamma(k/2)}\left[\left(\varepsilon^2 m \ln\ln n\right)/2\right]^{k/2-1} \exp\left\{-\varepsilon^2 m \ln\ln n/2\right\}$$

By choosing m and r appropriately and using the above estimates for (I) and (II), it is easily seen that for $r > 1$ and sufficiently large n one has

$$\mathsf{P}\{\zeta_n(\cdot) \notin K_\varepsilon\} \leq \exp\left\{-r^2 \ln\ln n\right\} \tag{9.39}$$

If $n_j := \left[c^j\right] + 1$ where $c > 1$, then

$$\sum_j \mathsf{P}\left\{\zeta_{n_j}(\cdot) \notin K_\varepsilon\right\} \leq (\log c)^{-r^2} \sum_j j^{-r^2} < \infty$$

so that eventually $\zeta_{n_j}(\cdot) \in K_\varepsilon$ a.s., and for c sufficiently close to 1 this implies that eventually $\zeta_n(\cdot) \in K_{2\varepsilon}$ with probability 1. This shows that a.s. at most the points of K are limit points of $\{\zeta_n\}_{n\geq 3}$ and also that a.s. this sequence is relatively compact (for $\{\zeta_n : n \geq 3\}$ is totally bounded).

2. To prove the theorem it is therefore sufficient (because of the compactness of K) to show the following fact: *given $x \in K$ and $\varepsilon > 0$, the probability that ζ_n is infinitely often in the open ε-sphere $\{x\}_\varepsilon$ around x equals* 1. Let now $m \geq 1$ be an integer, $\delta \in (0,1)$ and x^l, ζ_n^l be the l-th coordinate of x and ζ_n respectively. Denote the event

$$\left|\zeta_n^l\left(\frac{i}{2m}\right) - \zeta_n^l\left(\frac{i-1}{2m}\right) - \left[x^l\left(\frac{i}{2m}\right) - x^l\left(\frac{i-1}{2m}\right)\right]\right| < \delta$$

by A_n. Then, defining $\varphi_{m,n}(\delta) := \sqrt{2m \ln\ln n}\left(\left|x^l\left(\frac{i}{2m}\right) - x^l\left(\frac{i-1}{2m}\right)\right| + \delta\right)$, we have

$$\begin{aligned}
\mathsf{P}\{A_n\} &\geq \prod_{i=2}^{m}\prod_{l=1}^{k}\frac{1}{\sqrt{2\pi}}\int_{\varphi_n(0)}^{\varphi_n(\delta)} e^{-s^2/2}ds \\
&\geq \text{Const}\prod_{i=2}^{m}\prod_{l=1}^{k}\frac{\exp\left\{-m\left[x^l\left(\frac{i}{2m}\right) - x^l\left(\frac{i-1}{2m}\right)\right]^2 \ln\ln n\right\}}{\sqrt{m\ln\ln n}}, \quad \text{Const} > 0
\end{aligned}$$

The last inequality for sufficiently large n is based on the estimate

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-s^2/2}ds \geq \frac{1}{b\sqrt{2\pi}}e^{-a^2/2}\left(1 - e^{-(b^2-a^2)/2}\right)$$

valid for any $a \leq [0,b)$. So, by summing up the exponents and using Cauchy–Schwartz's inequality we get

$$\mathsf{P}\{A_n\} \geq \frac{\text{Const}}{\ln n\sqrt{m\ln\ln n}} \tag{9.40}$$

Putting then $n_j = m^j$ we see that A_{n_j}'s are mutually independent and

$$\sum_{j=1}^{\infty}\mathsf{P}\left\{A_{n_j}\right\} \geq \text{Const}\sum_{j=1}^{\infty}\frac{1}{j\sqrt{\ln j}} = \infty$$

Hence, by Borel–Cantelli's Lemma 6.2 infinitely many events A_n happen almost surely. By what we previously have proved ζ_n is eventually close to K, and therefore (a.s.) we have eventually

$$\|\zeta_n(t) - \zeta_n(s)\| \leq \sqrt{|t-s|} + \delta \tag{9.41}$$

for all $s, t \in [0,1]$. Now if $y \in C^k[0,1]$ the following two statements

$$\|y(t) - y(s)\| \leq \sqrt{|t-s|} + \delta, \quad \forall s, t \in [0,1]$$

$$\left| y^l\left(\frac{i}{2m}\right) - y^l\left(\frac{i-1}{2m}\right) - \left[x^l\left(\frac{i}{2m}\right) - x^l\left(\frac{i-1}{2m}\right)\right]\right| < \delta$$
$$2 \le i \le m, \quad l = 1, \ldots, k$$

imply

$$\|y(t) - x(t)\| < \varepsilon$$

provided m is sufficiently large and δ is sufficiently small. Looking at the definition of A_n and in view of the fact that it happens infinitely often (a.s.) and by (9.41) we conclude that

$$\mathsf{P}\{|\zeta_n - x| < \varepsilon \text{ infinitely often}\} = 1$$

This completes the proof. □

The discreteness of n is inessential for the previous considerations. So, if $u > e$ is real and we put

$$\boxed{\zeta_u(t) := (2u \ln\ln u)^{-1/2} W_{ut}} \tag{9.42}$$

for $t \in [0, 1]$ we have the following statement.

Corollary 9.1. *With probability* 1 *the net* $\{\zeta_u(\cdot)\}_{u \ge e}$ *is relatively norm-compact and the set of its norms-limit points coincides with* K*, so putting* $t = 1$ *and* $u = \tau$ *we conclude that*

$$\boxed{\mathsf{P}\left\{\limsup_{\tau\to\infty} \frac{\|W_\tau\|}{\sqrt{2\tau \ln\ln\tau}} = 1\right\} = 1} \tag{9.43}$$

This corollary proves the fact that almost all trajectories of a standard Wiener process $\{W_t(\omega)\}_{t\ge 0}$ remain to be inside of the 'ε-*tube*'

$$Q_{1+\varepsilon} := \left\{(t, y) : |y| \le (1+\varepsilon)\sqrt{2t \ln\ln t} \quad (\text{for } t \ge e)\right\}$$

for any $\varepsilon > 0$, and in the same time with probability 1 they '*get out*' of the 'tube'

$$Q_{1-\varepsilon} := \left\{(t, y) : |y| \le (1-\varepsilon)\sqrt{2t \ln\ln t} \quad (\text{for } t \ge e)\right\}$$

infinitely often (see Fig. 9.2).

9.3.7 'White noise' and its interpretation

Gaussian 'white' noise $\{\mathcal{N}(t, \omega)\}_{t\in(-\infty,\infty)}$, in fact, is a model for a 'completely' random process whose individual random variable

$$\mathcal{N}\left(t = t', \omega\right)$$

is normally distributed. As such, it is an idealization of stochastic phenomena encountered in engineering system analysis.

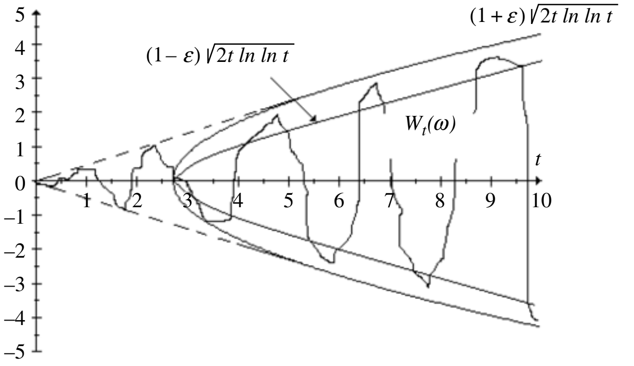

Fig. 9.2. The illustration of LIL for a standard Wiener process $W_t(\omega)$.

Definition 9.13. ***Gaussian 'white' noise*** $\{\mathcal{N}(t,\omega)\}_{t\in(-\infty,\infty)}$ *is defined (basically, in application literature) as a scalar stationary Gaussian process with* ***zero-mean mathematical expectation*** *and a constant* ***spectral density function*** $f(\lambda)$ *on the entire real line. that is,*

$$\boxed{\mathsf{E}\{\mathcal{N}(t,\omega)\} = 0 \quad \textit{for any} \quad t \in (-\infty,\infty)} \tag{9.44}$$

and if

$$\boxed{C(t) := \mathsf{E}\{\mathcal{N}(s,\omega)\,\mathcal{N}(s+t,\omega)\}} \tag{9.45}$$

then

$$\boxed{f(\lambda) := \frac{1}{2\pi}\int_{t=-\infty}^{\infty} e^{-i\lambda t} C(t)\,dt = \frac{c}{2\pi}, \quad \lambda \in R, c = \text{const}} \tag{9.46}$$

Remark 9.5. *Since the spectral density function may be interpreted as measuring the relative contribution of frequency* λ *to the oscillatory make-up of* $C(t)$*, the last equation (9.46), yielding that all frequencies are present equally, justifies the name 'white' noise in analogy with white light in physics.*

Remark 9.6. *The formula (9.46) also implies that*

$$C(t) = \delta(t)$$

where $\delta(t)$ *is the Dirac delta-function verifying the identity*

$$\int_{t=-\infty}^{\infty} \varphi(t)\,\delta(t-t_0) = \varphi(t_0)$$

for any continuous function $\varphi : \mathbb{R} \to \mathbb{R}$*, and hence,* $\{\mathcal{N}(t,\omega)\}_{t\in(-\infty,\infty)}$ *is uncorrelated at distinct times, so as a consequence, it is independent at distinct times since it is Gaussian.*

Remark 9.7. *Since*

$$C(t) = \int_{t=-\infty}^{\infty} e^{i\lambda t} f(\lambda)\, d\lambda$$

then, in particular,

$$\operatorname{var}(\mathcal{N}(s,\omega)) = C(0) = \int_{t=-\infty}^{\infty} f(\lambda)\, d\lambda = \infty$$

So, the nature of the covariance $C(t)$ indicates that ***such a process cannot be realized,*** *and that* ***'white' noise is not a stochastic process in the usual sense****.*

The relation between Gaussian white noise $\{\mathcal{N}(t,\omega)\}_{t\in(-\infty,\infty)}$ and the standard scalar Wiener process $\{W_t\}_{t\geq 0}$ can be understood formally as follows:

$$\mathcal{N}(t,\omega) \overset{a.s.}{=} \dot{W}_t \tag{9.47}$$

where $\dot{W}_t$ is the time-derivative of the Wiener process W_t. We have to emphasize that the formula (9.47) should be understood '*symbolically*' since, as we have already established, *the Wiener process is never differentiable*. The symbolic sense of the relation (9.47) means that the time integration of white noise processes has a specific format, namely,

$$\int_{\tau=0}^{t} \beta(\tau,\omega)\, \mathcal{N}(\tau,\omega)\, d\tau$$

it should be $\underset{=}{\text{understood}}$ as

$$\int_{\tau=0}^{t} \beta(\tau,\omega) \overset{\lambda=1/2}{\circ} dW_\tau(\omega)$$

where the right-hand side is *the Stratonovich integral* which will be introduced and discussed in detail in Chapter 11.

This page intentionally left blank

10 Markov Processes

Contents

A stochastic dynamic system satisfies *the Markov property*,[1]

if the probable (future) state of the system at any time $t > s$ is independent of the (past) behavior of the system at times $t < s$, given the present state at time s.

This property can be nicely illustrated by considering a classical movement of a particle whose trajectory after time s depends only on its coordinates (position) and velocity at time s, so that its behavior before time s has absolutely no affect on its dynamic after time s. In fact, this property for stochastic systems is completely shared with one for solutions of initial value problems involving ordinary differential equations. So, stochastic processes satisfying this property arise naturally as solutions of stochastic differential equations obtained from ordinary ones but containing a stochastic perturbation term in the right-hand side. These equations will be considered in Chapter 13.

In this chapter the Markov property is made precise and the basic properties of Markov processes are discussed.

10.1 Definition of Markov property

10.1.1 Main definition

Let $x(t, \omega) \in \mathbb{R}^n$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathsf{P})$ with state space $\mathbb{R}^n$ and the index set $J := [t_0, T] \subseteq [0, \infty)$ where $T = \infty$ may be considered. For any $t_1, t_2 \in J$ define

$$\mathcal{F}_{[t_1,t_2]} := \sigma \{x(t, \omega), t_1 \leq t \leq t_2\} \tag{10.1}$$

where $\sigma \{x(t, \omega), t_1 \leq t \leq t_2\}$ is a minimal sigma-algebra generated by the set of 'intervals' (rectangles, etc.) of the form

[1] This definition was introduced by A.A. Markov in 1906.

$$\begin{array}{l}\{\omega : x(\tau_1, \omega) \in B_1, \ldots, x(\tau_m, \omega) \in B_m\} \\ \quad t_1 \le \tau_1 < \tau_2 < \cdots < \tau_m \le t_2 \\ B_i \in \mathcal{B}^n \text{ is a Borel sets of the dimension } n \\ \quad m \text{ is any positive integer}\end{array} \tag{10.2}$$

Definition 10.1. $\{x(t,\omega)\}_{t\in J}$ *is called a* ***Markov process (MP)*** *if the following* ***Markov property*** *holds: for any* $t_0 \le \tau \le t \le T$ *and all* $A \in \mathcal{B}^n$

$$\boxed{\mathsf{P}\left\{x(t,\omega) \in A \mid \mathcal{F}_{[t_0,\tau]}\right\} \overset{a.s.}{=} \mathsf{P}\{x(t,\omega) \in A \mid x(\tau,\omega)\}} \tag{10.3}$$

The statement below, based on the properties of the conditional mathematical expectations, seems to be evident.

Claim 10.1. *The following are each equivalent to the Markov property (10.3):*

- *for any* $t_0 \le \tau \le t \le T$ *and all* $A \in \mathcal{F}_{[t,T]}$

$$\boxed{\mathsf{P}\left\{A \mid \mathcal{F}_{[t_0,\tau]}\right\} \overset{a.s.}{=} \mathsf{P}\{A \mid x(\tau,\omega)\}} \tag{10.4}$$

- *for any* $t_0 \le \tau \le t \le T$ *and all* $Y \in \mathcal{F}_{[t,T]}$

$$\boxed{\mathsf{E}\left\{Y \mid \mathcal{F}_{[t_0,\tau]}\right\} \overset{a.s.}{=} \mathsf{E}\{Y \mid x(\tau,\omega)\}} \tag{10.5}$$

- *for any* $t_0 \le t_1 \le t \le t_2 \le T$ *and all* $A_1 \in \mathcal{F}_{[t_0,t_1]}$ *and* $A_2 \in \mathcal{F}_{[t_2,T]}$

$$\boxed{\mathsf{P}\{A_1 \cap A_2 \mid x(t,\omega)\} \overset{a.s.}{=} \mathsf{P}\{A_1 \mid x(t,\omega)\}\,\mathsf{P}\{A_2 \mid x(t,\omega)\}} \tag{10.6}$$

10.1.2 Criterion for a process to have the Markov property

Lemma 10.1. *A stochastic process* $x(t,\omega) \in \mathbb{R}^n$*, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*, with state space* $\mathbb{R}^n$ *and the index set* $J := [t_0, T] \subseteq [0, \infty)$ *is a Markov process* ***if and only if*** *for any* $t \in J$ *and all bounded* $F \in \mathcal{F}_{[t_0,t]}$ *(F from* $\mathbb{R}^k$*) and* $G \in \mathcal{F}_{[t,T]}$ *(F from* $\mathbb{R}^l$*) the following identity holds:*

$$\boxed{\mathsf{E}\{FG^\top \mid x(t,\omega)\} \overset{a.s.}{=} \mathsf{E}\{F \mid x(t,\omega)\}\,\mathsf{E}\{G^\top \mid x(t,\omega)\}} \tag{10.7}$$

Proof. Obviously (10.3) is a partial case of (10.6) when $k = l = 1$, $t_1 = t = t_2$ and

$$F = \chi\left(\omega \in A_1 \in \mathcal{F}_{[t_0,t]}\right), \quad G = \chi\left(\omega \in A_2 \in \mathcal{F}_{[t,T]}\right)$$

By the linear property of the conditional mathematical expectation, namely,

$$\mathsf{E}\left\{\alpha S_1 + \beta S_2 \mid \mathcal{F}_{[t_0,t]}\right\} \overset{a.s.}{=} \alpha\mathsf{E}\left\{S_1 \mid \mathcal{F}_{[t_0,t]}\right\} + \beta\mathsf{E}\left\{S_2 \mid \mathcal{F}_{[t_0,t]}\right\}$$

valid for all $\alpha, \beta \in \mathbb{R}$, by the boundedness property for the considered vector-functions F and G, it seems to be evident that the property is valid for 'simple' functions having the representation

$$F := \sum_{i=1}^{M} c_i \chi\left(\omega \in A_i \in \mathcal{F}_{[t_0,t]}\right), \quad G = \sum_{j=1}^{N} \chi\left(\omega \in A_j \in \mathcal{F}_{[t,T]}\right)$$

Then by Theorem 2.3 on the monotone approximation it follows that any $\mathcal{F}_{[t_0,t]}$ (or $\mathcal{F}_{[t,T]}$) measurable bounded function h such that

$$\sup_{\omega \in \Omega} \|h(\omega)\| \leq H < \infty$$

may be uniformly on $\omega \in \Omega$ (for almost all ω) approximated by a system of 'simple' functions that proves the desired result (10.7). □

Lemma 10.2. *A stochastic process $x(t,\omega) \in \mathbb{R}$, defined on $(\Omega, \mathcal{F}, \mathsf{P})$, with state space $\mathbb{R}^n$ and the index set $J := [t_0, T] \subseteq [0, \infty)$ is a Markov process* ***if and only if*** *for any integer m, n and any*

$$t_0 \leq s_1 < \cdots < s_m \leq t \leq t_1 < \cdots < t_n \leq T$$

and any bounded (scalar valued) functions

$$g_i(\omega, x(s_i,\omega)) \in \mathcal{F}_{[t_0,s_i]} \ (i = 1, \ldots, m)$$
$$f_j\left(\omega, x\left(t_j,\omega\right)\right) \in \mathcal{F}_{[t_j,T]} \ (j = 1, \ldots, n)$$

the following property holds:

$$\boxed{\begin{aligned} &\mathsf{E}\left\{\prod_{i=1}^{m} g_i(\omega, x(s_i,\omega)) \prod_{j=1}^{n} g_j\left(\omega, x\left(t_j,\omega\right)\right) \right. \\ &\qquad\left. \Big| \, x(s_1,\omega), \ldots, x(s_m,\omega); x(t,\omega)\right\} \\ &\quad \overset{a.s.}{=} \mathsf{E}\left\{\prod_{i=1}^{m} g_i(\omega, x(s_i,\omega)) \prod_{j=1}^{n} g_j\left(\omega, x\left(t_j,\omega\right)\right) \Big| \, x(t,\omega)\right\}\end{aligned}} \tag{10.8}$$

Proof. First, notice that (10.7) holds in the component-wise sense; therefore it is sufficient to check this property for scalar valued functions. Remember that for the conditional mathematical expectation the following 'smoothing' property (7.11) holds: if we $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$, then

$$\mathsf{E}\{\xi/\mathcal{F}_0\} \overset{a.s.}{=} \mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_0\}/\mathcal{F}_1\} \overset{a.s.}{=} \mathsf{E}\{\mathsf{E}\{\xi/\mathcal{F}_1\}/\mathcal{F}_0\}$$

(a) Suppose that (10.7) holds. Then taking

$$F := \prod_{i=1}^{m} g_i(\omega, x(s_i,\omega)), \quad G := \prod_{j=1}^{n} g_j\left(\omega, x\left(t_j,\omega\right)\right)$$

we obtain (10.8), and hence, the necessity property follows.

(b) Let now (10.8) hold. Then (the arguments are omitted) taking

$$f_i = \chi_i^- := \chi\left(\omega \in A_i^- \in \mathcal{F}_{[t_0,s_i]}\right)$$
$$g_j = \chi_j^+ := \chi\left(\omega \in A_j^+ \in \mathcal{F}_{[t_0,t_j]}\right)$$
$$\mathring{A}_1 := A_1^- \times \cdots \times A_m^-, \quad \mathring{A}_2 := A_1^+ \times \cdots \times A_n^+$$

we have

$$\begin{aligned}
&\mathsf{P}\left\{\mathring{A}_1 \cap \mathring{A}_2 \mid x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\prod_{i=1}^{m} \chi_i^- \prod_{j=1}^{n} \chi_j^+ \,\middle|\, x(s_1,\omega),\ldots,x(s_m,\omega); x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\prod_{i=1}^{m} f_i \prod_{j=1}^{n} g_j \,\middle|\, x(s_1,\omega),\ldots,x(s_m,\omega); x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\prod_{i=1}^{m} f_i \prod_{j=1}^{n} g_j \,\middle|\, x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\mathsf{E}\left\{\prod_{i=1}^{m} \chi_i^- \prod_{j=1}^{n} \chi_j^+ \,\middle|\, x(s_1,\omega),\ldots,x(s_m,\omega)\right\} \,\middle|\, x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\prod_{i=1}^{m} \chi_i^- \mathsf{E}\left\{\prod_{j=1}^{n} \chi_j^+ \,\middle|\, x(s_1,\omega),\ldots,x(s_m,\omega)\right\} \,\middle|\, x(t,\omega)\right\} \\
&\quad = \mathsf{E}\left\{\prod_{i=1}^{m} \chi_i^- \,\middle|\, x(t,\omega)\right\} \mathsf{E}\left\{\prod_{j=1}^{n} \chi_j^+ \,\middle|\, x(t,\omega)\right\} \\
&\quad = \mathsf{P}\left\{\mathring{A}_1 \,\middle|\, x(t,\omega)\right\} \mathsf{P}\left\{\mathring{A}_2 \,\middle|\, x(t,\omega)\right\}
\end{aligned}$$

which proves the sufficiency property. □

The following corollary evidently follows.

Corollary 10.1. *A stochastic process* $x(t,\omega) \in \mathbb{R}$*, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*, with state space* $\mathbb{R}^n$ *and the index set* $J := [t_0, T] \subseteq [0, \infty)$ *is a Markov process* ***if and only if*** *for any integer m and* $n = 1$ *the property (10.8) holds.*

The next criterion is most important in applications.

Corollary 10.2. *A stochastic process* $x(t,\omega) \in \mathbb{R}$*, defined on* $(\Omega, \mathcal{F}, \mathsf{P})$*, with state space* $\mathbb{R}^n$ *and the index set* $J := [t_0, T] \subseteq [0, \infty)$ *is a Markov process* ***if and only if*** *for any integer m, all*

$$t_0 \le s_1 < \cdots < s_m \le t \le u \le T$$

and any $A \in \mathcal{B}^n$

$$\boxed{\begin{array}{c} \mathsf{P}\{x(u,\omega) \in A \mid x(s_1,\omega), \ldots, x(s_m,\omega), x(t,\omega)\} \\ \overset{a.s.}{=} \mathsf{P}\{x(u,\omega) \in A \mid x(t,\omega)\} \end{array}} \tag{10.9}$$

Proof. It follows directly from Lemma 10.2 if we take there $n = 1$ and

$$g_i(\omega, x(s_i,\omega)) := \chi(\omega : x(s_i,\omega) \in A)$$ □

10.2 Chapman–Kolmogorov equation and transition function

10.2.1 Transition probability and its four main properties

Definition 10.2. *A function* $\mathsf{P}\{s, x, t, A\}$ *as a function of four variables, with* $t_0 \le s \le t \le T$, $x \in \mathbb{R}^n$ *and* $A \in \mathcal{B}^n$ *is called the* ***transition probability*** *or* ***transition function*** *of a stochastic process* $\{x(t,\omega)\}_{t\in[t_0,T]}$ *if the following four properties are satisfied:*

1. *For any fixed* $s \le t$ *and any fixed* $x \in \mathbb{R}^n$ *the function* $\mathsf{P}\{s, x, t, \cdot\}$ *is a* ***probability*** *on* $\mathcal{B}^n$*;*
2. *For any fixed* $s \le t$ *and any fixed set* $A \in \mathcal{B}^n$ *the function* $\mathsf{P}\{s, \cdot, t, A\}$ *is* $\mathcal{B}^n$*-measurable;*
3. *For all* $s \in [t_0, T]$*, all* $x \in \mathbb{R}^n$ *and all* $A \in \mathcal{B}^n$

$$\boxed{\mathsf{P}\{s, x, s, A\} = \chi(\omega : x(s,\omega) \in A)} \tag{10.10}$$

4. *For any fixed* $s < u < t$*, any* $A \in \mathcal{B}^n$ *and all* $x \in \mathbb{R}^n$ *(except possibly for a* P_x*-null set)*

$$\boxed{\mathsf{P}\{s, x, t, A\} = \int\limits_{u\in\mathbb{R}^n} \mathsf{P}\{u, y, t, A\}\, \mathsf{P}\{s, x, u, dy\}} \tag{10.11}$$

   which is known as the ***Chapman–Kolmogorov equation****.*

Remark 10.1. *By the property 3 (10.10) the relation (10.11) will be fulfilled also for* $s \le u \le t$.

10.2.2 Two-step interpretation of the Chapman–Kolmogorov equation

A Markov process $\{x(t,\omega)\}_{t\in[t_0,T]}$ is said to have *the transition function* $\mathsf{P}\{s, x, t, A\}$ if for all $s \le t$ (both from $[t_0, T]$) and all $A \in \mathcal{B}^n$

$$\boxed{\mathsf{P}\{x(t,\omega) \in A \mid x(s,\omega) = x\} \overset{a.s.}{=} \mathsf{P}\{s, x, t, A\}} \tag{10.12}$$

(on the probabilistic measure $\mathsf{P}_{x(s,\omega)}$). In fact, it defines the probability to meet the stochastic process within a set A if at time s it has a value x, i.e., $x(s,\omega) = x$.

The Chapman–Kolmogorov equation (10.11) indicates that transition probability (10.12) can be decomposed into the state-space integral of products of probabilities to and from a location in state space, attained at an arbitrary intermediate fixed time in the parameter or index set, that is, *the one-step transition probability can be rewritten in terms of all possible combinations of two-step transition probabilities with respect to any arbitrary intermediate time*. This fact can be easily understood using the properties of the conditional mathematical expectation:

$$\begin{aligned}
\mathsf{P}\{s,x,t,A\} &\overset{a.s.}{=} \mathsf{P}\{x(t,\omega)\in A \mid x(s,\omega)=x\} \\
&= \mathsf{E}\{\chi(\omega : x(t,\omega)\in A) \mid x(s,\omega)=x\} \\
&= \mathsf{E}\{\mathsf{E}\{\chi(\omega : x(t,\omega)\in A) \mid x(s,\omega)=x; x(u,\omega)\} \mid x(s,\omega)=x\} \\
&= \mathsf{E}\{\mathsf{E}\{\chi(\omega : x(t,\omega)\in A) \mid x(u,\omega)\} \mid x(s,\omega)=x\} \\
&= \mathsf{E}\{\mathsf{P}\{x(t,\omega)\in A \mid x(u,\omega)\} \mid x(s,\omega)=x\} \\
&= \mathsf{E}\{\mathsf{P}\{u,x(u,\omega),t,A\} \mid x(s,\omega)=x\}
\end{aligned}$$

The importance of the transition probabilities for Markov processes is that all finite-dimensional distributions of the process can be obtained from them and the initial distribution at time t_0, namely, for $t_0 \le t_1 \le \cdots \le t_m \le T$ and $A_i \in \mathcal{B}^n$ one has

$$\boxed{\begin{gathered}
\mathsf{P}\{x(t_1,\omega)\in A_1,\ldots,x(t_m,\omega)\in A_m\} = \int\limits_{\mathbb{R}^n}\int\limits_{A_1}\cdots\int\limits_{A_{m-1}} \\
\mathsf{P}\{t_{m-1},x(t_{m-1},\omega),t_m,A_m\}\,\mathsf{P}\left\{t_{m-2},x_{m-2},t_{m-1},dx_{m-1}\right\} \\
\times\cdots\,\mathsf{P}\{t_0,x_0,t_1,dx_1\}\,\mathsf{P}\{x(t_0,\omega)\in dx_0\}
\end{gathered}} \tag{10.13}$$

This means exactly the following fact.

Claim 10.2. *A probability transition function* $\mathsf{P}\{s,x,t,A\}$ *and an initial distribution* $\mathsf{P}\{x(t_0,\omega)\in A_0\}$ *determines uniquely (up to stochastic equivalence) a Markov process.*

10.2.3 Homogeneous Markov processes

Definition 10.3. *A Markov process* $\{x(t,\omega)\}_{t\in[t_0,T]}$ *is called* ***homogeneous*** *(with respect to time index) if for all* $x\in\mathbb{R}^n$, $A\in\mathcal{B}^n$ *and any* $s, s+h\in[t_0,T]$

$$\boxed{\mathsf{P}\{s,x,s+h,A\} = \mathsf{P}\{0,x,h,A\}} \tag{10.14}$$

The right-hand side of (10.14) exactly defines the conditional probability for a stochastic process $\{x(t,\omega)\}_{t\in[t_0,T]}$ to appear in the set A after time h starting from the state x independently of the time $[s,s+h]\in[t_0,T]$ when this transition occurs.

For a homogeneous Markov process $\{x(t,\omega)\}_{t\in[t_0,T]}$ the properties in Definition 10.2 can be rewritten in the following form:

$1'$. *For any fixed t and any fixed $x \in R^n$ the function $P\{0,x,t,\cdot\}$ is a probability on B^n*;
$2'$. *For any fixed t and any fixed set $A \in B^n$ the function $P\{0,\cdot,t,A\}$ is B^n-measurable*;
$3'$. *For all $x \in R^n$ and all $A \in B^n$*

$$\mathsf{P}\{0,x,0,A\} = \chi(\omega : x(0,\omega) \in A)$$

$4'$. *For any fixed $s < t$, any $A \in B^n$ and all $x \in R^n$ (except possibly for a P_x-null set)*

$$\mathsf{P}\{0,x,s+t,A\} = \int\limits_{u\in\mathbb{R}^n} \mathsf{P}\{0,y,t,A\}\,\mathsf{P}\{0,x,s,dy\}$$

In fact, the last property ($4'$) permits the application of the power apparatus of the *semi-group theory* to analyze the behavior of homogenous Markov processes, but this is out with the scope of this book.

Below we will present the following interesting fact.

Lemma 10.3. *Any homogeneous Markov process is a stochastic process with a* ***stationary increment****.*

Proof. Obviously for any $t_0 \le s+u < t+u \le T$, any $x \in R^n$ *and all* $A \in B^n$

$$\begin{aligned}
\mathsf{P}\{s+u,x,t+u,A\} &= \mathsf{P}\{x(t+u,\omega) \in A \mid x(s+u,\omega) = x\} \\
&= \mathsf{P}\{x(t+u,\omega) - x(s+u,\omega) \in A - x \mid x(s+u,\omega) = x\} \\
&\quad \text{by Definition 10.3} \\
&= \mathsf{P}\{s,x,t,A\} \\
&= \mathsf{P}\{x(t,\omega) \in A \mid x(s,\omega) = x\} \\
&= \mathsf{P}\{x(t,\omega) - x(s,\omega) \in A - x \mid x(s,\omega) = x\} = \mathsf{P}\{s,x,t,A\}
\end{aligned}$$

which proves the statement. □

Since $\mathsf{P}\{s,x,t,A\}$ for homogeneous Markov processes depends only on $(t-s)$, x and A, one can use the notation

$$\mathsf{P}\{s,x,t,A\} := \mathsf{P}\{t-s,x,A\}$$

and the *Chapman–Kolmogorov equation* in $4'$ becomes

$$\boxed{\mathsf{P}\{t+s,x,A\} = \int\limits_{u\in\mathbb{R}^n} \mathsf{P}\{s,y,A\}\,\mathsf{P}\{t,x,dy\}} \tag{10.15}$$

10.2.4 Process with independent increments as MP

Here we show that a wide class of stochastic processes considered before, namely, processes with independent increments, are, in fact, Markov processes.

Theorem 10.1. *Let $\{x(t,\omega)\}_{t\in[t_0,T]}$ be a process with* ***independent increment*** *(see Definition 9.9) taking values in $\mathbb{R}^n$ and such that $x(t,\omega)$ is $\mathcal{F}_t$-measurable with*

$$\mathcal{F}_t = \sigma(\{x(\tau,\omega)\}, \tau \in [t_0, t]), \quad t \in [t_0, T]$$

Then $\{x(t,\omega)\}_{t\in[t_0,T]}$ is a ***Markov process****.*

Proof. For

$$t_0 \le s_1 < \cdots < s_m \le t \le u \le T$$

define

$$\begin{aligned} &\varsigma_1 := x(s_1,\omega), \quad \varsigma_2 := x(s_2,\omega) - x(s_1,\omega), \ldots, \\ &\varsigma_m := x(s_m,\omega) - x(s_{m-1},\omega), \quad \varsigma_{m+1} := x(t,\omega) - x(s_m,\omega) \\ &\varsigma := x(u,\omega) - x(t,\omega) \end{aligned}$$

Then for any bounded, $\mathcal{B}^n$-measurable function g we have

$$\begin{aligned} &\mathsf{E}\{g(x(u,\omega)) \mid x(s_1,\omega), \ldots, x(s_m,\omega), x(t,\omega)\} \\ &\quad = \mathsf{E}\left\{ g\left(\varsigma + \sum_{i=1}^{m+1} \varsigma_i\right) \middle| \varsigma_1, \ldots, \varsigma_{m+1} \right\} \end{aligned}$$

so that

$$\begin{aligned} &\mathsf{E}\left\{ g\left(\varsigma + \sum_{i=1}^{m+1} \varsigma_i\right) \middle| \varsigma_1 = y_1, \ldots, \varsigma_{m+1} = y_{m+1} \right\} \\ &\quad = \mathsf{E}\left\{ g\left(\varsigma + \sum_{i=1}^{m+1} y_i\right) \right\} \overset{a.s.}{=} \psi\left(\sum_{i=1}^{m+1} y_i\right) \end{aligned}$$

where ψ is a Borel function. On the other hand, by the properties of the conditional mathematical expectation,

$$\begin{aligned} &\mathsf{E}\{g(x(u,\omega)) \mid x(t,\omega)\} = \mathsf{E}\left\{ g\left(\varsigma + \sum_{i=1}^{m+1} \varsigma_i\right) \middle| \sum_{i=1}^{m+1} \varsigma_i \right\} \\ &\mathsf{E}\left\{ \mathsf{E}\left\{ g\left(\varsigma + \sum_{i=1}^{m+1} \varsigma_i\right) \middle| \varsigma_1, \ldots, \varsigma_{m+1} \right\} \middle| \sum_{i=1}^{m+1} \varsigma_i \right\} \\ &\quad = \mathsf{E}\left\{ \psi\left(\sum_{i=1}^{m+1} \varsigma_i\right) \middle| \sum_{i=1}^{m+1} \varsigma_i \right\} \overset{a.s.}{=} \psi\left(\sum_{i=1}^{m+1} \varsigma_i\right) \end{aligned}$$

which means that

$$\begin{aligned} &\mathsf{E}\{g(x(u,\omega)) \mid x(s_1,\omega), \ldots, x(s_m,\omega), x(t,\omega)\} \\ &\quad \overset{a.s.}{=} \mathsf{E}\{g(x(u,\omega)) \mid x(t,\omega)\} \end{aligned}$$

Theorem is proven. □

Proposition 10.1. *Both a **Brownian motion** W_t (or Wiener process) (see Definition 9.11) and a **Poisson process** N_t (9.24) are **Markov processes**.*

This result follows directly from the property of the Wiener process as well as Poisson process that both are processes with independent increments.

Example 10.1. *For a standard Wiener process $W_t(\omega)$, defined by (Definition 9.11), is a homogeneous Markov process with the transition (stationary) probability*

$$\boxed{\mathsf{P}\{t, x, A\} = \int_{y \in A} (2\pi t)^{-n/2} \exp\left\{-|y-x|^2/2t\right\} dy} \tag{10.16}$$

It is possible to show that a standard Wiener process satisfies the so-called '*strong Markov property*', the exact definition of which is given below.

10.2.5 Strong Markov property

Consider a continuous-time stochastic Markov process $\{x(t,\omega)\}_{t\in[t_0,T]}$, taking values in $\mathbb{R}^n$, and a closed set $\mathbb{F} \subset \mathbb{R}^n$. Let

$$\boxed{\tau = \tau(\omega) := \inf\{t : x(t,\omega) \in \mathbb{F}\}} \tag{10.17}$$

be a Markov time (see Chapter 7, Section 2.4) called the *first hitting time* of the set $\mathbb{F}$ or the *first exit time of the complementary set* $\mathbb{F}^c$.

Definition 10.4. *A continuous-time stochastic Markov process is called a **strong Markov process**, or **having the strong Markov property**, if the Markov property (10.3) holds for any Markov time, namely, when for any Markov time τ (10.17), and any Borel set $A \in \mathcal{B}^n$,*

$$\boxed{\mathsf{P}\left\{x(t+\tau,\omega) \in A \mid \mathcal{F}_{[t_0,\tau]}\right\} \overset{a.s.}{=} \mathsf{P}\{x(t+\tau,\omega) \in A \mid x(\tau,\omega)\}} \tag{10.18}$$

for any $t > 0$ such that $t + \tau \in [t_0, T]$.

10.3 Diffusion processes

10.3.1 Main definition

Definition 10.5. *A continuous-time stochastic Markov process $\{x(t,\omega)\}_{t\in[t_0,T]}$, taking values in $\mathbb{R}^n$, is called a **diffusion process** if its transition probability $\mathsf{P}\{s,x,t,A\}$ (see Definition 10.2) is smooth enough in the sense that is satisfies the following three conditions for every $s \in [t_0,T]$, $x \in \mathbb{R}^n$ and $\varepsilon > 0$:*

$$\lim_{t \to s+0} \frac{1}{t-s} \int_{|y-x|>\varepsilon} \mathsf{P}\{s,x,t,dy\} = 0 \tag{10.19}$$

$$\lim_{t \to s+0} \frac{1}{t-s} \int_{|y-x| \le \varepsilon} (y-x) \mathsf{P}\{s, x, t, dy\} = a(s, x) \tag{10.20}$$

$$\lim_{t \to s+0} \frac{1}{t-s} \int_{|y-x| \le \varepsilon} (y-x)(y-x)^{\mathsf{T}} \mathsf{P}\{s, x, t, dy\} = B(s, x) \tag{10.21}$$

where $a(s, x)$ and $B(s, x)$ represent well-defined $\mathbb{R}^n$ and $\mathbb{R}^{n \times n}$-valued functions respectively. These functions are called the coefficients of the given diffusion process:

- *$a(s, x)$ is referred to as the **drift vector**;*
- *$B(s, x)$ is referred to as the **diffusion matrix** which is symmetric and nonnegative definite for each admissible s, x.*

Obviously, the properties (10.19)–(10.21) can be rewritten, similarly, as follows:

$$\boxed{\begin{aligned} & \mathsf{P}\{|x(t+h,\omega) - x(t,\omega)| > \varepsilon \mid x(s,\omega) = x\} = o(h) \\ & \mathsf{E}\{[x(t+h,\omega) - x(t,\omega)] \times \\ & \quad \chi(|x(t+h,\omega) - x(t,\omega)| \le \varepsilon) \mid x(s,\omega) = x\} = a(s,x)h + o(h) \\ & \mathsf{E}\{[x(t+h,\omega) - x(t,\omega)][x(t+h,\omega) - x(t,\omega)]^{\mathsf{T}} \times \\ & \quad \chi(|x(t+h,\omega) - x(t,\omega)| \le \varepsilon) \mid x(s,\omega) = x\} \\ & \quad = B(s,x)h + o(h) \end{aligned}} \tag{10.22}$$

10.3.2 Kolmogorov´s backward and forward differential equation

Let us consider a scalar random Markov process $\{x(t, \omega)\}_{t \in [t_0, T]}$ with transition function $\mathsf{P}\{s, x, t, A\}$ where in this case

$$A := \{y \le x(t, \omega) \le y + \Delta y\}$$

Suppose also that there exists a function $p\{s, x, t, y\}$ such that

$$\boxed{\mathsf{P}\left\{s, x, t, y' \le x(t, \omega) \le y''\right\} = \int_{y'}^{y''} p(s, x, t, y)\, dy} \tag{10.23}$$

This function $p\{s, x, t, y\}$ is called the *density of transition probability* of the given Markov process, and, in fact, it defines the conditional density of the probability for the random variable $x(t, \omega)$ for the given $t \in [t_0, T]$ and any values $x(\tau, \omega)$ $(\tau \le s)$ including the condition $x(s, \omega) = x$ for $\tau = s$.

Lemma 10.4. *The density $p(s, x, t, y)$ of transition probability also satisfies the* ***Chapman –Kolmogorov equation*** ***for densities****:*

$$\boxed{p(s,x,t,y) = \int\limits_{z\in\mathbb{R}^n} p(s,x,u,z)\, p(u,z,t,y)\, dz} \tag{10.24}$$

valid for any $s \le u \le t$.

Proof. The joint density of distribution for the random variable $x(u,\omega)$ and $x(t,\omega)$, in view of independence of $x(t,\omega)$ on $x(s,\omega) = x$, is equal to

$$p(s,x,u,z)\, p(u,z,t,y), \quad z, y \in \mathbb{R}$$

Integrating by z this product, we obtain (10.24). □

10.3.2.1 Kolmogorov's backward equation

Theorem 10.2. *Let the density $p(s,x,t,y)$ of transition probability for a Markov process, with the drift vector $a(s,x)$ and the diffusion matrix $b(s,x)$, have the derivatives*

$$\frac{\partial}{\partial s} p(s,x,t,y), \quad \frac{\partial}{\partial x} p(s,x,t,y) \quad \text{and} \quad \frac{\partial^2}{\partial x^2} p(s,x,t,y)$$

which are uniformly continuous in y at any finite interval $y' \le y \le y''$. Then for any $t \in [a,b]$ and any $y \in \mathbb{R}$ it satisfies the following partial differential equation

$$\boxed{\begin{aligned} -\frac{\partial}{\partial s} p(s,x,t,y) = a(s,x) \frac{\partial}{\partial x} p(s,x,t,y) \\ + \frac{1}{2} b(s,x) \frac{\partial^2}{\partial x^2} p(s,x,t,y) \end{aligned}} \tag{10.25}$$

*known as the **Kolmogorov backward equation**.*

Proof. Consider any continuous function $\phi_c(x)$ which is equal to zero outside some finite interval, i.e.,

$$\phi_{[a,b]}(x) := \begin{cases} \phi(x) & \text{if } a \le x \le b \\ 0 & \text{if } x \in [a,b] \end{cases}$$

and denote

$$\varphi(s,x,t) := \int\limits_{y=-\infty}^{\infty} \phi_{[a,b]}(y)\, p\{s,x,t,y\}\, dy = \mathsf{E}_{s,x}\left\{\phi_{[a,b]}(x(t,\omega))\right\} \tag{10.26}$$

From the Chapman–Kolmogorov equation for densities (10.24) it follows that for any $t_0 \le s \le u \le t \le T$

$$\varphi(s,x,t) = \int\limits_{y=-\infty}^{\infty} \phi_{[a,b]}(y)\, p(s,x,t,y)\, dy$$

$$= \int_{y=-\infty}^{\infty} \phi_{[a,b]}(y) \left[\int_{z \in \mathbb{R}^n} p(s,x,u,z)\, p(u,z,t,y)\, dz \right] dy$$
$$= \int_{z \in \mathbb{R}^n} p(s,x,u,z) \left[\int_{y=-\infty}^{\infty} \phi_{[a,b]}(y)\, p(u,z,t,y)\, dy \right] dz$$
$$= \int_{z \in \mathbb{R}^n} p(s,x,u,z)\, \varphi(u,z,t)\, dz \tag{10.27}$$

Obviously, $\varphi(s,x,t)$ has continuous partial derivatives $\frac{\partial}{\partial s}\varphi(s,x,t)$, $\frac{\partial}{\partial x}\varphi(s,x,t)$ and $\frac{\partial^2}{\partial x^2}\varphi(s,x,t)$, and therefore it can be approximated by the first two terms of the Taylor expansion in the neighborhood of the point x (under a fixed u and t):

$$\varphi(u,z,t) - \varphi(u,x,t) = \frac{\partial \varphi(u,x,t)}{\partial x}(z-x) + \frac{1}{2}\left[\frac{\partial^2 \varphi(u,x,t)}{\partial x^2} + O(\delta_\varepsilon(u,x,t)) \right](z-x)^2$$

where

$$\delta_\varepsilon(u,x,t) = \sup_{|z-x| \le \varepsilon = b-a} \left| \frac{\partial^2 \varphi(u,z,t)}{\partial x^2} - \frac{\partial^2 \varphi(u,x,t)}{\partial x^2} \right| \underset{\varepsilon \to 0}{\to} 0$$

The presentations (10.19)–(10.21) imply

$$\varphi(s,x,t) - \varphi(u,x,t) = \int_{z \in \mathbb{R}^n} [\varphi(s,z,t) - \varphi(u,x,t)]\, p(s,x,u,z)\, dz$$
$$= \int_{|z-x| \le \varepsilon} [\varphi(s,z,t) - \varphi(u,x,t)]\, p(s,x,u,z)\, dz + o(|u-s|)$$
$$= \frac{\partial \varphi(u,x,t)}{\partial x} \int_{|z-x| \le \varepsilon} (z-x)\, p(s,x,u,z)\, dz$$
$$+ \frac{1}{2}\left[\frac{\partial^2 \varphi(u,x,t)}{\partial x^2} + O(\delta_\varepsilon(u,x,t)) \right] \int_{|z-x| \le \varepsilon} (z-x)^2\, p(s,x,u,z)\, dz$$
$$= \frac{\partial \varphi(u,x,t)}{\partial x} a(s,x)$$
$$+ \frac{1}{2}\left[\frac{\partial^2 \varphi(u,x,t)}{\partial x^2} + O(\delta_\varepsilon(u,x,t)) \right] b(s,x) + o(|u-s|)$$

which leads to the following identity

$$0 = \lim_{u \downarrow s} [\varphi(s, x, t) - \varphi(u, x, t)] = -\frac{\partial \varphi(s, x, t)}{\partial t} + \frac{\partial \varphi(s, x, t)}{\partial x} a(s, x) + \frac{1}{2} \frac{\partial^2 \varphi(s, x, t)}{\partial x^2} b(s, x)$$

Taking into account the definition (10.26), the last equation can be rewritten as

$$\int_{y=-\infty}^{\infty} \phi_{[a,b]}(y) \left[\frac{\partial p(s, x, t, y)}{\partial t} + \frac{\partial p(s, x, t, y)}{\partial x} a(s, x) + \frac{1}{2} \frac{\partial^2 p(s, x, t, y)}{\partial x^2} b(s, x) \right] dy = 0$$

Remembering that $\phi_{[a,b]}(y)$ is any continuous function, equal to zero outside of $[a, b]$, and extending this interval, we obtain (10.25). Theorem is proven. □

Remark 10.2. *Notice that the density $p(s, x, t, y)$ (10.23) of transition probability coincides with the so-called fundamental solution of the* ***elliptic partial differential equation*** *(10.25) which is characterized by the condition (10.27), namely, when for any continuous bounded function we have*

$$\varphi(s, x, t) = \int_{z \in \mathbb{R}^n} \varphi(u, z, t)\, p(s, x, u, z)\, dz$$

Resulting from the multi-dimensional Taylor series expansion, the following generalization of the Kolmogorov equation (10.25) for continuous-time stochastic Markov processes $\{x(t, \omega)\}_{t \in [t_0, T]}$, taking values in $\mathbb{R}^n$, seems to be evident:

$$\boxed{-\frac{\partial}{\partial s} p(s, x, t, y) = a^{\mathsf{T}}(s, x) \frac{\partial}{\partial x} p(s, x, t, y) + \frac{1}{2} \operatorname{tr} \left\{ b(s, x) \frac{\partial^2}{\partial x^2} p(s, x, t, y) \right\}} \tag{10.28}$$

10.3.2.2 Kolmogorov's (or Fokker–Planck) forward equation

Theorem 10.3. *Suppose that the density $p(s, x, t, y)$ of transition probability for a Markov process, with the drift vector $a(s, x)$ and the diffusion matrix $b(s, x)$, has the derivatives*

$$\frac{\partial}{\partial t} p(s, x, t, y), \quad \frac{\partial}{\partial y} [a(t, y)\, p(s, x, t, y)] \quad \textit{and} \quad \frac{\partial^2}{\partial y^2} [b(t, y)\, p(s, x, t, y)]$$

which are uniformly continuous in y at any finite interval $y' \leq y \leq y''$. Then for any $t \in [a, b]$ and any $y \in \mathbb{R}$ it satisfies the following partial differential equation:

$$\boxed{\begin{aligned}\frac{\partial}{\partial t} p(s, x, t, y) = &-\frac{\partial}{\partial y}[a(t, y) p(s, x, t, y)] \\ &+\frac{1}{2} \frac{\partial^2}{\partial y^2}[b(t, y) p(s, x, t, y)]\end{aligned}} \tag{10.29}$$

*known as **Kolmogorov's** or **the Fokker–Planck forward equation**.*

Proof. Analogously to the proof of the previous theorem and in view of the accepted assumptions, we have that for any twice differential function $\varphi(x)$ the following identity holds:

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{y=-\infty}^{\infty} \varphi(y) p(s, x, t+\Delta t, y) dy - \varphi(x)\right] \\ &= a(t, x) \varphi'(x) + \frac{1}{2} b(t, x) \varphi''(x)\end{aligned}$$

Therefore

$$\begin{aligned}&\frac{\partial}{\partial t} \int_{y=-\infty}^{\infty} \varphi(y) p(s, x, t, y) dy \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{y=-\infty}^{\infty} \varphi(y) p(s, x, t+\Delta t, y) dy - \int_{z=-\infty}^{\infty} \varphi(z) p(s, x, t, z) dz\right] \\ &= \int_{z=-\infty}^{\infty} p(s, x, t, z) \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{y=-\infty}^{\infty} \varphi(y) p(s, x, t+\Delta t, y) dy - \varphi(z)\right] dz \\ &= \int_{z=-\infty}^{\infty} p(s, x, t, z)\left[a(t, z) \varphi'(z) + \frac{1}{2} b(t, z) \varphi''(z)\right] dz\end{aligned}$$

Integrating the right-hand side of this equation by parts implies

$$\begin{aligned}&\frac{\partial}{\partial t} \int_{y=-\infty}^{\infty} \varphi(y) p(s, x, t, y) dy = \int_{y=-\infty}^{\infty} \varphi(y) \frac{\partial}{\partial t} p(s, x, t, y) dy \\ &= \int_{y=-\infty}^{\infty}\left(-\frac{\partial}{\partial y}[a(t, y) p(s, x, t, y)] + \frac{1}{2} \frac{\partial^2}{\partial y^2}[b(t, y) p(s, x, t, y)]\right) \varphi(y) dy\end{aligned}$$

Since the last integral relation holds for any continuous bounded function $\varphi(y)$ the relation (10.29) follows. Theorem is proven. □

As will be shown below, the *solutions of stochastic differential equations are Markov diffusion processes*.

10.4 Markov chains

10.4.1 Main definitions

Let the *phase space* of a Markov process $\{x(t,\omega)\}_{t\in\mathcal{T}}$ *be discrete*, that is,

$$x(t,\omega) \in \mathcal{X} := \{(1,2,\ldots,r) \text{ or } \mathbb{N}\cup\{0\}\}$$
$$\mathbb{N} = 1,2,\ldots \quad \text{is a countable set}$$

Definition 10.6. *A Markov process $\{x(t,\omega)\}_{t\in\mathcal{T}}$ with a discrete phase space $\mathcal{X}$ is said to be a **Markov chain***

(a) ***in continuous time** if*

$$\mathcal{T} := [t_0, T), \quad T \text{ is admitted to be } \infty$$

(b) ***in discrete time** if*

$$\mathcal{T} := \{t_0, t_1, \ldots, t_T\}, \quad T \text{ is admitted to be } \infty$$

The main Markov property (10.9) for this particular case looks as follows: for any $i, j \in \mathcal{X}$ and any $s_1 < \cdots < s_m < s \le t \in \mathcal{T}$

$$\begin{aligned} &\mathsf{P}\{x(t,\omega) = j \mid x(s_1,\omega) = i_i, \ldots, x(s_m,\omega) = i_m, x(s,\omega) = i\} \\ &\quad \overset{a.s.}{=} \mathsf{P}\{x(t,\omega) = j \mid x(s,\omega) = i\} \end{aligned} \tag{10.30}$$

if

$$\mathsf{P}\{x(s_1,\omega) = i_i, \ldots, x(s_m,\omega) = i_m, x(s,\omega) = i\} \neq 0$$

Definition 10.7. *Let*

$$\mathcal{X}_s := \{i \in \mathcal{X} : \mathsf{P}\{x(s,\omega) = i\} \neq 0, \quad s \in \mathcal{T}\}$$

For $s \le t$ $(s,t \in \mathcal{T})$ and $i \in \mathcal{X}_s$, $j \in \mathcal{X}$ define the conditional probabilities

$$\boxed{\pi_{i,j}(s,t) := \mathsf{P}\{x(t,\omega) = j \mid x(s,\omega) = i\}} \tag{10.31}$$

*which we will call the **transition probabilities** of a given Markov chain defining the conditional probability for a process $\{x(t,\omega)\}_{t\in\mathcal{T}}$ to be in the state j at time t under the condition that it was in the state i at time $s < t$.*

Definition 10.7 obviously implies that the function $\pi_{i,j}(s,t)$ (10.31) for any $i \in \mathcal{X}_s$, $j \in \mathcal{X}$ and any $s \le t$ $(s,t \in \mathcal{T})$ should satisfy the following four conditions:

1. $\pi_{i,j}(s,t)$ is a conditional probability, and hence, is nonnegative, that is,

$$\boxed{\pi_{i,j}(s,t) \ge 0} \tag{10.32}$$

2. starting from any state $i \in \mathcal{X}_s$ the Markov chain will obligatorily occur in some state $j \in \mathcal{X}_t$, i.e.,

$$\boxed{\sum_{j \in \mathcal{X}_t} \pi_{i,j}(s,t) = 1} \tag{10.33}$$

3. if no transitions, the chain remains in its starting state with probability 1, that is,

$$\boxed{\pi_{i,j}(s,s) = \delta_{ij}} \tag{10.34}$$

   for any $i, j \in \mathcal{X}_s$, $j \in \mathcal{X}$ and any $s \in \mathcal{T}$;
4. the chain can occur in the state $j \in \mathcal{X}_t$ passing through any intermediate state $k \in \mathcal{X}_u$ $(s \le u \le t)$, i.e.,

$$\boxed{\pi_{i,j}(s,t) = \sum_{k \in \mathcal{X}_u} \pi_{i,k}(s,u)\, \pi_{k,j}(u,t)} \tag{10.35}$$

   The relation (10.35) is known as the *Markov equation*, and in fact, represents the discrete analog of the Chapman–Kolmogorov equation (10.11).

Since for homogeneous Markov chains the transition probabilities $\pi_{i,j}(s,t)$ depend only on the difference $(t - s)$, below we will use the notation

$$\boxed{\pi_{i,j}(t - s) := \pi_{i,j}(s,t)} \tag{10.36}$$

In this case the Markov equation becomes

$$\boxed{\pi_{i,j}(h_1 + h_2) = \sum_{k \in \mathcal{X}} \pi_{i,k}(h_1)\, \pi_{k,j}(h_2)} \tag{10.37}$$

valid for any $h_1, h_2 \ge 0$.

10.4.2 Expectation time before changing a state

10.4.2.1 Exponential law

Consider now a *homogeneous Markov chain* $\{x(t,\omega)\}_{t \in \mathcal{T}}$ with a *discrete phase space* $\mathcal{X}$ and suppose that at time $s \in \mathcal{T}$ it is in a state, say, $x(s,\omega) = i$.

Consider now the time τ (after the time s) just before changing the current state i, i.e., $\tau > s$. By the homogeneity property it follows that distribution function of the time τ_1 (after the time $s_1 := s + u$, $x(s + u, \omega) = i$) is the same as for the τ (after the time s,

$x(s,\omega)=i$) that leads to the following identity:

$$\begin{aligned}&\mathsf{P}\{\tau > v \mid x(s,\omega)=i\} = \mathsf{P}\{\tau_1 > v \mid x(s_1,\omega)=i\}\\ &\mathsf{P}\{\tau > v+u \mid x(s+u,\omega)=i\}\\ &\quad = \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i, \tau > u \geq s\}\end{aligned} \tag{10.38}$$

since the event $\{x(s,\omega)=i, \tau > u\}$ includes as a subset the event $\{x(s+u,\omega)=i\}$.

Lemma 10.5. *The* ***expectation time*** τ *(of the homogeneous Markov chain* $\{x(t,\omega)\}_{t\in\mathcal{T}}$ *with a discrete phase space* $\mathcal{X}$*) to be in the current state* $x(s,\omega)=i$ *before its changing has the exponential distribution*

$$\boxed{\mathsf{P}\{\tau > v \mid x(s,\omega)=i\} = e^{-\lambda_i t}} \tag{10.39}$$

where λ_i *is a nonnegative constant whose inverse value characterizes the average expectation time before the changing the state* $x(s,\omega)=i$*, namely,*

$$\boxed{\frac{1}{\lambda_i} = \mathsf{E}\{\tau \mid x(s,\omega)=i\}} \tag{10.40}$$

The constant λ_i *is usually called the '****exit density****'.*

Proof. Define the function $f_i(u)$ as

$$f_i(u) := \mathsf{P}\{\tau > u \mid x(s,\omega)=i\}$$

By formula (1.68) of the conditional probability we have

$$\begin{aligned}f_i(u+v) &:= \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i\}\\ &= \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i, \tau > u\}\,\mathsf{P}\{\tau > u \mid x(s,\omega)=i\}\\ &= \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i, \tau > u\}\,f_i(u)\end{aligned}$$

Applying (10.38) one has

$$\begin{aligned}f_i(u+v) &:= \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i\}\\ &= \mathsf{P}\{\tau > u+v \mid x(s,\omega)=i, \tau > u\}\,f_i(u)\\ &= \mathsf{P}\{\tau > v \mid x(s,\omega)=i\}\,f_i(u) = f_i(v)\,f_i(u)\end{aligned}$$

which means that

$$\ln f_i(u+v) = \ln f_i(u) + \ln f_i(v) \tag{10.41}$$

such that

$$f_i(0) = \mathsf{P}\{\tau > 0 \mid x(s,\omega)=i\} = 1$$

Differentiation of (10.41) by u gives

$$\frac{f_i'(u+v)}{f_i(u+v)} = \frac{f_i'(u)}{f_i(u)}$$

which for $u = 0$ becomes

$$\frac{f_i'(v)}{f_i(v)} = \frac{f_i'(0)}{f_i(0)} = f_i'(0) := -\lambda_i$$

The last ODE implies

$$f_i(v) = e^{-\lambda_i t}$$

or, equivalently (10.39). To prove (10.40) it is sufficient to notice that

$$\begin{aligned} \mathsf{E}\{\tau \mid x(s,\omega) = i\} &= \int_{t=0}^{\infty} t d[-f_i(t)] \\ &= \left[-te^{-\lambda_i t}\right]_{t=0}^{\infty} - \int_{t=0}^{\infty} \left[-e^{-\lambda_i t}\right] dt = \int_{t=0}^{\infty} e^{-\lambda_i t} dt = \lambda_i^{-1} \end{aligned}$$

Lemma is proven. □

Remark 10.3. *For* ***discrete time Markov chains*** *the relation (10.41) applying for $u = t = 1, 2, \ldots$ and $v = 1$ becomes*

$$\begin{aligned} \ln f_i(t+1) &= \ln f_i(t) + \ln f_i(1) \\ &= \ln f_i(t-1) + 2\ln f_i(1) = \cdots = (t+1)\ln f_i(1) \\ f_i(1) &= 1 - \mathsf{P}\{\tau = 1 \mid x(s,\omega) = i\} \end{aligned}$$

and therefore,

$$\boxed{\begin{aligned} f_i(t) &= [f_i(1)]^t = [1 - \mathsf{P}\{\tau = 1 \mid x(s,\omega) = i\}]^t \\ &= q^t = e^{(\ln q)t} = e^{-|\ln q| t} \\ q &= 1 - \mathsf{P}\{\tau = 1 \mid x(s,\omega) = i\} \end{aligned}} \tag{10.42}$$

10.4.2.2 Returnable (recurrent) and non-returnable (non-recurrent) states

Definition 10.8. *If in (10.40)*

(a)

$$\boxed{\lambda_i = 0} \tag{10.43}$$

or equivalently,

$$\mathsf{E}\{\tau \mid x(s,\omega) = i\} = \infty \tag{10.44}$$

then the state i is called ***absorbing*** *(****non-recurrent****) or* ***null*** *since after some time the process never can not leave this state with probability* 1*;*

(b)

$$\boxed{\lambda_i > 0} \tag{10.45}$$

or equivalently,

$$\mathsf{E}\{\tau \mid x(s,\omega) = i\} < \infty \tag{10.46}$$

then the state i is called ***non-absorbing*** *(****recurrent****) or* ***positive*** *since the process returns to this state infinitely many times with probability* 1.

10.4.3 Ergodic theorem

The result below shows that there exists the class of homogeneous Markov chains, called *ergodic*, which satisfy some additional conditions provided that after a long time such chains '*forget*' the initial states from which they have started.

Theorem 10.4. (The ergodic theorem) *Let for some state $j_0 \in \mathcal{X}$ of a homogeneous Markov chain and some $h > 0$, $\delta \in (0,1)$ for all $i \in \mathcal{X}$*

$$\boxed{\pi_{i,j_0}(h) \geq \delta} \tag{10.47}$$

Then for any initial state distribution $\mathsf{P}\{x(0,\omega) = i\}$ $(i \in \mathcal{X})$ for any $i, j \in \mathcal{X}$ there exists the limit

$$\boxed{p_j^* := \lim_{t\to\infty} \pi_{i,j}(t)} \tag{10.48}$$

such that for any $t \geq 0$ this limit is reachable with an ***exponential rate****, namely,*

$$\boxed{\left|\pi_{i,j}(t) - p_j^*\right| \leq (1-\delta)^{[t/h]} = e^{-\alpha[t/h]}} \tag{10.49}$$

where $\alpha := |\ln(1-\delta)|$ and $[z]$ is the integer part of $z \in \mathbb{R}$.

Proof.

(a) For any $t \geq 0$ define

$$m_j(t) := \inf_{i\in\mathcal{X}} \pi_{i,j}(t) \quad \text{and} \quad M_j(t) := \sup_{i\in\mathcal{X}} \pi_{i,j}(t)$$

which evidently satisfy

$$m_j(t) \le \pi_{i,j}(t) \le M_j(t)$$

for any $i, j \in \mathcal{X}$ and any $t \ge 0$. Show that $m_j(t)$ monotonically increases and $M_j(t)$ monotonically decreases such that

$$M_j(t) - m_j(t) \underset{t\to\infty}{\to} 0 \tag{10.50}$$

since, having (10.50), we obtain (10.48). Using the property (10.37) one has

$$m_j(s+t) := \inf_{i\in\mathcal{X}} \sum_{k\in\mathcal{X}} \pi_{i,k}(s)\pi_{k,j}(t) \ge m_j(t) \inf_{i\in\mathcal{X}} \sum_{k\in\mathcal{X}} \pi_{i,k}(s) = m_j(t)$$
$$M_j(s+t) := \sup_{i\in\mathcal{X}} \sum_{k\in\mathcal{X}} \pi_{i,k}(s)\pi_{k,j}(t) \le M_j(t) \sup_{i\in\mathcal{X}} \sum_{k\in\mathcal{X}} \pi_{i,k}(s) = M_j(t)$$

Next, for $0 \le h \le t$

$$\begin{aligned}
M_j(t) - m_j(t) &= \sup_{i\in\mathcal{X}} \pi_{i,j}(t) + \sup_{l\in\mathcal{X}} \left[-\pi_{l,j}(t)\right] \\
&= \sup_{i,l\in\mathcal{X}} \left[\pi_{i,j}(t) - \pi_{l,j}(t)\right] \\
&= \sup_{i,l\in\mathcal{X}} \sum_{k\in\mathcal{X}} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right]\pi_{k,j}(t-h) \\
&= \sup_{i,l\in\mathcal{X}} \left\{ \sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right]\pi_{k,j}(t-h) \right. \\
&\qquad \left. + \sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right]\pi_{k,j}(t-h) \right\} \\
&\le \sup_{i,l\in\mathcal{X}} \left\{ M_j(t-h) \sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] \right. \\
&\qquad \left. + \, m_j(t-h) \sum_{k\in\mathcal{X}_-} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] \right\}
\end{aligned}$$

Here

$$\mathcal{X}_+ := \left\{k \in \mathcal{X} : \pi_{i,k}(h) - \pi_{l,k}(h) \ge 0\right\}$$
$$\mathcal{X}_- := \left\{k \in \mathcal{X} : \pi_{i,k}(h) - \pi_{l,k}(h) < 0\right\}$$

So, evidently

$$\sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] + \sum_{k\in\mathcal{X}_-} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right]$$

$$= \sum_{k\in\mathcal{X}_+} \pi_{i,k}(h) + \sum_{k\in\mathcal{X}_-} \pi_{i,k}(h) - \sum_{k\in\mathcal{X}_+} \pi_{l,k}(h) - \sum_{k\in\mathcal{X}_-} \pi_{l,k}(h)$$
$$= \sum_{k\in\mathcal{X}} \pi_{i,k}(h) - \sum_{k\in\mathcal{X}} \pi_{l,k}(h) = 1 - 1 = 0$$

and therefore

$$M_j(t) - m_j(t) \le \left[M_j(t-h) - m_j(t-h)\right] \sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] \tag{10.51}$$

Now notice that if $j_0 \notin \mathcal{X}_+$ then

$$\sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] \le \sum_{k\in\mathcal{X}_+} \pi_{i,k}(h) \le 1 - \pi_{i,j_0}(h) \le 1 - \delta$$

and if $j_0 \in \mathcal{X}_+$ then

$$\sum_{k\in\mathcal{X}_+} \left[\pi_{i,k}(h) - \pi_{l,k}(h)\right] \le \sum_{k\in\mathcal{X}_+} \pi_{i,k}(h) - \pi_{i,j_0}(h) \le 1 - \delta$$

Therefore (10.51) leads to

$$M_j(t) - m_j(t) \le (1-\delta)\left[M_j(t-h) - m_j(t-h)\right]$$

Iterating back this inequality $[t/h]$-times and using the estimate

$$M_j(v) - m_j(v) \le 1 \text{ if } v = t - h[t/h]$$

we get

$$M_j(t) - m_j(t) \le (1-\delta)^{[t/h]} \tag{10.52}$$

which proves (10.50) and, consequently, (10.48).
(b) In view of the inequality

$$\left|\pi_{i,j}(t) - p_j^*\right| \le M_j(t) - m_j(t)$$

and using (10.52) we obtain (10.49). Theorem is proven. □

Corollary 10.3. (On a stationary state distribution) *Suppose that (10.47) holds. Then for any $j \in \mathcal{X}$ and for any*

$$\boxed{p_j(t) := \mathsf{P}\{x(t,\omega) = j\}} \tag{10.53}$$

the following property holds

$$\boxed{\left|p_j(t) - p_j^*\right| \le (1-\delta)^{[t/h]}} \tag{10.54}$$

where p_j^* as in (10.48).

Proof. The existence of p_j^* follows from Theorem 10.4, and the formula (10.54) results from

$$\left|p_j(t) - p_j^*\right| = \left|\sum_{i\in\mathcal{X}} \pi_{i,j}(t)\, p_i(0) - p_j^*\right| = \left|\sum_{i\in\mathcal{X}} \left[\pi_{i,j}(t) - p_j^*\right] p_i(0)\right|$$
$$\leq \sum_{i\in\mathcal{X}} \left|\pi_{i,j}(t) - p_j^*\right| p_i(0) \leq (1-\delta)^{[t/h]} \sum_{i\in\mathcal{X}} p_i(0) = (1-\delta)^{[t/h]}$$

which proves the corollary. □

Definition 10.9. *Homogeneous Markov chains, satisfying (10.47), are called* ***ergodic***.[2]

For ergodic Markov chains the following property holds.

Corollary 10.4. *For any $j \in \mathcal{X}$ of an ergodic Markov chain the values p_j^* $(j \in \mathcal{X})$ satisfy*

$$\boxed{p_j^* = \sum_{i\in\mathcal{X}} p_i^* \pi_{i,j}(t)} \tag{10.55}$$

or equivalently, in the vector format

$$\boxed{\begin{array}{c} p^* = \Pi^\intercal(t)\, p^* \\ p^* := \left(p_1^*, \ldots, p_n^*, \ldots\right)^\intercal, \quad \Pi(t) := \left\|\pi_{i,j}(t)\right\|_{i,j\in\mathcal{X}} \end{array}} \tag{10.56}$$

that is, the vector p^* is the eigenvector of the matrix $\Pi^\intercal(t)$ corresponding to its eigenvalue equal to 1.

Proof. By (10.48) we have

$$p_j^* = \lim_{s\to\infty} p_j(s+t) = \lim_{s\to\infty} \sum_{i\in\mathcal{X}} \pi_{i,j}(t)\, p_i(s)$$
$$\geq \lim_{s\to\infty} \sum_{i\leq N} \pi_{i,j}(t)\, p_i(s) = \sum_{i\leq N} \pi_{i,j}(t) \lim_{s\to\infty} p_i(s) = \sum_{i\leq N} \pi_{i,j}(t)\, p_i^*$$

Hence, tending $N \to \infty$ we get

$$p_j^* \geq \sum_{i\leq N} \pi_{i,j}(t)\, p_i^* \tag{10.57}$$

Suppose $p_j^* > \sum_{i\leq N} \pi_{i,j}(t)\, p_i^*$ for some $j = j_*$ and some $t = t_* \geq 0$. But, using (10.48), we have

$$\sum_{i\leq N} p_i^* = \lim_{t\to\infty} \sum_{i\leq \mathcal{N}} \pi_{i,j}(t) \leq 1 \tag{10.58}$$

[2]Information concerning the class of controllable ergodic Markov chains can be found in Poznyak et al. (2000).

Hence, by the supposition

$$\begin{aligned}\sum_{j\in\mathcal{X}} p_j^* &\geq \sum_{j\leq N} p_j^* = \sum_{j\leq N, j\neq j_*} p_j^* + p_{j_*}^* \\ &> \sum_{j\leq N, j\neq j_*} \sum_{i\in\mathcal{X}, j\neq j_*} \pi_{i,j_*}(t_*)\, p_i^* + \sum_{i\in\mathcal{X}, j_*} \pi_{i,j_*}(t_*)\, p_i^* \\ &= \sum_{i\in\mathcal{X}, j\neq j_*} \sum_{j\leq N, j\neq j_*} \pi_{i,j_*}(t_*)\, p_i^* + \sum_{i\in\mathcal{X}, j_*} \pi_{i,j_*}(t_*)\, p_i^* \\ &= \sum_{i\in\mathcal{X}} \sum_{j\leq N} \pi_{i,j}(t_*)\, p_i^* \underset{N\to\infty}{\to} \sum_{i\in\mathcal{X}} \sum_{j\in\mathcal{X}} \pi_{i,j}(t_*)\, p_i^* = \sum_{i\in\mathcal{X}} p_i^*\end{aligned}$$

which leads to a contradiction since $\sum_{j\in\mathcal{X}} p_j^* > \sum_{i\in\mathcal{X}} p_i^*$ and concludes the proof. □

Corollary 10.5. *For any ergodic Markov chain*

(a) or

$$\boxed{\sum_{j\in\mathcal{X}} p_j^* = 1} \tag{10.59}$$

i.e., p_j^* $(j \in \mathcal{X})$ form the probability distribution which is ***unique*** *and called* ***stationary*** *one;*

(b) or

$$\boxed{\sum_{j\in\mathcal{X}} p_j^* = 0} \tag{10.60}$$

i.e., all $p_j^* = 0$ $(j \in \mathcal{X})$.

Proof. The case (b) is evident. Let us prove (a). If $\sum_{j\in\mathcal{X}} p_j^* > 0$, then take

$$p_i(0) := p_i^* / \sum_{l\in\mathcal{X}} p_l^*$$

and, then

$$p_j(t) = \sum_{i\in\mathcal{X}} \pi_{i,j}(t)\, p_i(0) = \sum_{i\in\mathcal{X}} \pi_{i,j}(t)\, p_i^* / \sum_{l\in\mathcal{X}} p_l^* = p_j^* / \sum_{l\in\mathcal{X}} p_l^* = p_j(0)$$

which, by (10.54), implies

$$\lim_{t\to\infty} p_j(t) = p_j^* = p_j(0)$$

Therefore

$$\sum_{j\in\mathcal{X}} p_j^* = \sum_{j\in\mathcal{X}} p_j(0) = 1$$

which proves (10.59). Suppose now that there exist two stationary distributions p^* and p^{**} which, obviously, both satisfy (10.54). Then

$$\begin{aligned}\left|p_j^{**} - p_j^*\right| &= \left|\left[p_j^{**} - p_j(t)\right] + \left[p_j(t) - p_j^*\right]\right| \\ &\le \left|p_j^{**} - p_j(t)\right| + \left|p_j(t) - p_j^*\right| \le 2(1-\delta)^{[t/h]} \underset{t\to\infty}{\to} 0\end{aligned}$$

and hence, $p_j^{**} = p_j^*$ for any $j \in \mathcal{X}$. Corollary is proven. □

11 Stochastic Integrals

Contents

In this section we will study the following most important constructions of stochastic integrals:

- the ***time-integral*** *of a sample path of a second order (s.o.) stochastic process* $\{x(t,\omega)\}_{t\geq 0}$, i.e.,

$$\boxed{I_{[a,b]}(x) = \int_{\tau=a}^{b} x(\tau,\omega)\, d\tau} \tag{11.1}$$

- the so-called ***Wiener integral*** of a deterministic function $f(t)$ with respect to an s.o. stationary orthogonal increment random process $W_\tau(\omega)$, i.e.,

$$\boxed{I_{[a,b]}^{W}(f) = \int_{\tau=a}^{b} f(\tau)\, dW_\tau(\omega)} \tag{11.2}$$

  In this case, $I_t^W(f)$, being associated with the Lebesgue integral, is usually referred to as a stochastic integral with respect to an '*orthogonal random measure*' $x(\tau,\omega) = W_\tau(\omega)$.
- the so-called ***Itô integral*** of a random function $g(t,\omega)$ with respect to an s.o. stationary orthogonal increment random process $W_t(\omega)$, i.e.,

$$\boxed{I_{[a,b]}(g) = \int_{\tau=a}^{b} g(\tau,\omega)\, dW_\tau(\omega)} \tag{11.3}$$

  Obviously, the Wiener integral (11.2) is a partial case of the Itô integral (11.3) when $g(t,\omega) = f(t)$.
- the so-called ***Stratonovich integral*** of a random function $g(t,\omega)$ with respect to an s.o. stationary orthogonal increment random process $W_t(\omega)$, i.e.,

$$\boxed{I_{[a,b]}^{S}(g) = \int_{\tau=a}^{b} g(\tau,\omega) \overset{\lambda=1/2}{\circ} dW_\tau(\omega)} \tag{11.4}$$

where the summation in the right-hand side is taken in a special sense discussed below.

11.1 Time-integral of a sample-path

11.1.1 Integration of a simple stochastic trajectory

Here we will discuss how to understand (11.1)

$$\boxed{I_{[a,b]}(x) = \int_{\tau=a}^{b} x(\tau,\omega)\, d\tau, \quad x(\tau,\omega) \in \mathbb{R}} \tag{11.5}$$

Define the usual partial Darboux sum as

$$y_n(\omega) := \sum_{i=0}^{k_n-1} x\left(t_i^n,\omega\right)\left[t_{i+1}^n - t_i^n\right] \tag{11.6}$$

where the partition $\left\{t_i^n\right\}_{i=\overline{0,k_n}}$ of the time-interval $[a,b]$ is defined as follows:

$$a = t_0^n < t_1^n < \cdots < t_{k_n}^n = b \tag{11.7}$$

such that

$$\delta_n := \max_{i=\overline{0,k_n}} \left|t_{i+1}^n - t_i^n\right| \underset{n\to\infty}{\to} 0 \tag{11.8}$$

Then one has

$$\begin{aligned}
\mathsf{E}\left\{y_n^2(\omega)\right\} &= \mathsf{E}\left\{\sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} x\left(t_i^n,\omega\right) x\left(t_j^n,\omega\right)\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right]\right\} \\
&= \sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} \mathsf{E}\left\{x\left(t_i^n,\omega\right) x\left(t_j^n,\omega\right)\right\}\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right] \\
&= \sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} \rho\left(t_i^n, t_j^n\right)\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right]
\end{aligned}$$

where

$$\rho\left(t_i^n, t_j^n\right) := \mathsf{E}\left\{x\left(t_i^n,\omega\right) x\left(t_j^n,\omega\right)\right\} \tag{11.9}$$

Lemma 11.1. *If* $\{x(t,\omega)\}_{t\geq 0}$ *is a second-order mean-square continuous process, then there exists a unique (in a stochastic sense) mean-square limit of the partial Darboux sums (11.6), i.e., there exists*

$$\boxed{y(\omega) = \text{l.i.m.}\, y_n(\omega) := \int\limits_{\tau=a}^{b} x(\tau,\omega)\, d\tau = I_{[a,b]}(\omega)} \tag{11.10}$$

which is called the time-integral on the interval $[a,b]$ *of a second-order stochastic mean-square continuous process* $\{x(t,\omega)\}_{t\geq 0}$.

Proof. By Lemma 9.1 and considering that $\{x(t,\omega)\}_{t\geq 0}$ is a second-order mean-square continuous process it follows that $\rho\left(t_i^n, t_j^n\right)$ is continuous with respect to both arguments. So

$$\mathsf{E}\left\{y_n^2(\omega)\right\} \to \int\limits_{t=a}^{b}\int\limits_{t=a}^{b} \rho(t,s)\, dt\, ds$$

where the right-hand side is the usual Riemann integral of a continuous function. But this implies that a limit also exists

$$y(\omega) = \text{l.i.m.}\, y_n(\omega) := \int\limits_{\tau=a}^{b} x(\tau,\omega)\, d\tau = I_{[a,b]}(x)$$

or equivalently, $\mathsf{E}\left\{(y_n(\omega) - y(\omega))^2\right\} \underset{n\to\infty}{\to} 0$. To prove the uniqueness assume that two limits exist

$$y(\omega) = \text{l.i.m.}\, y_n(\omega), \quad y'(\omega) = \text{l.i.m.}\, y_n(\omega)$$

such that $\mathsf{E}\left\{\left(y'(\omega) - y(\omega)\right)^2\right\} > 0$. Then it follows that

$$\begin{aligned}
0 \underset{n\to\infty}{\leftarrow} \mathsf{E}\left\{(y_n(\omega) - y(\omega))^2\right\} &= \mathsf{E}\left\{\left(\left[y_n(\omega) - y'(\omega)\right] + \left[y'(\omega) - y(\omega)\right]\right)^2\right\} \\
&= \mathsf{E}\left\{\left(y_n(\omega) - y'(\omega)\right)^2\right\} + \mathsf{E}\left\{\left(y'(\omega) - y(\omega)\right)^2\right\} \\
&\quad + 2\mathsf{E}\left\{\left[y_n(\omega) - y'(\omega)\right]\left[y'(\omega) - y(\omega)\right]\right\} \\
&\underset{n\to\infty}{\to} \mathsf{E}\left\{\left(y'(\omega) - y(\omega)\right)^2\right\} > 0
\end{aligned}$$

since by the Cauchy–Schwartz inequality

$$\begin{aligned}
&\mathsf{E}\left\{\left[y_n(\omega) - y'(\omega)\right]\left[y'(\omega) - y(\omega)\right]\right\} \\
&\quad \leq \mathsf{E}\left\{\left(y_n(\omega) - y'(\omega)\right)^2\right\}\mathsf{E}\left\{\left(y'(\omega) - y(\omega)\right)^2\right\} \underset{n\to\infty}{\to} 0
\end{aligned}$$

which leads to a contradiction. Lemma is proven. □

Remark 11.1. *There exist infinitely many random variables $\tilde{y}(\omega)$ which can be considered as the time-integral of $\{x(t,\omega)\}_{t\geq 0}$ on the interval $[a,b]$ and which differ in terms of $y(\omega)$ (11.10) at the subset $\Omega_0 \subset \Omega$ zero-measure, i.e.,*

$$\mathsf{P}\{\omega \in \Omega_0 \mid \tilde{y}(\omega) \neq y(\omega)\} = 0$$

This means that when we are writing $\int_{\tau=a}^{b} x(\tau,\omega)\,d\tau$ we deal with a family of random variable such that

$$\boxed{\begin{aligned} \mathsf{E}\left\{\int_{\tau=a}^{b} x(\tau,\omega)\,d\tau\right\} &= \int_{\tau=a}^{b} \mathsf{E}\{x(\tau,\omega)\}\,d\tau \\ \mathsf{E}\left\{\left(\int_{\tau=a}^{b} x(\tau,\omega)\,d\tau\right)^2\right\} &= \int_{t=a}^{b}\int_{t=a}^{b} \mathsf{E}\{x(t,\omega)\,x(s,\omega)\}\,dt\,ds \end{aligned}} \tag{11.11}$$

11.1.2 Integration of the product of a random process and a deterministic function

To deal with integrals $\int_{\tau=a}^{b} g(t)\,x(\tau,\omega)\,d\tau$ we may follow the same scheme as above.

Lemma 11.2. *Suppose that*

1. *$\{x(t,\omega)\}_{t\geq 0}$ is a second-order mean-square continuous process;*
2. *$g:[a,b]\to \mathsf{R}$ is a quadratically integrable partially continuous function on $[a,b]$, i.e., $g \in L_2[a,b]$ which means*

$$\int_{t=a}^{b} g^2(t)\,dt < \infty$$

Then there exists a family of random variables

$$\boxed{y(\omega) = \text{l.i.m.}\, y_n(\omega) := \int_{\tau=a}^{b} g(t)\,x(\tau,\omega)\,d\tau} \tag{11.12}$$

where

$$y_n(\omega) := \sum_{i=0}^{k_n-1} g\left(t_i^n\right) x\left(t_i^n,\omega\right)\left[t_{i+1}^n - t_i^n\right]$$

such that

$$\boxed{\begin{aligned} &\mathsf{E}\left\{\int_{\tau=a}^{b} g(\tau)\, x(\tau,\omega)\, d\tau\right\} = \int_{\tau=a}^{b} g(\tau)\, \mathsf{E}\{x(\tau,\omega)\}\, d\tau \\ &\mathsf{E}\left\{\left(\int_{\tau=a}^{b} g(\tau)\, x(\tau,\omega)\, d\tau\right)^2\right\} \\ &\qquad = \int_{t=a}^{b}\int_{t=a}^{b} g(t)\, g(s)\, \mathsf{E}\{x(t,\omega)\, x(s,\omega)\}\, dt\, ds \end{aligned}} \tag{11.13}$$

Proof. Following the proof of the previous lemma we have

$$\begin{aligned} &\mathsf{E}\left\{y_n^2(\omega)\right\} \\ &= \mathsf{E}\left\{\sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} g\left(t_i^n\right) g\left(t_j^n\right) x\left(t_i^n,\omega\right) x\left(t_j^n,\omega\right)\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right]\right\} \\ &= \sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} \mathsf{E}\left\{x\left(t_i^n,\omega\right) x\left(t_j^n,\omega\right)\right\} g\left(t_i^n\right) g\left(t_j^n\right)\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right] \\ &= \sum_{i=0}^{k_n-1}\sum_{j=0}^{k_n-1} \rho\left(t_i^n, t_j^n\right) g\left(t_i^n\right) g\left(t_j^n\right)\left[t_{i+1}^n - t_i^n\right]\left[t_{j+1}^n - t_j^n\right] \\ &\underset{n\to\infty}{\to} \int_{t=a}^{b}\int_{t=a}^{b} g(t)\, g(s)\, \rho(t,s)\, dt\, ds \end{aligned}$$

The right-hand side exists since

$$\begin{aligned} &\int_{t=a}^{b}\int_{t=a}^{b} g(t)\, g(s)\, \rho(t,s)\, dt\, ds \\ &= \int_{t=a}^{b}\int_{t=a}^{b} g(t)\, g(s)\, \mathsf{E}\{x(t,\omega)\, x(s,\omega)\}\, dt\, ds \\ &\le \int_{t=a}^{b}\int_{t=a}^{b} g(t)\, g(s)\sqrt{\mathsf{E}\left\{x^2(t,\omega)\right\}\mathsf{E}\left\{x^2(s,\omega)\right\}} dt\, ds \\ &= \left(\int_{t=a}^{b} g(t)\sqrt{\mathsf{E}\left\{x^2(t,\omega)\right\}} dt\right)^2 \le \int_{t=a}^{b} g^2(t)\, dt \int_{t=a}^{b} \mathsf{E}\left\{x^2(t,\omega)\right\} dt < \infty \end{aligned}$$

Lemma is proven. □

11.2 λ-stochastic integrals

11.2.1 General discussion

Here we will discuss the question how should the integral

$$\int_{\tau=a}^{b} g(\tau,\omega)\, dW_\tau(\omega)$$

be defined. We start with the simplest case when

$$g(t,\omega) = W_t(\omega)$$

where $\{W_t(\omega)\}_{t\geq 0}$ is a standard one-dimensional Wiener process. It seems reasonable to start by defining the integral for random step functions considered as an approximation of $\{W_t(\omega)\}_{t\geq 0}$. Specifically, for any fixed $\lambda \in [0,1]$ the random step function approximation $\varphi_n^\lambda(t,\omega)$ may be defined as

$$\begin{aligned}
&\varphi_n^\lambda(t,\omega) := \lambda W_{t_k^n}(\omega) + (1-\lambda) W_{t_{k-1}^n}(\omega) \\
&t \in \left[t_{k-1}^n, t_k^n\right), \quad k = 0,\ldots,n \\
&a = t_0^n < t_1^n < \cdots < t_{k_n}^n = b
\end{aligned} \tag{11.14}$$

where $\left\{t_k^n\right\}_{k=\overline{0,k_n}}$ is a sequence of partitions of the given interval $[a,b]$, for which the following definition is natural:

$$\begin{aligned}
&\int_{\tau=a}^{b} \varphi_n^\lambda(\tau,\omega)\, dW_\tau(\omega) := \sum_{k=1}^{n} \varphi_n^\lambda(t_{k-1},\omega)\, \Delta W_k^n(\omega) \\
&\Delta W_k^n(\omega) := W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)
\end{aligned} \tag{11.15}$$

Below we will show that taking in (11.8) $\delta_n \to 0$ as $n \to \infty$ we get

$$\boxed{\begin{aligned}
&\int_{\tau=a}^{b} W_\tau(\omega) \overset{\lambda\in[0,1]}{\circ} dW_\tau(\omega) \\
&\qquad := \operatorname{l.i.m} \int_{\tau=a}^{b} \varphi_n^\lambda(\tau,\omega)\, dW_\tau(\omega) \\
&\qquad = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] + (\lambda - 1/2)(b-a)
\end{aligned}} \tag{11.16}$$

that demonstrates that there are infinitely many (indexed by the parameter λ) definitions, using the random step function approximations (11.14), leading to distinct stochastic integrals understood as their mean square limits of the random step function approximations.

The symbol $\binom{\lambda\in[0,1]}{\circ}$ emphasizes the fact that the integral $\int\limits_{\tau=a}^{b} W_\tau(\omega)\, dW_\tau(\omega)$ interpretation may be different, yielding

- for $\lambda = 0$ the so called *Itô (non-anticipating) integral calculus*;
- for $\lambda = 1/2$ the so called *Stratonovich integral calculus*.

Let us discuss all these possibilities in detail.

11.2.2 Variation of the sample path of a standard one-dimensional Wiener process

To begin, the next result has a consequence that $\{W_t(\omega)\}_{t\geq 0}$ is of unbounded variation on any finite-time interval with probability 1 which shows that

the definition of a stochastic integral $\int\limits_{\tau=a}^{b} W_\tau(\omega)\, dW_\tau(\omega)$
as a Stieltjes integral is impossible!

Lemma 11.3. *For* $\Delta W_k^n(\omega)$ *(11.15) and* $\delta_n \underset{n\to\infty}{\to} 0$ *the following property holds:*

$$\boxed{\text{l.i.m} \sum_{k=1}^{n} \left(\Delta W_k^n(\omega)\right)^2 = b - a} \tag{11.17}$$

If, additionally, δ_n *(11.8) satisfies*

$$\sum_{n=0}^{\infty} \delta_n < \infty \tag{11.18}$$

then

$$\boxed{\lim_{n\to\infty} \sum_{k=1}^{n} \left(\Delta W_k^n(\omega)\right)^2 \overset{a.s}{=} b - a} \tag{11.19}$$

Proof. Define

$$S_n(\omega) := \sum_{k=1}^{n} \left(\Delta W_k^n(\omega)\right)^2$$

So, from (9.33) for any n one has

$$\begin{aligned}
\mathsf{E}\{S_n(\omega)\} &= \sum_{k=1}^{n} \mathsf{E}\left\{\left(\Delta W_k^n(\omega)\right)^2\right\} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{\left(W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)\right)^2\right\}
\end{aligned}$$

$$= \sum_{k=1}^{n} \mathsf{E}\left\{W_{t_k^n}^2(\omega) + W_{t_{k-1}^n}^2(\omega) - 2W_{t_k^n}(\omega)\, W_{t_{k-1}^n}(\omega)\right\}$$
$$= \sum_{k=1}^{n} \mathsf{E}\left\{W_{t_k^n}^2(\omega) + W_{t_{k-1}^n}^2(\omega) - 2W_{t_k^n}(\omega)\, W_{t_{k-1}^n}(\omega)\right\}$$
$$= \sum_{k=1}^{n} \left(t_k^n + t_{k-1}^n - 2t_k^n \wedge t_{k-1}^n\right) = \sum_{k=1}^{n} \left(t_k^n - t_{k-1}^n\right) = t_n^n - t_0^n = b - a$$

Now, since $\left(\Delta W_i^n(\omega)\right)^2$ and $\left(\Delta W_j^n(\omega)\right)^2$ are independent if $i \neq j$ and since $\Delta W_i^n(\omega)$ are Gaussian (so, $\mathsf{E}\left\{\left(\Delta W_k^n(\omega)\right)^4\right\} = 3\left(t_k^n - t_{k-1}^n\right)^2$) then one has

$$\begin{aligned} \operatorname{var}\left(S_n(\omega)\right) &= \sum_{k=1}^{n} \operatorname{var}\left(\left(\Delta W_k^n(\omega)\right)^2\right) \\ &= \sum_{k=1}^{n} \left[\mathsf{E}\left\{\left(\Delta W_k^n(\omega)\right)^4\right\} - \left(t_k^n - t_{k-1}^n\right)^2\right] \\ &= 2\sum_{k=1}^{n} \left(t_k^n - t_{k-1}^n\right)^2 \le 2\delta_n (b-a) \to 0 \text{ as } \delta_n \to 0 \end{aligned} \tag{11.20}$$

Therefore, the first part of the lemma is verified. Then, by Fatou's Lemma 6.4 and in view of (11.20) we have

$$\begin{aligned} \mathsf{E}\left\{\sum_{n=1}^{\infty} \left(S_n(\omega) - \mathsf{E}\{S_n(\omega)\}\right)^2\right\} &= \mathsf{E}\left\{\liminf_{N\to\infty} \sum_{n=1}^{N} \left(S_n(\omega) - \mathsf{E}\{S_n(\omega)\}\right)^2\right\} \\ &\le \liminf_{N\to\infty} \mathsf{E}\left\{\sum_{n=1}^{N} \left(S_n(\omega) - \mathsf{E}\{S_n(\omega)\}\right)^2\right\} \\ &= \sum_{n=1}^{\infty} \operatorname{var}\left(S_n(\omega)\right) \le 2(b-a) \sum_{n=1}^{\infty} \delta_n < \infty \end{aligned}$$

and, hence, by Theorem 6.8

$$\mathsf{P}\left\{\omega : \sum_{n=1}^{\infty} \left(S_n(\omega) - \mathsf{E}\{S_n(\omega)\}\right)^2 < \infty\right\} = 1$$

which completes the proof. □

Corollary 11.1. *For m-dimensional standard Wiener process $\{W_t(\omega)\}_{t\ge 0}$ it follows that*

$$\boxed{\begin{aligned} \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \sum_{k=1}^{n} \Delta W_k^n(\omega) \left[\Delta W_k^n(\omega)\right]^{\mathsf{T}} &= (b-a)\, I_{m\times m} \\ \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \sum_{k=1}^{n} \left\|\Delta W_k^n(\omega)\right\|^2 &= m(b-a) \end{aligned}} \tag{11.21}$$

Proof. The first formula in (11.21) is evident. The second one is obtained by taking the trace of the first relation. □

Corollary 11.2. *If δ_n satisfies (11.18) then the sample paths of $\{W_t(\omega)\}_{t\geq 0}$ are of* ***unbounded variation*** *on every finite interval with probability* 1*, that is,*

$$\boxed{\sum_{k=1}^{n} \left\| \Delta W_k^n(\omega) \right\| \underset{n\to\infty}{\overset{a.s.}{\to}} \infty} \tag{11.22}$$

Proof. Since

$$\sum_{k=1}^{n} \left\| \Delta W_k^n(\omega) \right\|^2 \leq \max_{k=1,\dots,n} \left\| \Delta W_k^n(\omega) \right\| \sum_{k=1}^{n} \left\| \Delta W_k^n(\omega) \right\|$$

and, due to uniformity of the partition and the continuity of sample paths,

$$\max_{k=1,\dots,n} \left\| \Delta W_k^n(\omega) \right\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

for large enough n it follows that

$$\sum_{k=1}^{n} \left\| \Delta W_k^n(\omega) \right\| \geq \frac{\sum\limits_{k=1}^{n} \left\| \Delta W_k^n(\omega) \right\|^2}{\max\limits_{k=1,\dots,n} \left\| \Delta W_k^n(\omega) \right\|} \overset{a.s.}{=} \frac{b-a+o_\omega(1)}{\max\limits_{k=1,\dots,n} \left\| \Delta W_k^n(\omega) \right\|} \underset{n\to\infty}{\overset{a.s.}{\to}} \infty$$

which establishes the corollary. □

The last corollary exactly confirms the fact that the definition of a stochastic integral $\int\limits_{\tau=a}^{b} W_\tau(\omega)\, dW_\tau(\omega)$ as a Stieltjes integral is impossible.

11.2.3 Mean square λ-approximation

Theorem 11.1. *As (11.8) holds, that is, as*

$$\boxed{\delta_n := \max_{i=\overline{0,k_n}} \left| t_{i+1}^n - t_i^n \right| \underset{n\to\infty}{\to} 0}$$

then

$$\boxed{\begin{aligned} &\underset{\delta_n \underset{n\to\infty}{\to} 0}{\text{l.i.m}} \sum_{k=1}^{n} \varphi_n^\lambda(t_{k-1},\omega)\, \Delta W_k^n(\omega) \\ &\quad = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] + \left(\lambda - \frac{1}{2}\right)(b-a) \end{aligned}} \tag{11.23}$$

where

$$\varphi_n^\lambda(t,\omega) := \lambda W_{t_k^n}(\omega) + (1-\lambda)\, W_{t_{k-1}^n}(\omega)$$

$$t \in \left[t_{k-1}^n, t_k^n\right), \quad k = 0,\dots,n, \quad a = t_0^n < t_1^n < \cdots < t_{k_n}^n = b$$

Proof.

(a) *Case* $\lambda = 0$. We have

$$\sum_{k=1}^{n} \varphi_n^{\lambda}(t_{k-1}, \omega)\, \Delta W_k^n(\omega) = \sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\, \Delta W_k^n(\omega)$$

and

$$\begin{aligned}
\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\, \Delta W_k^n(\omega) &= \sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\left[W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)\right] \\
&= \sum_{k=1}^{n}\left[W_{t_{k-1}^n}(\omega) - W_{t_k^n}(\omega) + W_{t_k^n}(\omega)\right] \\
&\quad \times \left[W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)\right] \\
&= \sum_{k=1}^{n} W_{t_k^n}(\omega)\, \Delta W_k^n(\omega) - \sum_{k=1}^{n}\left(\Delta W_k^n(\omega)\right)^2
\end{aligned} \tag{11.24}$$

Also one has

$$\begin{aligned}
\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\, \Delta W_k^n(\omega) &= \sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\left[W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)\right] \\
&= \sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\, W_{t_k^n}(\omega) - \sum_{k=1}^{n} W_{t_{k-1}^n}^2(\omega) \\
&= -\sum_{k=1}^{n} \Delta W_k^n(\omega)\, W_{t_k^n}(\omega) + \sum_{k=1}^{n} W_{t_k^n}^2(\omega) \\
&\quad - \sum_{k=1}^{n} W_{t_{k-1}^n}^2(\omega)
\end{aligned} \tag{11.25}$$

Adding both sides of (11.24) and (11.25) we obtain

$$\begin{aligned}
2\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\, \Delta W_k^n(\omega) &= \sum_{k=1}^{n} W_{t_k^n}^2(\omega) - \sum_{k=1}^{n} W_{t_{k-1}^n}^2(\omega) \\
&= W_b^2(\omega) - W_a^2(\omega) - \sum_{k=1}^{n}\left(\Delta W_k^n(\omega)\right)^2
\end{aligned} \tag{11.26}$$

Then taking into account the property (11.17) it follows that

$$\underset{\delta_n \underset{n\to\infty}{\to} 0}{\text{l.i.m}} \sum_{k=1}^{n} \varphi_n^{\lambda=0}(t_{k-1}, \omega)\, \Delta W_k^n(\omega) = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] - \frac{1}{2}(b-a) \tag{11.27}$$

(b) Case $\lambda = 1$. Notice that

$$W_{t_k^n}(\omega)\, \Delta W_k^n(\omega) = W_{t_{k-1}^n}(\omega)\, \Delta W_k^n(\omega) + \left(\Delta W_k^n(\omega)\right)^2$$

Again, by the property (11.17) and in view of (11.27) it follows that

$$\begin{aligned}
&\underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \sum_{k=1}^{n} \varphi_n^{\lambda=1}(t_{k-1},\omega)\,\Delta W_k^n(\omega) \\
&\quad = \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \sum_{k=1}^{n} \varphi_n^{\lambda=0}(t_{k-1},\omega)\,\Delta W_k^n(\omega) + \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \sum_{k=1}^{n} \left(\Delta W_k^n(\omega)\right)^2 \\
&\quad = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] - \frac{1}{2}(b-a) + (b-a) \\
&\quad = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] + \frac{1}{2}(b-a)
\end{aligned} \tag{11.28}$$

The theorem now follows by adding (11.27), multiplied by $(1-\lambda)$, and (11.28), multiplied by λ. □

11.2.4 Itô (non-anticipating) case ($\lambda = 0$)

When the random variable

$$\varphi_n^{\lambda}(t,\omega) := \lambda W_{t_k^n}(\omega) + (1-\lambda)\,W_{t_{k-1}^n}(\omega)$$

is measurable with respect to the σ-algebra, generated by $W_{t_{k-1}^n}(\omega)$, we deal with the so-called *non-anticipating* case when, in fact $\varphi_n^{\lambda}(t,\omega)$ is independent of $\Delta W_k^n(\omega)$. This occurs only when

$$\boxed{\lambda = 0} \tag{11.29}$$

which leads to the *Itô case* with the random step function

$$\varphi_n^{\lambda=0}(t,\omega) = W_{t_{k-1}^n}(\omega), \quad t_{k-1}^n \le t < t_k^n$$

in (11.23), namely,

$$\boxed{\begin{aligned}
\int_{\tau=a}^{b} W_\tau(\omega)\,dW_\tau(\omega) &:= \int_{\tau=a}^{b} W_\tau(\omega) \overset{\lambda=0}{\circ} dW_\tau(\omega) \\
&= \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \int_{\tau=a}^{b} \varphi_n^{\lambda=0}(\tau,\omega)\,dW_\tau(\omega) \\
&= \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] - \frac{(b-a)}{2}
\end{aligned}} \tag{11.30}$$

Remark 11.2. *Notice that (11.30) holds with probability* 1 *if we take* $\{\delta_n\}_{n\ge1}$ *such that* $\sum_{n=1}^{\infty} \delta_n < \infty$.

To conclude this section the following important properties of the integral (11.30) are verified below.

Lemma 11.4. *The integral (11.30), interpreted in the Itô sense, satisfies*

$$\boxed{\mathsf{E}\left\{\int_{\tau=a}^{b} W_{\tau}(\omega)\, dW_{\tau}(\omega)\right\} = 0} \tag{11.31}$$

and

$$\boxed{\mathsf{E}\left\{\left[\int_{\tau=a}^{b} W_{\tau}(\omega)\, dW_{\tau}(\omega)\right]^2\right\} = \frac{\left(b^2 - a^2\right)}{2}} \tag{11.32}$$

Proof.
(a) Since $W_{t_{k-1}^n}(\omega)$ and $\Delta W_k^n(\omega)$ are independent with mean-zero for any k and n, one has

$$\mathsf{E}\left\{\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\,\Delta W_k^n(\omega)\right\} = \sum_{k=1}^{n} \mathsf{E}\left\{W_{t_{k-1}^n}(\omega)\right\} \mathsf{E}\left\{\Delta W_k^n(\omega)\right\} = 0$$

Then (11.31) follows proceeding to the limit (in the mean square sense) when $\delta_n \underset{n\to\infty}{\to} 0$.
(b) One has

$$\left[\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\,\Delta W_k^n(\omega)\right]^2 = \sum_{k=1}^{n}\left[W_{t_{k-1}^n}(\omega)\right]^2\left[\Delta W_k^n(\omega)\right]^2 + 2\sum_{\substack{k=1\\k<j}}^{n}\sum_{j=1}^{n} W_{t_{k-1}^n}(\omega)\,\Delta W_k^n(\omega)\, W_{t_{j-1}^n}(\omega)\,\Delta W_j^n(\omega)$$

Taking into account that $W_{t_{k-1}^n}(\omega)\,\Delta W_k^n(\omega)\, W_{t_{j-1}^n}(\omega)$ and $\Delta W_j^n(\omega)$ are independent with the latter having mean-zero, it follows that

$$\begin{aligned}
\mathsf{E}\left\{\left[\sum_{k=1}^{n} W_{t_{k-1}^n}(\omega)\,\Delta W_k^n(\omega)\right]^2\right\} &= \sum_{k=1}^{n} \mathsf{E}\left\{\left[W_{t_{k-1}^n}(\omega)\right]^2\right\} \mathsf{E}\left\{\left[\Delta W_k^n(\omega)\right]^2\right\} \\
&= \sum_{k=1}^{n} t_{k-1}^n\left(t_{k+1}^n - t_k^n\right) \underset{\delta_n \underset{n\to\infty}{\to} 0}{\to} \int_a^b t\,dt \\
&= \frac{\left(b^2 - a^2\right)}{2}
\end{aligned}$$

which gives (11.32). □

11.2.5 Stratonovich case ($\lambda = 1/2$)

Taking in (11.23)

$$\boxed{\lambda = 1/2} \tag{11.33}$$

the random step function becomes

$$\varphi_n^{\lambda=0}(t,\omega) = \frac{1}{2}\left[W_{t_k^n}(\omega) + W_{t_{k-1}^n}(\omega)\right], \quad t_{k-1}^n \le t < t_k^n$$

and therefore

$$\boxed{\begin{aligned} &\int_{\tau=a}^{b} W_\tau(\omega) \overset{\lambda=1/2}{\circ} dW_\tau(\omega) \\ &\quad = \underset{\substack{\delta_n \to 0 \\ n\to\infty}}{\text{l.i.m}} \int_{\tau=a}^{b} \varphi_n^{\lambda=1/2}(\tau,\omega)\, dW_\tau(\omega) = \frac{1}{2}\left[W_b^2(\omega) - W_a^2(\omega)\right] \end{aligned}} \tag{11.34}$$

which indicates that the corresponding stochastic calculus in this case will be analogous to the Riemann–Stieltjes calculus.

11.3 The Itô stochastic integral

In this section we will introduce the Itô stochastic integral for a fairly general class of non-anticipation random functions in such a way as to preserve the basic Wiener process properties such as (11.31) and (11.32). The properties of the developing integral will follow the lines of general Lebesgue-type integration theory in such a way that, for a fixed parameter interval, this integral will be defined as a continuous linear random function (or, in other words, linear mapping) transforming a space of random functions (complete metric space) into a space of random variables.

11.3.1 The class of quadratically integrable random non-anticipating step functions

Let $\{W_t(\omega)\}_{t\ge 0}$ be a standard Wiener process defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and let $\{\mathcal{F}(t) : t \in [a,b]\}$ be a family (*flow*) of sub-σ-algebras of $\mathcal{F}$ satisfying

•

$$\mathcal{F}(t_1) \subseteq \mathcal{F}(t_2) \quad \text{if } t_1 \le t_2 \tag{11.35}$$

• $W_t(\omega)$ is $\mathcal{F}(t)$-measurable, i.e.,

$$W_t(\omega) \in \mathcal{F}(t) \tag{11.36}$$

• For any $s > 0$

$$\left[W_{t+s}(\omega) - W_t(\omega)\right] \text{ is independent of } \mathcal{F}(t) \tag{11.37}$$

Definition 11.1. *A random function $f(t,\omega)$ is said to be* ***non-anticipating*** *if $f(t,\omega)$ is $\mathcal{F}(t)$-measurable.*

Definition 11.2. *Denote by $\mathcal{L}^2[a,b]$ the class of random functions $f(t,\omega)$ satisfying the conditions:*

1. $f(t,\omega)$ is measurable on $[a,b]\times\Omega$;
2. $f(t,\omega)$ is non-anticipating, that is, $f(t,\omega)$ is $\mathcal{F}(t)$-measurable;
3.

$$\int_{t=a}^{b} \mathsf{E}\left\{f^2(t,\omega)\right\} dt < \infty \tag{11.38}$$

Let also $\mathcal{E}$ denotes de subclass of $\mathcal{L}^2[a,b]$ consisting of the random step functions, namely, $\mathcal{E}$ is the class of random functions $f(t,\omega)$ satisfying the conditions 1–3 given above together with

$$f(t,\omega) = f(t_i,\omega), \quad t_i \le t < t_{i+1} \tag{11.39}$$

for some partition $a = t_0 < t_1 < \cdots < t_n = b$.

The space $\mathcal{L}^2[a,b]$ equipped with the scalar product

$$\boxed{\langle f, g\rangle := \int_{t=a}^{b} \mathsf{E}\{f(t,\omega)\, g(t,\omega)\}\, dt} \tag{11.40}$$

and the corresponding norm

$$\boxed{\|f\|_{\mathcal{L}^2[a,b]} := \left(\int_{t=a}^{b} \mathsf{E}\left\{f^2(t,\omega)\right\} dt\right)^{1/2}} \tag{11.41}$$

forms a real **Hilbert space**.

Lemma 11.5. *$\mathcal{E}$ is dense in $\mathcal{L}^2[a,b]$.*

Proof. Proving the lemma requires that for any $f \in \mathcal{L}^2[a,b]$ there exists a sequence $\{f_n\} \subseteq \mathcal{E}$ such that

$$\|f_n - f\|_{\mathcal{L}^2[a,b]} \underset{n\to\infty}{\to} 0 \tag{11.42}$$

(a) First, by setting $f(t,\omega) \equiv 0$ for all $\omega \in \Omega$ if $t \neq [a,b]$, we may consider $f(t,\omega)$ on the entire line $(-\infty,\infty)$. Then by Fubini's theorem and the property 3 (11.38) it follows that

$$\mathsf{E}\left\{\int_{t=-\infty}^{\infty} f^2(t,\omega)\,dt\right\} = \int_{t=-\infty}^{\infty} \mathsf{E}\left\{f^2(t,\omega)\right\}dt < \infty$$

so that with probability 1,

$$\int_{t=-\infty}^{\infty} f^2(t,\omega)\,dt \overset{a.s.}{<} \infty$$

Now, by the application of the dominated convergence Theorem 6.2, it follows that

$$\mathsf{E}\left\{\int_{t=-\infty}^{\infty} |f(t+h,\omega) - f(t,\omega)|^2\,dt\right\} \underset{h\to 0}{\to} 0 \tag{11.43}$$

This results from the following consideration:

$$\int_{t=-\infty}^{\infty} |f(t+h,\omega) - f(t,\omega)|^2\,dt \leq 4\int_{t=-\infty}^{\infty} f^2(t,\omega)\,dt < \infty \tag{11.44}$$

which, by the dominated convergence Theorem 6.2 for the Lebesgue integration along the sample paths, implies

$$\int_{t=-\infty}^{\infty} |f(t+h,\omega) - f(t,\omega)|^2\,dt \underset{h\to 0}{\overset{a.s.}{\to}} 0 \tag{11.45}$$

Then, by (11.44), it follows that

$$\mathsf{E}\left\{\int_{t=-\infty}^{\infty} |f(t+h,\omega) - f(t,\omega)|^2\,dt\right\} \leq 4\mathsf{E}\left\{\int_{t=-\infty}^{\infty} f^2(t,\omega)\,dt\right\} < \infty$$

So the last expression permits the application of the dominated convergence Theorem 6.2 in the probability space yielding (11.43).
(b) Define then the functions

$$\varphi_j(t) := [jt]/t$$

where $j = 1, 2, \ldots$ and $[\cdot]$ means the 'greatest integer in' function. Since

$$\varphi_j(t) \to t \text{ as } j \to \infty$$

the expression (11.43) can be replaced by

$$\mathsf{E}\left\{\int_{s=-\infty}^{\infty}\left|f\left(s+\varphi_j(t),\omega\right)-f(s+t,\omega)\right|^2 ds\right\}<\infty$$

for any fixed t. Again, the application of the dominated convergence Theorem 6.2 together with Fubini's theorem yields

$$\int_{s=-\infty}^{\infty}\int_{t=a-1}^{b}\mathsf{E}\left\{\left|f\left(s+\varphi_j(t),\omega\right)-f(s+t,\omega)\right|^2\right\}dt\,ds$$
$$=\int_{t=a-1}^{b}\int_{s=-\infty}^{\infty}\mathsf{E}\left\{\left|f\left(s+\varphi_j(t),\omega\right)-f(s+t,\omega)\right|^2\right\}dsdt\underset{j\to\infty}{\to}0$$

Hence, there exists a sub-sequence $\{j_n\}$ such that

$$\int_{t=a-1}^{b}\mathsf{E}\left\{\left|f\left(s+\varphi_{j_n}(t),\omega\right)-f(s+t,\omega)\right|^2\right\}dt\underset{n\to\infty}{\to}0$$

for almost all s. Fixing then such $s\in[0,1]$ and replacing t by $(t-s)$ in the last formula, one has

$$\int_{t=a-1+s}^{b+s}\mathsf{E}\left\{\left|f\left(s+\varphi_{j_n}(t-s),\omega\right)-f(t,\omega)\right|^2\right\}dt\underset{n\to\infty}{\to}0$$

which in view of the property

$$f(t,\omega)\equiv 0 \quad \text{for all} \quad \omega\in\Omega \text{ if } t\neq[a,b]$$

and the identity

$$\int_{t=a-1+s}^{b+s}(\cdot)=\underbrace{\int_{t=a-1+s}^{a}(\cdot)}_{0}+\int_{t=a}^{b}(\cdot)+\underbrace{\int_{t=b}^{b+s}(\cdot)}_{0}=\int_{t=a}^{b}(\cdot)$$

implies

$$\int_{t=a}^{b}\mathsf{E}\left\{\left|f\left(s+\varphi_{j_n}(t-s),\omega\right)-f(t,\omega)\right|^2\right\}dt\underset{n\to\infty}{\to}0 \tag{11.46}$$

Finally, taking

$$f_n(t,\omega):=f\left(s+\varphi_{j_n}(t-s),\omega\right)$$

we have that $f_n(t,\omega)\in\mathcal{E}$, and therefore, (11.46) states (11.42). The proof is complete. □

Denoting now by $L^2(\Omega)$ the usual Hilbert space of random variables f on $(\Omega, \mathcal{F}, \mathsf{P})$ with finite second moments and norm

$$\|f\|_2 := \left(\mathsf{E}\left\{f^2\right\}\right)^{1/2} < \infty \tag{11.47}$$

we may define the *Itô integral* $I_{[a,b]}(\omega)$ on $\mathcal{E}$ as

$$\boxed{I_{[a,b]}(f) := \sum_{k=1}^{n} f(t_{k-1}, \omega) \Delta W_k^n(\omega)} \tag{11.48}$$

where $f(t, \omega) = f(t_{k-1}, \omega)$ for $t_{k-1} \le t < t_k$ is in $\mathcal{E}$ and $\Delta W_k^n(\omega) := W_{t_k^n}(\omega) - W_{t_{k-1}^n}(\omega)$ as in (11.15).

Definition 11.3. *The* ***Itô integral*** $I_{[a,b]}(f)$ *on* $\mathcal{L}^2[a,b]$ *is defined as the extension of (11.48) given on* $\mathcal{E}$ *when* $n \to \infty$ *as a continuous linear random functional action from* $\mathcal{L}^2[a,b]$ *on* $L^2(\Omega)$.

The next theorem constitutes the central result of this section concerning the main properties of the Itô integral.

Theorem 11.2. *The Itô integral*

$$\boxed{I_{[a,b]}(f) = \int_{\tau=a}^{b} f(\tau, \omega)\, dW_\tau(\omega)} \tag{11.49}$$

on $\mathcal{L}^2[a,b]$ *satisfies*

$$\boxed{\mathsf{E}\left\{\int_{\tau=a}^{b} f(\tau, \omega)\, dW_\tau(\omega)\right\} = 0} \tag{11.50}$$

and

$$\boxed{\begin{aligned}\left\|I_{[a,b]}(f)\right\|_2 &= \left(\mathsf{E}\left\{\left(\int_{\tau=a}^{b} f(\tau, \omega)\, dW_\tau(\omega)\right)^2\right\}\right)^{1/2} \\ &= \left(\int_{\tau=a}^{b} \mathsf{E}\left\{f^2(\tau, \omega)\right\} d\tau\right)^{1/2} = \|f\|_{\mathcal{L}^2[a,b]}\end{aligned}} \tag{11.51}$$

Proof. For $f(t,\omega) \in \mathcal{E}$ $I_{[a,b]}(\omega)$ (11.48) is obviously linear random functional. It also follows that

$$\begin{aligned}
\left\| I_{[a,b]}(f) \right\|_2^2 &= \mathsf{E}\left\{ \left(\sum_{k=1}^{n} f(t_{k-1},\omega)\,\Delta W_k^n(\omega) \right)^2 \right\} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{ f^2(t_{k-1},\omega) \right\} \mathsf{E}\left\{ \left[\Delta W_k^n(\omega)\right]^2 \right\} \\
&= \sum_{k=1}^{n} \mathsf{E}\left\{ f^2(t_{k-1},\omega) \right\} \Delta t_k = \|f\|^2_{\mathcal{L}^2[a,b]}
\end{aligned}$$

since $f(t_{k-1},\omega)$ and $\Delta W_k^n(\omega)$ are independent and $\mathsf{E}\left\{\Delta W_k^n(\omega)\right\} = 0$.

Let now $f(t,\omega) \in \mathcal{L}^2[a,b]$. Choosing $\{f_n\} \subseteq \mathcal{E}$ such that $f_n \to f$ whereas $n \to \infty$, and since $\{f_n\}$ is a Cauchy sequence (because of the density property of $\mathcal{E} \subseteq \mathcal{L}^2[a,b]$) one can define

$$\int_{\tau=a}^{b} f(\tau,\omega)\,dW_\tau(\omega) := \lim_{n\to\infty} \int_{\tau=a}^{b} f_n(\tau,\omega)\,dW_\tau(\omega)$$

which completes the proof. □

Remark 11.3. *The properties (11.50) and (11.51) of the Itô stochastic integral are precisely the extensions of the properties (11.31) and (11.32), respectively, of the integral* $\int_{\tau=a}^{b} W_\tau(\omega)\,dW_\tau(\omega)$ *when in* $\int_{\tau=a}^{b} f(\tau,\omega)\,dW_\tau(\omega)$ *one takes* $f(\tau,\omega) = W_\tau(\omega)$.

11.3.2 The Itô stochastic integral as the function of the upper limit

Definition 11.4. *Let $\mathcal{B} \subseteq [a,b]$ be a Borel set. Then the* ***generalized Itô stochastic integral*** $\int_{\tau\in\mathcal{B}} f(\tau,\omega)\,dW_\tau(\omega)$ *may be defined as follows:*

$$\boxed{\int_{\tau\in\mathcal{B}} f(\tau,\omega)\,dW_\tau(\omega) := \int_{\tau=a}^{b} \chi(t\in\mathcal{B})\, f(\tau,\omega)\,dW_\tau(\omega)} \tag{11.52}$$

where $\chi(t\in\mathcal{B})$ is the characteristic function of the event $\{t\in\mathcal{B}\}$.

Claim 11.1. *The following* ***additivity property*** *seems to be evident:*

$$\boxed{\begin{aligned}
&\int_{\tau\in\mathcal{B}_1\cup\mathcal{B}_2} f(\tau,\omega)\,dW_\tau(\omega) \\
&\quad = \int_{\tau\in\mathcal{B}_1} f(\tau,\omega)\,dW_\tau(\omega) + \int_{\tau\in\mathcal{B}_2} f(\tau,\omega)\,dW_\tau(\omega)
\end{aligned}} \tag{11.53}$$

for any Borel disjoint subsets $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{B}$ such that $\mathcal{B}_1 \cap \mathcal{B}_2 = \varnothing$.

Proof. It results directly from the additivity property of the sum in (11.48). □

Corollary 11.3. *For any $t_1, t_2 \in [a, b]$ such that $a \le t_1 < t_2 \le b$ one has*

$$\boxed{\int_{\tau=a}^{t_2} f(\tau,\omega)\, dW_\tau(\omega) = \int_{\tau=a}^{t_1} f(\tau,\omega)\, dW_\tau(\omega) + \int_{\tau=t_1}^{t_2} f(\tau,\omega)\, dW_\tau(\omega)} \tag{11.54}$$

Setting

$$\boxed{x(t,\omega) := \int_{\tau=a}^{t} f(\tau,\omega)\, dW_\tau(\omega)} \tag{11.55}$$

the relation (11.54) becomes

$$x(t_2,\omega) = x(t_1,\omega) + \int_{\tau=t_1}^{t_2} f(\tau,\omega)\, dW_\tau(\omega)$$

which shows that $\{x(t,\omega)\}_{t\in[a,b]}$ is a *Markov process.*

Lemma 11.6. *The stochastic process $\{x(t,\omega)\}_{t\in[a,b]}$ given by (11.55) is a martingale with respect to the flow of sub-σ-algebras $\{\mathcal{F}(t)\}_{t\in[a,b]}$ satisfying (11.35)–(11.37), that is, for any $s > 0$ ($t + s \in [a, b]$) and $t \in [a, b]$*

$$\boxed{\mathsf{E}\{x(t+s,\omega) \mid \mathcal{F}(t)\} \overset{a.s.}{=} x(t,\omega)} \tag{11.56}$$

Proof.

(a) First, notice that $\mathsf{E}\{|x(t,\omega)|\} < \infty$ follows from $x(t,\omega) \in L^2(\Omega)$. Since $[W_{t+s}(\omega) - W_t(\omega)]$ is independent of $\mathcal{F}(t)$, then for step functions $f_n \in \mathcal{E}$ the independence of the increment

$$[x_n(t+s,\omega) - x_n(t,\omega)]$$

of $\mathcal{F}(t)$ with

$$x_n(t,\omega) := \int_{\tau=a}^{t} f_n(\tau,\omega)\, dW_\tau(\omega)$$

and the mean-zero property (11.50) imply

$$\mathsf{E}\{x_n(t+s,\omega) - x_n(t,\omega) \mid \mathcal{F}(t)\} \overset{a.s.}{=} \mathsf{E}\{x_n(t+s,\omega) - x_n(t,\omega)\} = 0 \tag{11.57}$$

And since $x_n(t,\omega)$ is measurable with respect to $\mathcal{F}(t)$ the result (11.56) follows.

(b) For a general process $f(t,\omega) \in \mathcal{L}^2[a,b]$ the stochastic integral $x(t,\omega)$ is also from $L^2(\Omega)$, and hence, there exists a sequence $\{f_n(t,\omega)\}_{n\geq 1}$ of step functions such that $f_n(t,\omega) \to f(t,\omega)$ in $\mathcal{L}^2[a,b]$ and $x_n(t,\omega) \to x(t,\omega)$ in $L^2(\Omega)$. The result follows from (11.57) by passage to the limit. Lemma is proven. □

Corollary 11.4. *The stochastic process* $\left\{|x(t,\omega)|^p\right\}_{t\in[a,b]}$ *with* $p \geq 1$, *given by (11.55) with* $f(t,\omega) \in \mathcal{L}^2[a,b]$, *is a submartingale, that is,*

$$\boxed{\mathsf{E}\left\{|x(t+s,\omega)|^p \mid \mathcal{F}(t)\right\} \overset{a.s.}{\geq} |x(t,\omega)|^p} \tag{11.58}$$

such that for any $r > 0$

$$\boxed{\mathsf{P}\left\{\sup_{t\in[a,b]} |x(t,\omega)| > r\right\} \leq \frac{1}{r^2}\int_{\tau=a}^{b} \mathsf{E}\left\{f^2(\tau,\omega)\right\} d\tau} \tag{11.59}$$

and

$$\boxed{\mathsf{E}\left\{\sup_{t\in[a,b]} |x(t,\omega)|^2\right\} \leq 4\mathsf{E}\left\{\int_{\tau=a}^{b} f^2(\tau,\omega)\,d\tau\right\}} \tag{11.60}$$

Proof. The properties (11.58) and (11.59) follow from (7.28), the Doob inequality (7.71) and the inequality (7.75) for the moments of the maximum modulus for $p = 2$. □

The next sub-section is the main one in this section.

11.3.3 The Itô formula

11.3.3.1 One-dimensional case

Theorem 11.3. (Ito, 1951) *Suppose* $\{x(t,\omega)\}_{t\in[a,b]}$ *has the* ***stochastic differential***

$$\boxed{x(t,\omega) \overset{a.s.}{=} x(a,\omega) + \int_{\tau=a}^{t} f(\tau,\omega)\,d\tau + \int_{\tau=a}^{t} g(\tau,\omega)\,dW_\tau(\omega)} \tag{11.61}$$

or, equivalently, in the ***symbolic form***

$$\boxed{dx(t,\omega) = f(t,\omega)\,dt + g(t,\omega)\,dW_t(\omega)} \tag{11.62}$$

Here $\{f(t,\omega)\}_{t\in[a,b]}$ *is a second-order mean-square continuous process, the first integral in the right-hand side of (11.61) is the usual time-integral (11.10) and the second* *integral is the Itô integral of the function* $g(\tau,\omega) \in \mathcal{L}^2[a,b]$. *If* $V = V(t,x)$ *is a real valued deterministic function defined for all* $t \in [a,b]$ *and* $x \in \mathbb{R}$ *with continuous partial derivatives* $\frac{\partial}{\partial t}V(t,x)$, $\frac{\partial}{\partial x}V(t,x)$ *and* $\frac{\partial^2}{\partial x^2}V(t,x)$, *then the stochastic process* $\{V(t,x(t,\omega))\}_{t\in[a,b]}$ *has a stochastic differential* $dV(t,x(t,\omega))$ *on* $[a,b]$ *given by the symbolic form*

$$\boxed{dV(t,x(t,\omega)) = v_1(t,x(t,\omega))\,dt + v_2(t,x(t,\omega))\,dW_t(\omega)} \tag{11.63}$$

with

$$\boxed{\begin{gathered} v_1(t,x) = \frac{\partial}{\partial t}V(t,x) + \frac{\partial}{\partial x}V(t,x)f + \frac{1}{2}\frac{\partial^2 V(t,x)}{\partial x^2}g^2 \\ v_2(t,x) = \frac{\partial V(t,x)}{\partial x}g \end{gathered}} \tag{11.64}$$

or, equivalently, in the 'open form'

$$\boxed{\begin{aligned} V(t,x(t,\omega)) &\overset{a.s.}{=} V(a,x(a,\omega)) \\ &+ \int_{\tau=a}^{t}\left[\frac{\partial}{\partial\tau}V(\tau,x(\tau,\omega)) + \frac{\partial V(\tau,x(\tau,\omega))}{\partial x}f(\tau,\omega)\,d\tau\right. \\ &+ \left.\frac{1}{2}\frac{\partial^2 V(\tau,x(\tau,\omega))}{\partial x^2}g^2(\tau,\omega)\right]d\tau \\ &+ \int_{\tau=a}^{t}\frac{\partial V(\tau,x(\tau,\omega))}{\partial x}g(\tau,\omega)\,dW_\tau(\omega) \end{aligned}} \tag{11.65}$$

Remark 11.4. *The stochastic differential (11.63) differs from 'usual differential' (obtained by the standard calculus) only as regards the specific term*

$$\boxed{\frac{1}{2}\frac{\partial^2 V(t,x)}{\partial x^2}g^2} \tag{11.66}$$

which is referred to as the ***Itô term****, the expression (11.65) as the* ***Itô formula****.*

Proof. (a) We start with the simplest case when $f(t,\omega)$ and $g(t,\omega)$ are *constants*, i.e.,

$$f(t,\omega) = f_0, \quad g(t,\omega) = g_0 \tag{11.67}$$

Then for a fixed partition (11.7)

$$\begin{aligned} V(t,x(t,\omega)) - V(s,x(s,\omega)) &= \sum_{k=1}^{n}\Delta V_k^n(\omega) \\ \Delta V_k^n(\omega) &:= V\left(t_k^n, x\left(t_k^n,\omega\right)\right) - V\left(t_{k-1}^n, x\left(t_{k-1}^n,\omega\right)\right) \end{aligned}$$

By Taylor's formula there exist constants θ_1 and θ_2 (not exceeding 1) such that

$$\begin{aligned}\Delta V_k^n(\omega) = & \frac{\partial}{\partial t} V\left(t_{k-1}^n + \theta_1 \Delta t_k, x\left(t_{k-1}^n, \omega\right)\right) \Delta t_k \\ & + \frac{\partial}{\partial x} V\left(t_{k-1}^n, x\left(t_{k-1}^n, \omega\right)\right) \Delta x_k \\ & + \frac{1}{2} \frac{\partial^2}{\partial x^2} V\left(t_{k-1}^n, x\left(t_{k-1}, \omega\right) + \theta_2 \Delta x_k\right) (\Delta x_k)^2\end{aligned}$$

where

$$\Delta t_k := \left(t_k^n - t_{k-1}^n\right), \quad \Delta x_k := x\left(t_{k-1}^n, \omega\right)$$

As $\delta_n := \max_{i=\overline{0,k_n}} \left|t_{i+1}^n - t_i^n\right| \underset{n\to\infty}{\to} 0$, by the continuity of $x\left(t_k^n, \omega\right)$, one has

$$\frac{\partial}{\partial t} V\left(t_{k-1}^n + \theta_1 \Delta t_k, x\left(t_{k-1}^n, \omega\right)\right) \overset{a.s.}{\to} \frac{\partial}{\partial t} V\left(t_{k-1}^n, x\left(t_{k-1}^n, \omega\right)\right)$$

$$\frac{\partial^2}{\partial x^2} V(t_{k-1}^n, x\left(t_{k-1}, \omega\right) + \theta_2 \Delta x_k) \overset{a.s.}{\to} \frac{\partial^2}{\partial x^2} V\left(t_{k-1}^n, x\left(t_{k-1}, \omega\right)\right)$$

By (11.67) it follows that

$$\Delta x_k = f_0 \Delta t_k + g_0 \Delta W_k^n(\omega)$$

so that

$$\sum_{k=1}^n \left((\Delta x_k)^2 - \left[g_0 \Delta W_k^n(\omega)\right]^2\right) = f_0^2 \sum_{k=1}^n (\Delta t_k)^2 + 2 f_0 g_0 \sum_{k=1}^n \Delta W_k^n(\omega) \Delta t_k \quad (11.68)$$

In view of the estimates, resulting from (11.19),

$$\sum_{k=1}^n (\Delta t_k)^2 \le \delta_n \sum_{k=1}^n \Delta t_k = \delta_n (b - a)$$

$$\begin{aligned}\left|\sum_{k=1}^n \Delta W_k^n(\omega) \Delta t_k\right| & \le \left[\sum_{k=1}^n \left(\Delta W_k^n(\omega)\right)^2 \sum_{k=1}^n (\Delta t_k)^2\right]^{1/2} \\ & \le \sqrt{\delta_n (b-a)} \left(\sum_{k=1}^n \left(\Delta W_k^n(\omega)\right)^2\right)^{1/2}\end{aligned}$$

it follows that the right-hand side of the identity (11.68) tends to zero with probability 1 which, after taking the limits where $n \to \infty$ ($\delta_n \to 0$, $\sum_{n=1}^{\infty} \delta_n < \infty$), implies

$$\begin{aligned}& V(t, x(t, \omega)) - V(s, x(s, \omega)) \\ & \quad \overset{a.s.}{=} \lim_{\delta_n \underset{n\to\infty}{\to} 0} \sum_{k=1}^n \Delta V_k^n(\omega)\end{aligned}$$

$$= \lim_{\substack{\delta_n \to 0 \\ n \to \infty}} \sum_{k=1}^{n} \left[\frac{\partial}{\partial t} V\left(t_{k-1}^n, x\left(t_{k-1}^n, \omega\right)\right) + \frac{\partial}{\partial x} V\left(t_{k-1}^n, x\left(t_{k-1}^n, \omega\right)\right) f_0 \right.$$
$$\left. + \frac{1}{2}\frac{\partial^2}{\partial x^2} V(t_{k-1}^n, x\left(t_{k-1}, \omega\right) g_0^2 \right] \Delta t_k$$
$$+ \lim_{\substack{\delta_n \to 0 \\ n \to \infty}} \sum_{k=1}^{n} \frac{\partial}{\partial x} V\left(t_{k-1}^n, x\left(t_{k-1}^n, \omega\right)\right) g_0 \Delta W_k^n\left(\omega\right)$$
$$+ \lim_{\substack{\delta_n \to 0 \\ n \to \infty}} \sum_{k=1}^{n} \left[\frac{1}{2}\frac{\partial^2}{\partial x^2} V(t_{k-1}^n, x\left(t_{k-1}, \omega\right) g_0^2 \right] \left[\left(\Delta W_k^n\left(\omega\right)\right)^2 - \Delta t_k \right] \quad (11.69)$$

The first two limits on the right-hand side of (11.69) are the terms on the right-hand side of (11.65). It remains to show that the last limit in (11.69) is zero. Let

$$\beta_k := \left(\Delta W_k^n\left(\omega\right)\right)^2 - \Delta t_k$$

Since $\{\beta_k\}_{k \geq 1}$ is the sequence of independent random variables, for all $k \geq 1$ it follows that

$$\mathsf{E}\left\{\beta_k\right\} = 0, \quad \mathsf{E}\left\{\beta_k^2\right\} = 2\left(\Delta t_k\right)^2$$

Setting $\chi_k\left(N\right)$ the characteristic function of the event

$$\left\{\omega : \left|x\left(t_i, \omega\right)\right| \leq N \text{ for all } i \leq k\right\}$$

we get

$$\mathsf{E}\left\{ \left(\frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \beta_k \right)^2 \right\}$$
$$= \mathsf{E}\left\{ \left(\frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \beta_k \right)^2 \chi_k\left(N\right) \right\}$$
$$+ \mathsf{E}\left\{ \left(\frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \beta_k \right)^2 \left[1 - \chi_k\left(N\right)\right] \right\}$$
$$\leq \mathsf{E}\left\{ \left(\frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \beta_k \right)^2 \chi_k\left(N\right) \right\}$$
$$= \mathsf{E}\left\{ \left(\frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \chi_k\left(N\right) \right)^2 \right\} \mathsf{E}\left\{\beta_k^2\right\}$$

By the continuity property of $\frac{\partial^2 V(t,x)}{\partial x^2}$ on $[a, b] \times [-N, N]$, it follows that

$$\left| \frac{\partial^2 V(t_{k-1}^n, x\left(t_{k-1}, \omega\right)}{\partial x^2} \chi_k\left(N\right) \right| \leq C = \text{const}$$

that's why

$$\underset{\delta_n \underset{n\to\infty}{\to} 0}{\text{l.i.m}} \sum_{k=1}^{n} \mathsf{E}\left\{\left(\frac{\partial^2 V(t_{k-1}^n, x(t_{k-1}, \omega))}{\partial x^2}\beta_k\right)^2\right\} \le 2C \lim_{\delta_n \underset{n\to\infty}{\to} 0} \sum_{k=1}^{n} (\Delta t_k)^2 = 0$$

But, we also have

$$\sum_{n=1}^{\infty}\left[\mathsf{E}\left\{\sum_{k=1}^{n}\left(\frac{\partial^2 V(t_{k-1}^n, x(t_{k-1}, \omega))}{\partial x^2}\beta_k\right)^2\right\}\right]^2 \le 4C^2 (b-a)^2 \sum_{n=1}^{\infty} \delta_n^2 < \infty$$

which, by Theorem 6.8, implies

$$\sum_{k=1}^{n}\left(\frac{\partial^2 V(t_{k-1}^n, x(t_{k-1}, \omega))}{\partial x^2}\beta_k\right)^2 \underset{\delta_n \underset{n\to\infty}{\to} 0}{\to} 0$$

So, the validity of the Itô formula (11.65) for the case of the constant functions f and g is proven.

(b) Obviously the case, when f and g are step functions f_n and g_n, results from the previous one if take into account the additivity property (11.54) of the Itô integral. We have

$$\begin{aligned} V(t, x(t,\omega)) \overset{a.s.}{=} {} & V(a, x(a,\omega)) \\ & + \int_{\tau=a}^{t}\left[\frac{\partial}{\partial \tau}V(\tau, x(\tau,\omega)) + \frac{\partial V(\tau, x(\tau,\omega))}{\partial x} f_n(\tau,\omega)\, d\tau\right. \\ & \left. + \frac{1}{2}\frac{\partial^2 V(\tau, x(\tau,\omega))}{\partial x^2} g_n^2(\tau,\omega)\right] d\tau \\ & + \int_{\tau=a}^{t} \frac{\partial}{\partial x} V(\tau, x(\tau,\omega)) g_n(\tau,\omega)\, dW_\tau(\omega) \end{aligned} \tag{11.70}$$

(c) For the general case suppose that

$$\int_{\tau=a}^{t} |f_n(\tau, x(\tau,\omega)) - f(\tau, x(\tau,\omega))|\, dt \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

$$\int_{\tau=a}^{t} |g_n(\tau, x(\tau,\omega)) - g(\tau, x(\tau,\omega))|\, dt \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

This can be done considering $\sum_{n=1}^{\infty} \delta_n^2 < \infty$ (see Lemma 11.3). Hence, the right-hand side of (11.70) tends with probability 1 to that of (11.65). Theorem is proven. □

Remark 11.5. *The appearance of the additional Itô term is related to the fact that the term* $\left(\Delta W_k^n(\omega)\right)^2$ *behaves like* Δt_k*, or roughly speaking,*

$$(dW_t(\omega))^2 \sim dt$$

The application of the Itô formula (11.65) is illustrated by the corollary given below.

Corollary 11.5. *For any twice continuously differential function $h = h(x)$ and a standard Wiener process $\{W_t(\omega)\}_{t\in[a,b]}$ the following identity for any $t \in [a,b]$ holds:*

$$\boxed{\begin{aligned}\int_{\tau=a}^{t} \frac{\partial h(W_\tau(\omega))}{\partial x} dW_\tau(\omega) &\overset{a.s.}{=} h(W_t(\omega)) - h(W_a(\omega)) \\ &- \frac{1}{2}\int_{\tau=a}^{t} \frac{\partial^2 h(W_\tau(\omega))}{\partial x^2} d\tau\end{aligned}} \tag{11.71}$$

Proof. Define $\{x(t,\omega)\}_{t\in[a,b]}$ as a stochastic process as follows:

$$x(t,\omega) = W_t(\omega)$$

Then its stochastic differential is

$$dx(t,\omega) = dW_t(\omega)$$

that is, in (11.62) $f(t,\omega) \equiv 0$ and $g(t,\omega) \equiv 1$. Then, taking $V(t,x) = h(x)$ and using the identity (11.65) one has

$$\begin{aligned}h(x(t,\omega)) &\overset{a.s.}{=} h(x(a,\omega)) + \int_{\tau=a}^{t} \frac{1}{2}\frac{\partial^2 h(x(\tau,\omega))}{\partial x^2} d\tau \\ &+ \int_{\tau=a}^{t} \frac{\partial h(x(\tau,\omega))}{\partial x} dW_\tau(\omega)\end{aligned}$$

which completes the proof. □

Example 11.1. *Let $h(x) = x^{n+1}$ (n is any positive integer). Then (11.71) becomes*

$$\boxed{\begin{aligned}&\int_{\tau=a}^{t} W_\tau^n(\omega)\, dW_\tau(\omega) \\ &\overset{a.s.}{=} \frac{1}{n+1}\left[W_t^{n+1}(\omega)) - W_a^{n+1}(\omega))\right] - \frac{n}{2}\int_{\tau=a}^{t} W_\tau^{n-1}(\omega))d\tau\end{aligned}} \tag{11.72}$$

Notice that for $n = 1$ the formula (11.72) coincides with (11.30) valid with probability 1 when $t = b$.

11.3.3.2 Multi-dimensional case

Consider now n-dimensional stochastic vector process $\{X(t,\omega)\}_{t\in[a,b]}$ with the stochastic differential

$$\boxed{\begin{aligned}dX(t,\omega) &= F(t,\omega)\,dt + G(t,\omega)\,dW_t(\omega) \\ X(t,\omega) &= (x_1(t,\omega), \ldots, x_n(t,\omega))^\intercal\end{aligned}} \tag{11.73}$$

where

$$W_t(\omega) = \left(W_{1,t}(\omega), \ldots, W_{m,t}(\omega)\right)^\top$$

is an m-dimensional standard Wiener process with independent components,

$$F(t,\omega) = (f_1(t,\omega), \ldots, f_n(t,\omega))^\top$$

is an n-dimensional vector function with components which are second-order mean-square continuous processes,

$$G(t,\omega) = \left[g_{ij}(t,\omega)\right]_{i=\overline{1,n};\, j=\overline{1,m}} \in \mathbb{R}^{n\times m}$$

is an $n \times m$-matrix function with components $g_{ij}(t,\omega) \in \mathcal{L}^2[a,b]$. In the component-wise form (11.73) looks like

$$dx_i(t,\omega) = f_i(t,\omega)\,dt + \sum_{j=1}^{m} g_{ij}(t,\omega)\,dW_{j,t}(\omega), \quad i = \overline{1,n} \tag{11.74}$$

since (in the symbolic form)

$$\begin{aligned} dx_i(t,\omega)\,dx_j(t,\omega) = {} & f_i(t,\omega)\,f_j(t,\omega)\,(dt)^2 \\ & + \sum_{k=1}^{m} [f_i(t,\omega)\,g_{ik}(t,\omega)]\,dW_{k,t}(\omega)\,dt \\ & + \sum_{k=1}^{m}\sum_{\substack{k=1 \\ k\neq h}}^{m} \left[g_{ik}(t,\omega)\,g_{jh}(t,\omega)\right] dW_{k,t}(\omega)\,dW_{h,t}(\omega) \\ & + \sum_{k=1}^{m} \left[g_{ik}(t,\omega)\,g_{jk}(t,\omega)\right] \left(dW_{k,t}(\omega)\right)^2 \end{aligned}$$

due to the independence of $W_{k,t}(\omega)$ and $W_{h,t}(\omega)$ and the mean-square calculus the $dW_{k,t}(\omega)\,dW_{h,t}(\omega)$ terms, for $k \neq h$, vanish. Thus the contribution of each of these second-order terms to the stochastic differential is $\frac{1}{2}\frac{\partial^2 V(t,X)}{\partial x_i \partial x_j}\sum_{k=1}^{m}\left[g_{ik}(t,\omega)\,g_{jk}(t,\omega)\right]dt$. So the Itô formula is established rigorously, similarly to the one-dimensional case (11.65) and looks as follows:

$$\begin{aligned} dV(t, X(t,\omega)) = {} & \left[\frac{\partial}{\partial t}V(t, X(t,\omega)) + \sum_{i=1}^{n}\frac{\partial}{\partial x_i}V(t, X(t,\omega))\,f_i(t,\omega)\right. \\ & \left. + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 V(t, X(t,\omega))}{\partial x_i \partial x_j}\sum_{k=1}^{m}\left[g_{ik}(t,\omega)\,g_{jk}(t,\omega)\right]\right] dt \\ & + \sum_{i=1}^{n}\frac{\partial V(t, X(t,\omega))}{\partial x_i}\sum_{k=1}^{m} g_{jk}(t,\omega)\,dW_{k,t}(\omega) \end{aligned}$$

or, in the equivalent vector-matrix notations,

$$\boxed{dV = \left[V_t + V_x^\intercal F + \frac{1}{2}\mathrm{tr}\left\{GG^\intercal V_{xx}\right\}\right]dt + V_x^\intercal G dW} \tag{11.75}$$

where V_t, V_x and V_{xx} denote the partial derivative on t, the gradient of V and the Hessian-matrix (the second partial derivatives) of V respectively, and $\mathrm{tr}\{\cdot\}$ is sum of the main diagonal entries.

Example 11.2. *Let $n = 2$ and*

$$V(t, X) = \frac{1}{2}X^\intercal \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} X = x_1 x_2$$

Then the application of (11.75) implies (the arguments are omitted)

1. $W_1 \neq W_2 : m = 2,\ G(t, \omega) = \begin{bmatrix} g_1(t, \omega) & 0 \\ 0 & g_2(t, \omega) \end{bmatrix}$ *and therefore*

$$dV = [x_2 f_1 + x_1 f_2]\, dt + x_2 g_1 dW_1 + x_1 g_2 dW_2 \tag{11.76}$$

2. $W_1 = W_2 = W : m = 1,\ G(t, \omega) = \begin{bmatrix} g_1(t, \omega) \\ g_2(t, \omega) \end{bmatrix}$ *and therefore*

$$\begin{aligned} dV &= \left[x_2 f_1 + x_1 f_2 + \frac{1}{2}\mathrm{tr}\left\{\begin{bmatrix} g_1^2 & g_1 g_2 \\ g_1 g_2 & g_2^2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\}\right]dt \\ &\quad + x_2 g_1 dW_1 + x_1 g_2 dW_2 \\ &= [x_2 f_1 + x_1 f_2 + g_1 g_2]\, dt + x_2 g_1 dW_1 + x_1 g_2 dW_2 \end{aligned} \tag{11.77}$$

Example 11.3. *For $G(t, \omega) \in \mathbb{R}^{n \times m}$ and $X(t, \omega)$ satisfying*

$$dX(t, \omega) = G(t, \omega)\, dW_t(\omega)$$

for $V(t, X) = \|X\|^{2k}$ one has

$$V_x = 2k\,\|X\|^{2k-1}\frac{X}{\|X\|} = 2k\,\|X\|^{2k-2} X$$

$$\begin{aligned} V_{xx} &= 2k(2k-2)\,\|X\|^{2k-3}\tfrac{X}{\|X\|}X^\intercal + 2k\,\|X\|^{2k-2} I \\ &= 4k(k-1)\,\|X\|^{2k-4} XX^\intercal + 2k\,\|X\|^{2k-2} I \end{aligned}$$

$$\begin{aligned} \frac{1}{2}\mathrm{tr}\{GG^\intercal V_{xx}\} &= 2k(k-1)\,\|X\|^{2k-4}(G^\intercal X)^2 \\ &\quad + k\,\|X\|^{2k-2}\,\mathrm{tr}\{GG^\intercal\} \end{aligned}$$

and therefore

$$\begin{aligned} d\|X\|^{2k} &= \left[\frac{1}{2}\operatorname{tr}\left\{GG^\top V_{xx}\right\}\right]dt + V_x^\top G dW \\ &= \left[2k(k-1)\|X\|^{2k-4}\left(G^\top X\right)^2 + k\|X\|^{2k-2}\operatorname{tr}\left\{GG^\top\right\}\right]dt \\ &\quad + 2k\|X\|^{2k-2}\left(G^\top X\right)dW \end{aligned} \tag{11.78}$$

The next lemma generalizes the example above.

Lemma 11.7. *Let $x(t,\omega) \in \mathbb{R}^n$ and $y(t,\omega) \in \mathbb{R}^n$ be the stochastic processes satisfying*

$$\boxed{\begin{aligned} dx(t,\omega) &= F^x(t,\omega)\,dt + G^x(t,\omega)\,dW_t^x(\omega) \\ x(0,\omega) &= x_0(\omega) \\ dy(t,\omega) &= F^y(t,\omega)\,dt + G^y(t,\omega)\,dW_t^y(\omega) \\ y(0,\omega) &= y_0(\omega) \end{aligned}} \tag{11.79}$$

where $W_t^x(\omega) \in \mathbb{R}^{n_x}$ and $W_t^y(\omega) \in \mathbb{R}^{n_y}$ are standard Wiener processes. Then

$$\boxed{\begin{aligned} & d\left[x^\top(t,\omega)\,y(t,\omega)\right] \\ &\quad = \left(x^\top(t,\omega)\,F^y(t,\omega) + y^\top(t,\omega)\,F^x(t,\omega)\right. \\ &\qquad \left. + \operatorname{tr}\left[G^{x\top}(t,\omega)\,G^y(t,\omega)\right]\right)dt \\ &\qquad + \left[x^\top(t,\omega)\,G^y(t,\omega)\,dW^y + y^\top(t,\omega)\,G^x(t,\omega)\,dW^x\right] \end{aligned}} \tag{11.80}$$

Proof. Define

$$\begin{aligned} z(t,\omega) &:= \left(x^\top(t,\omega),\, y^\top(t,\omega)\right)^\top \in \mathbb{R}^{2n} \\ W_t(\omega) &:= \left(W_t^x(\omega)^\top,\, W_t^y(\omega)^\top\right)^\top \in \mathbb{R}^{n_x+n_y} \end{aligned}$$

which, obviously, satisfies

$$\begin{aligned} dz(t,\omega) &= F^z(t,\omega)\,dt + G^z(t,\omega)\,dW_t(\omega) \\ z(0,\omega) &= z_0(\omega) = \begin{pmatrix} x_0(\omega) \\ y_0(\omega) \end{pmatrix} \\ F^z(t,\omega) &:= \begin{pmatrix} F^x(t,\omega) \\ F^y(t,\omega) \end{pmatrix}, \quad G^z(t,\omega) := \begin{pmatrix} G^x(t,\omega) & 0 \\ 0 & G^y(t,\omega) \end{pmatrix} \end{aligned}$$

Then applying the Itô formula (11.75)

$$dV = \left[V_t + V_x^\top F + \frac{1}{2}\operatorname{tr}\left\{GG^\top V_{xx}\right\}\right]dt + V_x^\top G dW$$

to the function

$$V(t,z) = \frac{1}{2}z^\top \begin{bmatrix} 0 & I_{n\times n} \\ I_{n\times n} & 0 \end{bmatrix} z = x^\top y$$

we get (omitting arguments)

$$\begin{aligned} dV &= z^{\mathsf{T}} \begin{bmatrix} 0 & I_{n\times n} \\ I_{n\times n} & 0 \end{bmatrix} F^{z} dt \\ &\quad + \frac{1}{2}\operatorname{tr}\left\{ G^{z} G^{z\mathsf{T}} \begin{bmatrix} 0 & I_{n\times n} \\ I_{n\times n} & 0 \end{bmatrix} \right\} dt + z^{\mathsf{T}} \begin{bmatrix} 0 & I_{n\times n} \\ I_{n\times n} & 0 \end{bmatrix} G^{z} dW \\ &= \left(x^{\mathsf{T}} F^{y} + y F^{x} \right) dt + \operatorname{tr}\left[G^{x\mathsf{T}} G^{y} \right] dt + \left[x^{\mathsf{T}} G^{y} dW^{y} + y^{\mathsf{T}} G^{x} dW^{x} \right] \end{aligned}$$

which complete the proof. □

11.3.3.3 Estimation of moments of a stochastic integral

One of the most important applications of the vector-form Itô formula (11.75) is related to the following estimate for the even order moments of stochastic Itô integral.

Theorem 11.4. (Gut, 2005) *Suppose $G = G(t,\omega)$ is an $n \times m$-matrix function with components $g_{ij}(t,\omega) \in \mathcal{L}^2[a,b]$ given on the interval $[a,b]$. If for a given positive integer k*

$$\int_{t=a}^{b} \mathsf{E}\left\{ \|G(t,\omega)\|^{2k} \right\} dt < \infty \tag{11.81}$$

then

$$\boxed{\begin{aligned} &\mathsf{E}\left\{ \left\| \int_{t=a}^{b} G(t,\omega)\, dW_t(\omega) \right\|^{2k} \right\} \\ &\qquad \le [k(2k-1)]^{k} (b-a)^{k-1} \int_{t=a}^{b} \mathsf{E}\left\{ \|G(t,\omega)\|^{2k} \right\} dt \end{aligned}} \tag{11.82}$$

Proof. Applying the integral version of (11.78) for

$$X(t,\omega) := \int_{\tau=a}^{t} G(\tau,\omega)\, dW_{\tau}(\omega)$$

it follows that

$$\begin{aligned} \left\| \int_{\tau=a}^{t} G(\tau,\omega)\, dW_{\tau}(\omega) \right\|^{2k} &= 2k \int_{s=a}^{t} \left\| \int_{r=a}^{s} G(r,\omega)\, dW_{r}(\omega) \right\|^{2k-2} \\ &\quad \times \left[\int_{r=a}^{s} G(r,\omega)\, dW_{r}(\omega) \right]^{\mathsf{T}} G(s,\omega)\, dW_{s}(\omega) \end{aligned}$$

$$+\int_{s=a}^{t}\left\{k\left\|\int_{r=a}^{s}G(r,\omega)\,dW_r(\omega)\right\|^{2k-2}\operatorname{tr}\left\{G(r,\omega)\,G^{\mathsf{T}}(r,\omega)\right\}\right.$$
$$+2k(k-1)\left\|\int_{r=a}^{s}G(r,\omega)\,dW_r(\omega)\right\|^{2k-4}$$
$$\left.\times\left\|\left[\int_{r=a}^{s}G(r,\omega)\,dW_r(\omega)\right]^{\mathsf{T}}G(r,\omega)\right\|^{2}\right\}ds$$

Taking the expected value of this relation and taking into account that the expectation of the first term in the right-hand side is zero and the second and third terms are the same except for the constant factors k and $2k(k-1)$, for $t=b$ by the Hölder inequality (13.73) one has

$$\mathsf{E}\left\{\left\|\int_{t=a}^{b}G(t,\omega)\,dW_t(\omega)\right\|^{2k}\right\}$$
$$=k(2k-1)\int_{t=a}^{b}\mathsf{E}\left\{\left\|\int_{\tau=a}^{t}G(\tau,\omega)\,dW_\tau(\omega)\right\|^{2k-2}\operatorname{tr}\left\{G(t,\omega)\,G^{\mathsf{T}}(t,\omega)\right\}\right\}dt$$
$$\le k(2k-1)\left(\int_{t=a}^{b}\mathsf{E}\left\{\left\|\int_{\tau=a}^{t}G(\tau,\omega)\,dW_\tau(\omega)\right\|^{2k}\right\}dt\right)^{(2k-2)/2k}$$
$$\times\left(\int_{t=a}^{b}\mathsf{E}\left\{\operatorname{tr}\left\{G(t,\omega)\,G^{\mathsf{T}}(t,\omega)\right\}^{2k}\right\}dt\right)^{2/2k}$$

Since $\mathsf{E}\left\{\left\|\int_{\tau=a}^{t}G(\tau,\omega)\,dW_\tau(\omega)\right\|^{2k}\right\}$ is a nondecreasing function we have

$$\mathsf{E}\left\{\left\|\int_{\tau=a}^{t}G(\tau,\omega)\,dW_\tau(\omega)\right\|^{2k}\right\}\le\mathsf{E}\left\{\left\|\int_{\tau=a}^{b}G(\tau,\omega)\,dW_\tau(\omega)\right\|^{2k}\right\}$$

Hence

$$\mathsf{E}\left\{\left\|\int_{t=a}^{b}G(t,\omega)\,dW_t(\omega)\right\|^{2k}\right\}$$

$$\leq k(2k-1)\left(\mathsf{E}\left\{\left\|\int_{\tau=a}^{b} G(\tau,\omega)\,dW_{\tau}(\omega)\right\|^{2k}\right\}(b-a)\right)^{(2k-2)/2k}$$

$$\times\left(\int_{t=a}^{b}\mathsf{E}\left\{\operatorname{tr}\left\{G(t,\omega)\,G^{\mathsf{T}}(t,\omega)\right\}^{2k}\right\}dt\right)^{2/2k}$$

Raising then both sides to the k-th power and dividing both sides by

$$\left(\mathsf{E}\left\{\left\|\int_{\tau=a}^{b} G(\tau,\omega)\,dW_{\tau}(\omega)\right\|^{2k}\right\}\right)^{k-1}$$

yields the result. □

11.4 The Stratonovich stochastic integral

11.4.1 Main property of λ-stochastic integrals

In Section 11.2.3 it is indicated that there are infinitely many possible choices for the interpretation of the λ-stochastic integral[1]

$$\boxed{\begin{aligned}\int_{\tau=a}^{b} W_{\tau}(\omega)\overset{\lambda\in[0,1]}{\circ} dW_{\tau}(\omega) &:= \underset{\substack{\delta_n \to 0\\ n\to\infty}}{\text{l.i.m}}\int_{\tau=a}^{b}\varphi_n^{\lambda}(\tau,\omega)\,dW_{\tau}(\omega)\\ &= \frac{1}{2}\left[W_b^2(\omega)-W_a^2(\omega)\right]+\left(\lambda-\frac{1}{2}\right)(b-a)\end{aligned}} \tag{11.83}$$

However, only two stochastic integrals (Ito, 1951) with $\lambda = 0$ and Stratonovich (1966) with $\lambda = 1/2$ have gained acceptance in the theoretical and application literature. The first one has been investigated just above. Here we will present the definition and the basic properties of the Stratonovich stochastic integral including a basic theorem showing that this integral satisfies the usual rules of calculus.

Suppose $\{x(t,\omega)\}_{t\in[a,b]}$ has the stochastic differential (11.61), i.e.,

$$x(t,\omega)\overset{a.s.}{=}x(a,\omega)+\int_{\tau=a}^{t} f(\tau,\omega)\,d\tau+\int_{\tau=a}^{t} g(\tau,\omega)\,dW_{\tau}(\omega)$$

or, equivalently, in the *symbolic form* (11.62)

$$dx(t,\omega)=f(t,\omega)\,dt+g(t,\omega)\,dW_t(\omega)$$

[1] If the mesh size δ_n of the corresponding partitions of the interval $[a,b]$ tends to zero such that $\sum_{n=1}^{\infty}\delta_n<\infty$ then the definition (11.83) can be understood with probability 1.

Here $\{f(t,\omega)\}_{t\in[a,b]}$ is a second-order mean-square continuous process, the first integral in the right-hand side of (11.61) is the usual simple-path time-integral (11.10) and the second integral is the Itô integral of the function $g(t,\omega)\in\mathcal{L}^2[a,b]$.

First, let us introduce a one-parametric family of stochastic integrals

$$\boxed{I^{\lambda}_{[a,b]}(h) := \int_{\tau=a}^{b} h(\tau, x(\tau,\omega)) \overset{\lambda\in[0,1]}{\circ} dW_\tau(\omega)} \tag{11.84}$$

containing the Stratonovich $I^{S}_{[a,b]}(g)$ (11.4) as well as the Itô (11.3) integrals, and for any fixed $\lambda\in[0,1]$ relates each integral in this family to the Itô integral.

Definition 11.5. *The* λ-***stochastic integral*** $I^{\lambda}_{[a,b]}(h)$ *of a function* $h(t,x(t,\omega))$ *from* $\mathcal{L}^2[a,b]$ *is defined for any* $\lambda\in[0,1]$ *as*

$$\boxed{I^{\lambda}_{[a,b]}(h) := \underset{\delta_n \underset{n\to\infty}{\to} 0}{\text{l.i.m}} \sum_{k=1}^{n} h(t_k, \lambda x(t_k,\omega) + (1-\lambda) x(t_{k-1},\omega)) \Delta W^n_k(\omega)} \tag{11.85}$$

given on $\mathcal{E}$ *when* $n\to\infty$ *as a continuous linear random functional action from* $\mathcal{L}^2[a,b]$ *on* $L^2(\Omega)$.

Theorem 11.5. *For any function* $h\in\mathcal{L}^2[a,b]$*, namely, such that*

$$\boxed{\int_{\tau=a}^{b} \mathsf{E}\left\{h^2(\tau, x(\tau,\omega))\right\} dt < \infty} \tag{11.86}$$

which is, additionally, differentiable on x*, and any fixed* $\lambda\in[0,1]$*, the stochastic integral (11.84) exists, and satisfies*

$$\boxed{\begin{aligned} I^{\lambda}_{[a,b]}(g) &:= \int_{\tau=a}^{b} h(\tau, x(\tau,\omega)) \overset{\lambda\in[0,1]}{\circ} dW_\tau(\omega) \\ &= \int_{\tau=a}^{b} h(\tau, x(\tau,\omega))\, dW_\tau(\omega) \\ &\quad + \lambda \int_{\tau=a}^{b} \frac{\partial}{\partial x} h(\tau, x(\tau,\omega))\, g(\tau,\omega)\, d\tau \end{aligned}} \tag{11.87}$$

where the integrals on the right-hand side are the corresponding Itô integral (11.49) and the ordinary time-integral (11.5) along the sample paths, respectively.

Proof. The existence of the first integral on the right-hand side of (11.87) as the mean-square limit (as the mesh size $\delta_n \underset{n\to\infty}{\to} 0$) of

$$\sum_{k=1}^{n} h\left(t_k, x\left(t_{k-1}, \omega\right)\right) \Delta W_k^n\left(\omega\right) \tag{11.88}$$

results from the condition (11.86) and the properties of the Itô integral (11.49). Therefore, to establish (11.87) it is sufficient to show that an a.s.-limit of the difference of the sums in (11.85) with any $\lambda \in [0, 1]$ and in (11.88) agrees with the ordinary sample-path time-integral in (11.87); that is, for sequences of the partition $\delta_n \underset{n\to\infty}{\to} 0$ we have to show that

$$\sum_{k=1}^{n} \left[h\left(t_k, \lambda x\left(t_k, \omega\right) + (1-\lambda) x\left(t_{k-1}, \omega\right)\right) - h\left(t_k, x\left(t_{k-1}, \omega\right)\right)\right] \Delta W_k^n\left(\omega\right)$$
$$\overset{a.s.}{\underset{\delta_n \underset{n\to\infty}{\to} 0}{\to}} \lambda \int_{\tau=a}^{b} \frac{\partial}{\partial x} h\left(\tau, x\left(\tau, \omega\right)\right) g\left(\tau, \omega\right) d\tau$$

By mean-value theorem (valid in view of the differentiability of $h(t, x)$ on x) we have

$$h\left(t_k, \lambda x\left(t_k, \omega\right) + (1-\lambda) x\left(t_{k-1}, \omega\right)\right) - h\left(t_k, x\left(t_{k-1}, \omega\right)\right)$$
$$= \lambda \frac{\partial}{\partial x} h\left(t_k, \theta_k\right) \left[x\left(t_k, \omega\right) - x\left(t_{k-1}, \omega\right)\right], \quad \theta_k \in [0, \lambda]$$

where

$$\frac{\partial}{\partial x} h\left(t_k, \theta_k\right) := \frac{\partial}{\partial x} h\left(t_k, \theta_k x\left(t_k, \omega\right) + (1-\theta_k) x\left(t_{k-1}, \omega\right)\right)$$

Therefore, to prove the result it suffices to show that

$$\sum_{k=1}^{n} \frac{\partial}{\partial x} h\left(t_k, \theta_k\right) \left[x\left(t_k, \omega\right) - x\left(t_{k-1}, \omega\right)\right] \Delta W_k^n\left(\omega\right)$$
$$\overset{a.s.}{\underset{\delta_n \underset{n\to\infty}{\to} 0}{\to}} \int_{\tau=a}^{b} \frac{\partial}{\partial x} h\left(\tau, x\left(\tau, \omega\right)\right) g\left(\tau, \omega\right) d\tau$$

By the continuity property of $h(t, x)$ on both arguments

$$\frac{\partial}{\partial x} h\left(t_k, \theta_k\right) \overset{a.s.}{\underset{\delta_n \underset{n\to\infty}{\to} 0}{\to}} \frac{\partial}{\partial x} h\left(t, x\left(t, \omega\right)\right)$$

But

$$\sum_{k=1}^{n} \left[x\left(t_k, \omega\right) - x\left(t_{k-1}, \omega\right)\right] \Delta W_k^n\left(\omega\right)$$
$$= \sum_{k=1}^{n} \left[f\left(t_k, \omega\right) \Delta t_k + g\left(t_k, \omega\right) \Delta W_k^n\left(\omega\right)\right] \Delta W_k^n\left(\omega\right)$$
$$\overset{a.s.}{\underset{\delta_n \underset{n\to\infty}{\to} 0}{\to}} \int_{\tau=a}^{b} g\left(\tau, \omega\right) d\tau$$

in view of the properties $(dW_t(\omega))^2 \sim dt$ and (11.35) if the mesh size δ_n of the corresponding partitions of the interval $[a, b]$ tends to zero such that $\sum_{n=1}^{\infty} \delta_n < \infty$. Theorem is proven. □

Corollary 11.6. *In the Stratonovich case (when $\lambda = 1/2$) the relation (11.87) becomes*

$$\boxed{\begin{aligned} I^S_{[a,b]}(g) &:= \int_{\tau=a}^{b} h(\tau, x(\tau,\omega)) \overset{\lambda=1/2}{\circ} dW_\tau(\omega) \\ &= \int_{\tau=a}^{b} h(\tau, x(\tau,\omega))\, dW_\tau(\omega) \\ &\quad + \frac{1}{2}\int_{\tau=a}^{b} \frac{\partial}{\partial x} h(\tau, x(\tau,\omega))\, g(\tau,\omega)\, d\tau \end{aligned}} \tag{11.89}$$

which states the direct relation between the Stratonovich (11.4) in the left-hand side and Itô (11.3) in the right-hand side integrals.

11.4.2 The Stratonovich differential

Now it becomes possible to define the Stratonovich differential analogously to (11.61).

Definition 11.6. *The stochastic* ***Stratonovich differential***

$$\boxed{d_S x(t,\omega) = f(t, x(t,\omega))\, dt + g(t, x(t,\omega)) \overset{\lambda=1/2}{\circ} dW_t(\omega)} \tag{11.90}$$

is defined as a stochastic process $\{x(t,\omega)\}_{t\geq 0}$ satisfying

$$\boxed{\begin{aligned} x(t,\omega) &\overset{a.s.}{=} x(a,\omega) + \int_{\tau=a}^{t} f(\tau, x(\tau,\omega))\, d\tau \\ &\quad + \int_{\tau=a}^{t} g(\tau, x(\tau,\omega)) \overset{\lambda=1/2}{\circ} dW_\tau(\omega) \end{aligned}} \tag{11.91}$$

Using the relation (11.89) for

$$h(t, x) = g(t, x)$$

one can express (11.91) as follows:

$$\boxed{\begin{aligned} x(t,\omega) &\overset{a.s.}{=} x(a,\omega) + \int_{\tau=a}^{t} f(\tau, x(\tau,\omega))\, d\tau \\ &+ \int_{\tau=a}^{t} g(\tau, x(\tau,\omega)) \overset{\lambda=1/2}{\circ} dW_\tau(\omega) = x(a,\omega) \\ &+ \int_{\tau=a}^{t} \left(f(\tau, x(\tau,\omega)) + \frac{1}{2}\left[\frac{\partial}{\partial x} g(\tau, x(\tau,\omega))\right] g(\tau, x(\tau,\omega)) \right) d\tau \\ &+ \int_{\tau=a}^{b} g(\tau, x(\tau,\omega))\, dW_\tau(\omega) \end{aligned}} \tag{11.92}$$

Thus $x(t,\omega)$ has also the Itô differential (11.61) which satisfies, as it follows from (11.92),

$$\boxed{\begin{aligned} dx(t,\omega) &= g(t, x(t,\omega))\, dW_t(\omega) \\ &\quad + \left(f(\tau, x(t,\omega)) + \frac{1}{2}\left[\frac{\partial}{\partial x} g(t, x(\tau,\omega))\right] g(t, x(\tau,\omega)) \right) dt \end{aligned}} \tag{11.93}$$

The theorem below shows that the Stratonovich stochastic integral satisfies the usual rules of calculus.

Theorem 11.6. *Suppose that the functions $g(x)$ and $h(x)$ are continuously differentiable and twice continuously differentiable respectively. Let also in (11.90) $f(t,x) \equiv 0$, that is,*

$$\boxed{d_S x(t,\omega) = g(x(t,\omega)) \overset{\lambda=1/2}{\circ} dW_t(\omega)} \tag{11.94}$$

*Then **for the Stratonovich differential (11.94) the usual chain-rule of calculus holds**, i.e.,*

$$\boxed{d_S h(x(t,\omega)) = \frac{\partial}{\partial x} h(x(t,\omega))\, d_S x(t,\omega)} \tag{11.95}$$

Proof. First, notice that from (11.93) for $f(t,x) \equiv 0$ it follows that

$$dx(t,\omega) = g(x(t,\omega))\, dW_t(\omega) + \frac{1}{2}\left[\frac{\partial}{\partial x} g(x(t,\omega))\right] g(x(t,\omega))\, dt$$

By the Itô formula (11.65) the stochastic differential (11.61) for h is (omitting arguments)

$$dh = \frac{1}{2}\left[h' g' g + h'' g^2\right] dt + h' g dW$$

By (11.87) the Itô term $h'gdW$ can be replaced by

$$h'gdW = h'g \overset{\lambda=1/2}{\circ} dW - \frac{1}{2}\left(h'g\right)' gdt$$
$$= h'g \overset{\lambda=1/2}{\circ} dW - \frac{1}{2}\left[h''g + h'g'\right] gdt$$

which gives

$$dh = \frac{1}{2}\left[h'g'g + h''g^2\right] dt + h'gdW$$
$$= \frac{1}{2}\left[h'g'g + h''g^2\right] dt + h'g \overset{\lambda=1/2}{\circ} dW - \frac{1}{2}\left[h''g + h'g'\right] gdt$$
$$= h'g \overset{\lambda=1/2}{\circ} dW = h'd_S x$$

Theorem is proven. □

11.4.3 Multidimensional case

Analogously to the Itô integral, the Stratonovich differential properties (11.90) and (11.95) can be extended for the multidimensional case in the following way.

The vector-form Stratonovich differential of

$$X(t,\omega) = (x_1(t,\omega), \ldots, x_n(t,\omega))^{\intercal}$$

is expressed as

$$d_S X(t,\omega) = F(t,\omega)\,dt + G(t,\omega) \overset{\lambda=1/2}{\circ} dW_t(\omega)$$

where

$$W_t(\omega) = \left(W_{1,t}(\omega), \ldots, W_{m,t}(\omega)\right)^{\intercal}$$
$$F(t,\omega) = (f_1(t,\omega), \ldots, f_n(t,\omega))^{\intercal}$$
$$G(t,\omega) = \left[g_{ij}(t,\omega)\right]_{i=\overline{1,n};\, j=\overline{1,m}} \in \mathbb{R}^{n\times m}, \quad g_{ij}(t,\omega) \in \mathcal{L}^2[a,b]$$

or equivalently, in the component-wise form as

$$d_S x_i(t,\omega) = f_i(t,\omega)\,dt + \sum_{j=1}^{m} g_{ij}(t,\omega) \overset{\lambda=1/2}{\circ} dW_{j,t}(\omega), \quad i = \overline{1,n}$$

which for $V = V(t,x)$ by (11.95) implies the following *Stratonovich differentiation rule*

$$\boxed{d_S V = V_t dt + V_x^{\intercal} d_S X(t,\omega)} \qquad (11.96)$$

12 Stochastic Differential Equations

Contents

12.1 Solution as a stochastic process

In this chapter we will discuss the class of so-called stochastic differential equations, introduced by K. Itô, whose basic theory was developed independently by Itô and I. Gihman during the 1940s. The extended version of this theory can be found in Ito (1951) and Gihman and Skorohod (1972). There the Itô-type integral calculus is applied. The principal motivation for choosing the *Itô approach* (as opposed to the Stratonovich calculus as another very popular interpretation of the stochastic integration) is that the Itô method extends to a broader class of equations and transforms the probability law of the Wiener process in a more natural way. This approach implements the so-called *diffusion approximation*, which arises from random difference equation models and has a wide application to control problems in engineering sciences motivated by the need for more sophisticated models, which spurred further work on these types of equations in the 1950s and 1960s.

12.1.1 Definition of a solution

Here we consider the single-dimensional case.

Definition 12.1. *By a* ***solution*** $x(t,\omega)$ *of the* ***stochastic differential equation***

$$\boxed{dx(t,\omega) = f(t, x(t,\omega))\,dt + g(t, x(t,\omega))\,dW_t(\omega)} \tag{12.1}$$

with a specific initial condition

$$\boxed{x(0,\omega) = x_0(\omega)} \tag{12.2}$$

is meant a stochastic process $\{x(t,\omega)\}_{t\geq 0}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *having the stochastic differential (12.1) as it was defined in the previous chapter, or equivalently,* *for all* $t \in [0, T]$ $(T < \infty)$ *the process* $\{x(t, \omega)\}_{t \geq 0}$ *satisfies the following stochastic integral equation*

$$\boxed{x(t, \omega) \overset{a.s.}{=} x_0(\omega) + \int_{\tau=0}^{t} f(\tau, x(\tau, \omega))\, d\tau + \int_{\tau=0}^{t} g(\tau, x(\tau, \omega))\, dW_\tau(\omega)} \tag{12.3}$$

Here the first integral in the right-hand side is an ordinary one along paths (11.1), and the second one is the Itô stochastic integral (11.3).

12.1.2 Existence and uniqueness

In this subsection the simplest version of the existence and uniqueness theorem is presented provided certain additional conditions concerning the dependence of f and on the variable x are imposed.

12.1.2.1 Result with the global Lipschitz condition

Theorem 12.1. *Let*

1. *the functions* $f(t, x)$ *and* $g(t, x)$ *be assumed to be* ***measurable*** *with respect to* $t \in [0, T]$ *and* $x \in \mathbb{R}$, *i.e., for any* $c \in \mathbb{R}$

$$\boxed{\begin{array}{l} \{(t, x) \mid f(t, x) < c\} \in [0, T] \times \mathbb{R} \\ \{(t, x) \mid g(t, x) < c\} \in [0, T] \times \mathbb{R} \end{array}} \tag{12.4}$$

2. *both functions satisfy the* ***uniform global Lipschitz condition*** *on the variable* x *for all* $t \in [0, T]$, *namely, for all* $t \in [0, T]$ *and all* $x, y \in \mathbb{R}$ *there exists a positive constant* K *such that*
   (a)

$$\boxed{|f(t, x) - f(t, y)| + |g(t, x) - g(t, y)| \leq K |x - y|} \tag{12.5}$$

   (b)

$$\boxed{|f(t, x)|^2 + |g(t, x)|^2 \leq K \left(1 + x^2\right)} \tag{12.6}$$

3. $x_0(\omega)$ *be* ***independent*** *of* $W_t(\omega)$ *for all* $t > 0$ *and have a* ***bounded second moment***, *i.e.,*

$$\boxed{\mathsf{E}\left\{x_0^2(\omega)\right\} < \infty} \tag{12.7}$$

Then there exists a solution $x(t, \omega)$ *of (12.3) defined for all* $t \in [0, T]$ *which is continuous with probability* 1 *and has the second bounded moment, that is,*

$$\boxed{\sup_{t\in[0,T]} \mathsf{E}\left\{x^2(t,\omega)\right\} < \infty} \tag{12.8}$$

*Furthermore, a solution with these properties is **pathwise unique** which means that if $x(t,\omega)$ and $y(t,\omega)$ are two solutions then*

$$\boxed{\mathsf{P}\left\{\sup_{t\in[0,T]} |x(t,\omega) - y(t,\omega)| = 0\right\} = 1} \tag{12.9}$$

Proof.

(a) *An auxiliary result.* Suppose that two stochastic processes $y_i(t,\omega)$ $(i = 1, 2)$ with bounded second moments, namely, satisfying

$$\sup_{t\in[0,T]} \mathsf{E}\left\{y_i^2(t,\omega)\right\} < \infty, \quad (i = 1, 2) \tag{12.10}$$

possess the following property: the σ-algebra generated by $y_i(s,\omega)$ $(i = 1, 2)$ and $W_s(\omega)$ with $s \le t$ is independent of the increment $\left[W_{t+h}(\omega) - W_t(\omega)\right]$ for any $h > 0$. Show that if $z_i(t,\omega)$ $(i = 1, 2)$ are defined by

$$z_i(t,\omega) \overset{a.s.}{=} x_0(\omega) + \int_{\tau=0}^{t} f(\tau, y_i(\tau,\omega))\,d\tau + \int_{\tau=0}^{t} g(\tau, y_i(\tau,\omega))\,dW_\tau(\omega) \tag{12.11}$$

then for some $L > 0$

$$\mathsf{E}\left\{|z_1(t,\omega) - z_2(t,\omega)|^2\right\} \le L \int_{\tau=0}^{t} \mathsf{E}\left\{|y_1(t,\omega) - y_2(t,\omega)|^2\right\} d\tau \tag{12.12}$$

Obviously, from (12.11) by the inequality $|a+b|^2 \le 2\left(|a|^2 + |b|^2\right)$ one has

$$\begin{aligned} &|z_1(t,\omega) - z_2(t,\omega)|^2 \\ &\quad \le 2\left(\int_{\tau=0}^{t} \left[f(\tau, y_1(\tau,\omega)) - f(\tau, y_2(\tau,\omega))\right] d\tau\right)^2 \\ &\qquad + 2\left(\int_{\tau=0}^{t} \left[g(\tau, y_1(\tau,\omega)) - g(\tau, y_2(\tau,\omega))\right] dW_\tau(\omega)\right)^2 \end{aligned} \tag{12.13}$$

Using the Cauchy–Schwartz inequality argument

$$\left(\int_{\tau=0}^{t} h(\tau)\,d\tau\right)^2 \le \int_{\tau=0}^{t} 1^2 d\tau \int_{\tau=0}^{t} h^2(\tau)\,d\tau = t\int_{\tau=0}^{t} h^2(\tau)\,d\tau$$

and applying the assumption (12.5) we get

$$\left(\int_{\tau=0}^{t}\left[f\left(\tau, y_1(\tau,\omega)\right)-f\left(\tau, y_2(\tau,\omega)\right)\right]d\tau\right)^2$$

$$\leq K^2 t\int_{\tau=0}^{t}\left[y_1(\tau,\omega)-y_2(\tau,\omega)\right]^2 d\tau \tag{12.14}$$

Again, by the assumptions (12.5) and (12.10)

$$\mathsf{E}\left\{\int_{\tau=0}^{t}\left[g\left(\tau, y_1(\tau,\omega)\right)-g\left(\tau, y_2(\tau,\omega)\right)\right]^2 d\tau\right\}$$

$$\leq K^2\mathsf{E}\left\{\int_{\tau=0}^{t}\left[y_1(\tau,\omega)-y_2(\tau,\omega)\right]^2 d\tau\right\}$$

$$\leq 2K^2\mathsf{E}\left\{\int_{\tau=0}^{t}\left[y_1^2(\tau,\omega)+y_2^2(\tau,\omega)\right]d\tau\right\}$$

$$\leq 2K^2 E\left\{\int_{\tau=0}^{T}\left[y_1^2(\tau,\omega)+y_2^2(\tau,\omega)\right]d\tau\right\}$$

$$= 2K^2\int_{\tau=0}^{T}\mathsf{E}\left\{\left[y_1^2(\tau,\omega)+y_2^2(\tau,\omega)\right]d\tau\right\}$$

$$4K^2 T\max_{i=1,2}\sup_{t\in[0,T]}\mathsf{E}\left\{y_i^2(t,\omega)\right\}<\infty$$

which means that

$$\left[g\left(\tau, y_1(\tau,\omega)\right)-g\left(\tau, y_2(\tau,\omega)\right)\right]\in\mathcal{L}^2[0,T]$$

and therefore, by the property (11.51) of the Itô integral it follows that

$$\mathsf{E}\left\{\left(\int_{\tau=0}^{t}\left[g\left(\tau, y_1(\tau,\omega)\right)-g\left(\tau, y_2(\tau,\omega)\right)\right]dW_\tau(\omega)\right)^2\right\}$$

$$= \mathsf{E}\left\{\int_{\tau=0}^{t}\left[g\left(\tau, y_1(\tau,\omega)\right)-g\left(\tau, y_2(\tau,\omega)\right)\right]^2 d\tau\right\}$$

$$\leq K^2\mathsf{E}\left\{\int_{\tau=0}^{t}\left[y_1(\tau,\omega)-y_2(\tau,\omega)\right]^2 d\tau\right\} \tag{12.15}$$

The result (12.12) follows now by taking the expected value in (12.13), using (12.14) and (12.15), and setting $L := 2(T+1)K^2$.

(b) *Uniform mean-square boundedness of* $\{x_n(t,\omega)\}_{n\geq 0}$. Consider then the sequence $\{x_n(t,\omega)\}_{n\geq 0}$ of successive approximations $x_n(t,\omega)$ defined by

$$x_n(t,\omega) \overset{a.s.}{=} x_0(\omega) + \int_{\tau=0}^{t} f(\tau, x_{n-1}(\tau,\omega))\, d\tau + \int_{\tau=0}^{t} g(\tau, x_{n-1}(\tau,\omega))\, dW_\tau(\omega) \tag{12.16}$$

and show that it converges to the unique solution of (12.3). First, let us demonstrate that $\{x_n(t,\omega)\}_{n\geq 0}$ is uniformly on t mean-square bounded on $[0,T]$, i.e., for all n

$$\sup_{t\in[0,T]} \mathsf{E}\left\{x_n^2(t,\omega)\right\} \leq \text{Const} < \infty \tag{12.17}$$

The relation (12.16) implies

$$\begin{aligned}
\mathsf{E}\left\{x_n^2(t,\omega)\right\} &\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + \mathsf{E}\left\{\left(\int_{\tau=0}^{t} f(\tau, x_{n-1}(\tau,\omega))\, d\tau\right)^2\right\}\right. \\
&\quad \left. + \mathsf{E}\left\{\left(\int_{\tau=0}^{t} g(\tau, x_{n-1}(\tau,\omega))\, dW_\tau(\omega)\right)^2\right\}\right) \\
&\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + T\,\mathsf{E}\left\{\int_{\tau=0}^{t} f^2(\tau, x_{n-1}(\tau,\omega))\, d\tau\right\}\right. \\
&\quad \left. + \mathsf{E}\left\{\int_{\tau=0}^{t} g^2(\tau, x_{n-1}(\tau,\omega))\, d\tau\right\}\right)
\end{aligned}$$

Applying then the assumption (12.6) we derive

$$\begin{aligned}
\mathsf{E}\left\{x_n^2(t,\omega)\right\} &\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + TK^2\int_{\tau=0}^{t}\left[1+\mathsf{E}\left\{x_{n-1}^2(\tau,\omega)\right\}\right] d\tau\right. \\
&\quad \left. + K^2\int_{\tau=0}^{t}\left[1+\mathsf{E}\left\{x_{n-1}^2(\tau,\omega)\right\}\right] d\tau\right) \\
&\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + L\int_{\tau=0}^{t}\left[1+\mathsf{E}\left\{x_{n-1}^2(\tau,\omega)\right\}\right] d\tau\right)
\end{aligned}$$

$$\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + Lt + L\int_{\tau=0}^{t} 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\} + L\tau + \int_{s=0}^{\tau}\mathsf{E}\left\{x_{n-2}^2(s,\omega)\right\}ds\right)d\tau\right)$$

$$\leq 3\left(\mathsf{E}\left\{x_0^2(\omega)\right\}[1+3Lt] + Lt + \frac{3(Lt)^2}{2} + 3L^2\int_{\tau=0}^{t}\left(\int_{s=0}^{\tau}\mathsf{E}\left\{x_{n-2}^2(s,\omega)\right\}ds\right)d\tau\right)$$

Using the identity

$$\int_{\tau=0}^{t}(t-\tau)^m\left[\int_{s=0}^{\tau}h(s)\,ds\right]d\tau = \int_{\tau=0}^{t}\frac{(t-\tau)^{m+1}}{m+1}h(\tau)\,d\tau \tag{12.18}$$

for $m = 0$, which results from integration by parts, the last term in the inequality above can be represented as

$$\int_{\tau=0}^{t}\left(\int_{s=0}^{\tau}\mathsf{E}\left\{x_{n-2}^2(s,\omega)\right\}ds\right)d\tau = \int_{\tau=0}^{t}(t-\tau)\,\mathsf{E}\left\{x_{n-2}^2(\tau,\omega)\right\}d\tau$$

and, the continuation of the iterations back leads to

$$\mathsf{E}\left\{x_n^2(t,\omega)\right\} \leq 3\left(1+3Lt+\cdots+\frac{(3Lt)^{n-1}}{(n-1)!}\right)\mathsf{E}\left\{x_0^2(\omega)\right\}$$
$$+3Lt+\cdots+\frac{(3Lt)^n}{n!} + (3L)^n\int_{\tau=0}^{t}\frac{(t-\tau)^{n-1}}{(n-1)!}\mathsf{E}\left\{x_0^2(\omega)\right\}d\tau$$
$$\leq \left(3\mathsf{E}\left\{x_0^2(\omega)\right\}+1\right)e^{3LT} := \text{Const} < \infty$$

which implies (12.17).

(c) *Uniform mean-square and a.s. convergence of* $\{x_n(t,\omega)\}_{n\geq 0}$. Using (12.11) being applied to (12.16) with

$$z_1(t,\omega) := x_n(t,\omega), \quad y_1(\tau,\omega) := x_{n-1}(t,\omega)$$
$$z_2(t,\omega) := x_{n-1}(t,\omega), \quad y_2(\tau,\omega) := x_{n-2}(t,\omega)$$

by the estimate (12.12) one has

$$\mathsf{E}\left\{|x_n(t,\omega) - x_{n-1}(t,\omega)|^2\right\} \leq L\int_{\tau=0}^{t}\mathsf{E}\left\{|x_{n-1}(\tau,\omega) - x_{n-2}(\tau,\omega)|^2\right\}d\tau \tag{12.19}$$

Iterating (12.19) and making use the identity (12.18) we arrive at

$$\begin{aligned} &\mathsf{E}\left\{\left|x_n(t,\omega)-x_{n-1}(t,\omega)\right|^2\right\} \\ &\qquad \le L^{n-1}\int\limits_{\tau=0}^{t}\frac{(t-\tau)^{n-2}}{(n-2)!}\mathsf{E}\left\{\left|x_1(\tau,\omega)-x_0(\tau,\omega)\right|^2\right\}d\tau \end{aligned} \tag{12.20}$$

Directly from (12.16) it follows that

$$\begin{aligned} \mathsf{E}\left\{\left|x_1(t,\omega)-x_0(t,\omega)\right|^2\right\} &\le L\int\limits_{\tau=0}^{t}\left[1+\mathsf{E}\left\{\left|x_0(t,\omega)\right|^2\right\}\right]d\tau \\ &\le LT\left[1+\mathsf{E}\left\{\left|x_0(t,\omega)\right|^2\right\}\right] := \text{Const} < \infty \end{aligned}$$

Substitution of this last inequality into (12.20) leads to

$$\sup_{t\in[0,T]}\mathsf{E}\left\{\left|x_n(t,\omega)-x_{n-1}(t,\omega)\right|^2\right\} \le \text{Const}\frac{(LT)^{n-1}}{(n-1)!}, \quad n\ge 1 \tag{12.21}$$

and, hence, to

$$\begin{aligned} &\sum_{n=1}^{\infty}\sup_{t\in[0,T]}\mathsf{E}\left\{\left|x_n(t,\omega)-x_{n-1}(t,\omega)\right|^2\right\} \\ &\qquad \le \text{Const}\sum_{n=1}^{\infty}\frac{(LT)^{n-1}}{(n-1)!} = \text{Const}\, e^{LT} < \infty \end{aligned} \tag{12.22}$$

which establishes uniform (on $t\in[0,T]$) mean square convergence of $\{x_n(t,\omega)\}_{n\ge 0}$ since

$$\begin{aligned} x_n(t,\omega) &= x_0(t,\omega)+\sum_{k=1}^{n}\left[x_k(t,\omega)-x_{k-1}(t,\omega)\right] \\ &\overset{L^2}{\underset{n\to\infty}{\to}} x_0(t,\omega)+\sum_{k=1}^{\infty}\left[x_k(t,\omega)-x_{k-1}(t,\omega)\right] \end{aligned} \tag{12.23}$$

But we also have

$$\begin{aligned} x_n(t,\omega) &= x_0(t,\omega)+\sum_{k=1}^{n}\left[x_k(t,\omega)-x_{k-1}(t,\omega)\right] \\ &\overset{a.s.}{\underset{n\to\infty}{\to}} x_0(t,\omega)+\sum_{k=1}^{\infty}\left[x_k(t,\omega)-x_{k-1}(t,\omega)\right] \end{aligned} \tag{12.24}$$

Indeed, for

$$Y_n := \sup_{t\in[0,T]}\left|x_n(t,\omega)-x_{n-1}(t,\omega)\right|$$

we have

$$Y_n \leq \int_{t=0}^{T} |f(t, x_n(t,\omega)) - f(t, x_{n-1}(t,\omega))|\, dt + \sup_{t\in[0,T]} \left| \int_{\tau=0}^{t} \left[g(\tau, x_n(t,\omega)) - g(\tau, x_{n-1}(t,\omega))\right] dW_\tau(\omega) \right|$$

and hence, applying the inequalities above,

$$\mathsf{E}\left\{Y_n^2\right\} \leq 2K^2 T \int_{t=0}^{T} \mathsf{E}\left\{|x_n(t,\omega) - x_{n-1}(t,\omega)|^2\right\} dt + 8K^2 \int_{t=0}^{T} \mathsf{E}\left\{|x_n(t,\omega) - x_{n-1}(t,\omega)|^2\right\} dt \leq \text{Const}\frac{(LT)^{n-1}}{(n-1)!}$$

So,

$$\sum_{n=1}^{\infty} \mathsf{E}\left\{Y_n^2\right\} \leq \text{Const} \sum_{n=1}^{\infty} \frac{(LT)^{n-1}}{(n-1)!} = \text{Const}e^{LT} < \infty$$

and, by Chebyshev's inequality (4.10),

$$\sum_{n=1}^{\infty} \mathsf{P}\left\{Y_n > n^{-2}\right\} \leq \sum_{n=1}^{\infty} n^4 \mathsf{E}\left\{Y_n^2\right\} \leq \text{Const} \sum_{n=1}^{\infty} n^4 \frac{(LT)^{n-1}}{(n-1)!} < \infty$$

which, applying the Borel–Cantelli Lemma 6.2 implies

$$\mathsf{P}\left\{\omega : Y_n \leq n^{-2} \text{ for sufficiently large } n\right\} = 1$$

and therefore

$$\left| \sum_{k=1}^{\infty} \left[x_k(t,\omega) - x_{k-1}(t,\omega)\right] \right| \leq \sum_{k=1}^{\infty} Y_k \overset{a.s.}{\leq} \sum_{k=1}^{n-1} Y_k + \sum_{k=n}^{\infty} Y_k \overset{a.s.}{<} \infty$$

which means that (12.24) holds. We have proved uniform, almost sure convergence of $\{x_n(t,\omega)\}_{n\geq 0}$, that is, there exists

$$x(t,\omega) := \lim_{n\to\infty} x_n(t,\omega) \text{ a.s.}$$

(d) To show that $x(t,\omega) \overset{a.s.}{=} \lim_{n\to\infty} x_n(t,\omega)$ satisfies (12.3) it is sufficient to take a limit in (12.16).

(e) *Almost sure uniqueness.* Suppose that $x(t,\omega)$ and $y(t,\omega)$ are both solutions satisfying (12.3). Then, by the property (12.12),

$$\mathsf{E}\left\{|x(t,\omega) - y(t,\omega)|^2\right\} \leq L \int_{\tau=0}^{t} \mathsf{E}\left\{|x(\tau,\omega) - y(\tau,\omega)|^2\right\}$$

By the Gronwall lemma (see, for example, Corollary 19.4 in Poznyak (2008))[1] we have

$$\mathsf{E}\left\{|x(t,\omega) - y(t,\omega)|^2\right\} = 0$$

for any $t \in [0, T]$, which implies

$$\mathsf{P}\{\omega : |x(t,\omega) - y(t,\omega)| = 0\} = 0$$

Theorem is proven. □

Remark 12.1. *The assumption 2 in Theorem 12.1 essentially requires that the functions f and g satisfy a Lipschitz condition and exhibit linear growth in the state variable, which is fairy restrictive. The conditions 2(a) and 2(b) facilitate an elementary proof of the existence and uniqueness of the solution (12.3) which is analogous to the classical Picard iteration proof (see Chapter 19 in Poznyak (2008)) designed in the ordinary differential equations theory. However, it is possible to show (Protter, 1977) that Theorem 12.1 remains valid even if the hypothesis 2(b) is removed: the global Lipschitz condition 2(a) suffices to guarantee the existence and uniqueness of the solution.*

Remark 12.2. *Notice also that if the Lipschitz condition 2(a) holds for a function f and*

1. *either f is independent of t,*
2. *or for some t the function f is bounded on x,*

then f will satisfy the growth condition 2(b) automatically.

12.1.2.2 Result with the local Lipschitz condition

Keeping the condition 2(b), the condition 2(a) can be relaxed to the local Lipshitz condition:

Condition 12.1. (2a′) *For any $N > 0$ there exists a constant K_N such that for all $t \in [0, T]$ and x, y satisfying $|x| \le N$, $|y| \le N$, the* ***local Lipschitz condition*** *holds, namely,*

$$\boxed{|f(t,x) - f(t,y)| + |g(t,x) - g(t,y)| \le K_N |x - y|} \tag{12.25}$$

[1] The Gronwall lemma states that if

$$x(t) \le v(t) + L \int_{s=0}^{t} x(s)\,ds$$

then

$$x(t) \le v(t) + L \int_{s=0}^{t} e^{L(t-s)} v(s)\,ds$$

for any measurable bounded functions $x(t)$ and $v(t)$.

Remark 12.3. *Note that the condition (12.25) holds whenever f and g are continuously differentiable in the second variable.*

We present the corresponding statement without proof.

Theorem 12.2. (Gihman and Skorohod, 1972) *The conclusion of Theorem 1 remains valid if the hypotheses other than 2(a) are satisfied, and 2(a′) holds.*

The local Lipschitz condition 2(a′), by itself, does not suffice to give global existence, however. The condition of global growth 2(b) cannot be removed from Theorem 12.2, which can be seen from the following simple ordinary value problem:

$$dx(t,\omega) = x^{\alpha}(t,\omega)\,dt, \quad x(0,\omega) = x_0(\omega), \quad \alpha > 1$$

which has the solution

$$x(t,\omega) = \begin{cases} 0 & \text{if} \quad x_0(\omega) = 0 \\ \left[1/x_0^{\alpha-1}(\omega) - (\alpha-1)t\right]^{1/(1-\alpha)} & \text{if} \quad x_0(\omega) \neq 0 \end{cases}$$

and exhibits an ‘*explosion*’ at time

$$t^* = \left[(\alpha-1)\,x_0^{\alpha-1}(\omega)\right]^{-1}$$

12.1.3 Dependence on parameters and on initial conditions

In this section we present the result which gives the conditions when the solutions $x(t,\omega)$ of (12.3) respond smoothly to smooth changes in the coefficient functions and initial conditions.

Theorem 12.3. (Gihman and Skorohod, 1972) *Suppose that the functions $f_n(t,x)$, $g_n(t,x)$ and $x_n(0,\omega)$ satisfy the conditions of Theorem 12.1 uniformly in n, that is, for the same K. Let $x_n(t,\omega)$ be the solution of the stochastic integral equation*

$$\boxed{\begin{aligned} x_n(t,\omega) \overset{a.s.}{=} x_n(0,\omega) + \int_{\tau=0}^{t} f_n(\tau, x_n(\tau,\omega))\,d\tau \\ + \int_{\tau=0}^{t} g_n(\tau, x_n(\tau,\omega))\,dW_{\tau}(\omega) \end{aligned}} \tag{12.26}$$

for all $n = 0, 1, \ldots$. If

1. for each N and each $t \in [0,T]$

$$\boxed{\begin{aligned} \lim_{n\to\infty} \sup_{x:|x|\le N} (|f_n(t,x) - f_0(t,x)| \\ + |g_n(t,x) - g_0(t,x)|) = 0 \end{aligned}} \tag{12.27}$$

2.

$$\boxed{\lim_{n\to\infty} \mathsf{E}\left\{|x_n(0,\omega) - x_0(0,\omega)|^2\right\} = 0} \tag{12.28}$$

Then

$$\boxed{\lim_{n\to\infty} \sup_{t\in[0,T]} \mathsf{E}\left\{|x_n(t,\omega) - x_0(t,\omega)|^2\right\} = 0} \tag{12.29}$$

Proof. The equation (12.26) can be represented as follows:

$$\begin{aligned} x_n(t,\omega) - x_0(t,\omega) \overset{a.s.}{=} & \int_{\tau=0}^{t} |f_n(\tau, x_n(\tau,\omega)) - f_n(\tau, x_0(\tau,\omega))|\, d\tau \\ & + \int_{\tau=0}^{t} |g_n(\tau, x_n(\tau,\omega)) - g_n(\tau, x_0(\tau,\omega))|\, dW_\tau(\omega) \\ & + \Delta y_n(t,\omega) \end{aligned}$$

where

$$\begin{aligned} \Delta y_n(t,\omega) \overset{a.s.}{=} & \; x_n(0,\omega) - x_0(0,\omega) \\ & + \int_{\tau=0}^{t} [f_n(\tau, x_0(\tau,\omega)) - f_0(\tau, x_0(\tau,\omega))]\, d\tau \\ & + \int_{\tau=0}^{t} [g_n(\tau, x_0(\tau,\omega)) - g_n(\tau, x_0(\tau,\omega))]\, dW_\tau(\omega) \end{aligned}$$

which, in view of the global Lipschitz condition and the estimates of the previous Theorem 12.1, leads to

$$\begin{aligned} \mathsf{E}\left\{|x_n(t,\omega) - x_0(t,\omega)|^2\right\} \le & \; 3\mathsf{E}\left\{[\Delta y_n(t,\omega)]^2\right\} \\ & + L\int_{\tau=0}^{t} \mathsf{E}\left\{|x_n(\tau,\omega) - x_0(\tau,\omega)|^2\right\} d\tau \end{aligned}$$

with $L = 3(T+1)K^2$. Applying again the Gronwall lemma we obtain

$$\begin{aligned} \mathsf{E}\left\{|x_n(t,\omega) - x_0(t,\omega)|^2\right\} \le & \; 3\mathsf{E}\left\{[\Delta y_n(t,\omega)]^2\right\} \\ & + L\int_{\tau=0}^{t} e^{L(t-\tau)}\mathsf{E}\left\{[\Delta y_n(\tau,\omega)]^2\right\} d\tau \end{aligned}$$

So, to demonstrate (12.29) it is sufficient to show that

$$\sup_{t\in[0,T]} \mathsf{E}\left\{[\Delta y_n(t,\omega)]^2\right\} \underset{n\to\infty}{\to} 0 \tag{12.30}$$

Toward establishing (12.30), the Cauchy–Schwartz inequality argument can be applied which gives

$$\mathsf{E}\left\{\left[\int_{\tau=0}^{t} [f_n(\tau, x_0(\tau,\omega)) - f_0(\tau, x_0(\tau,\omega))]\,d\tau\right]^2\right\}$$
$$\le T\mathsf{E}\left\{\int_{\tau=0}^{t} [f_n(\tau, x_0(\tau,\omega)) - f_0(\tau, x_0(\tau,\omega))]^2\,d\tau\right\}$$

By assumption (12.27) and the Lebesgue dominated convergence, Theorem 6.2, the right-hand side of the last inequality tends to zero, so that as $n \to \infty$,

$$\sup_{t\in[0,T]} \mathsf{E}\left\{\left[\int_{\tau=0}^{t} [f_n(\tau, x_0(\tau,\omega)) - f_0(\tau, x_0(\tau,\omega))]\,d\tau\right]^2\right\} \underset{n\to\infty}{\to} 0 \tag{12.31}$$

Since

$$h_n(t,\omega) := g_n(t, x_0(t,\omega)) - g_0(t, x_0(t,\omega)) \in \mathcal{L}^2[0,T]$$

one can apply Doob's inequality (11.71) for the moment of the maximum modulus (supremum) of a submartingale, which implies

$$\mathsf{E}\left\{\sup_{t\in[0,T]}\left[\int_{\tau=0}^{t} h_n(\tau,\omega)\,dW_\tau(\omega)\right]^2\right\} \le 4\mathsf{E}\left\{\int_{\tau=0}^{t} |h_n(\tau,\omega)|^2\,d\tau\right\} \underset{n\to\infty}{\to} 0 \tag{12.32}$$

by assumption (12.27) and, again, by the Lebesgue dominated convergence theorem 6.2. The property (12.32) together with (12.31) proves (12.30). Theorem is proven. □

The next corollary represents the more practical conditions to the functions involved which guarantees the continue dependence of the solution on an initial value.

Corollary 12.1. **(Gihman and Skorohod, 1972)** *Assume that functions $f(t,x)$ and $g(t,x)$ satisfy the conditions of Theorem 12.1. For any $z \in \mathbb{R}$ let $x(t,\omega \mid z)$ denote the solution of (12.3) defined for all $t \in [0,T]$ with the initial value*

$$\boxed{x(0,\omega) = x_0(\omega) = z} \tag{12.33}$$

namely

$$\boxed{\begin{aligned} x(t,\omega \mid z) \overset{a.s.}{=} z &+ \int_{\tau=0}^{t} f(\tau, x(\tau,\omega \mid z))\, d\tau \\ &+ \int_{\tau=0}^{t} g(\tau, x(\tau,\omega \mid z))\, dW_\tau(\omega) \end{aligned}} \tag{12.34}$$

Suppose also that for all $t \in [0, T]$ *and all* $x \in \mathbb{R}$ ***there exist the derivatives***

$$\frac{\partial}{\partial x} f(t,x) \quad \text{and} \quad \frac{\partial}{\partial x} g(t,x)$$

which are ***continuous*** *and* ***bounded****. Then the derivative of the solution with respect to the initial value (the '****coefficient of sensibility****')*

$$y(t,\omega \mid z) := \frac{\partial}{\partial z} x(t,\omega \mid z)$$

exists (in the mean-square sense) and satisfies the following ***linear stochastic integral equation*** *called* ***'the sensitivity equation'***

$$\boxed{\begin{aligned} y(t,\omega \mid z) = 1 &+ \int_{\tau=0}^{t} \frac{\partial}{\partial x} f(\tau, x(\tau,\omega \mid z))\, y(\tau,\omega \mid z)\, d\tau \\ &+ \int_{\tau=0}^{t} \frac{\partial}{\partial x} g(\tau, x(\tau,\omega \mid z))\, y(\tau,\omega \mid z)\, dW_\tau(\omega) \end{aligned}} \tag{12.35}$$

Proof. (the main scheme) By the conditions to the derivatives of the functions involved, it follows that for different z the solution of (12.34) satisfies all conditions of Theorem 12.3. Then, taking the limit in $[x(t,\omega \mid z+\delta) - x(t,\omega \mid z)]/\delta$ when $\delta \to 0$, we obtain the desired result. □

12.1.4 Moments of solutions

The result below demonstrates that the upper estimates for the even order moments of the solution of (12.3) are similar to the estimates obtained for the moments of stochastic integrals given in Theorem 11.4.

Theorem 12.4. (Gut, 2005) *Suppose that the functions* $f(t,x)$ *and* $g(t,x)$ *satisfy the conditions of Theorem 12.1 guaranteeing the existence and uniqueness of solutions of (12.3). If for some positive integer* n *we have*

$$\boxed{\mathsf{E}\left\{|x_0(\omega)|^{2n}\right\} < \infty} \tag{12.36}$$

then for any $t \in [0, T]$ the solution $x(t, \omega)$ of (12.3) with the initial value $x(0, \omega) = x_0(\omega)$ satisfies the inequalities

$$\boxed{\begin{aligned} \mathsf{E}\left\{|x(t,\omega)|^{2n}\right\} &\leq \left(1+\mathsf{E}\left\{|x_0(\omega)|^{2n}\right\}\right)e^{Ct} \\ \mathsf{E}\left\{|x(t,\omega)-x(0,\omega)|^{2n}\right\} &\leq D\left(1+\mathsf{E}\left\{|x_0(\omega)|^{2n}\right\}\right)t^n e^{Ct} \end{aligned}} \tag{12.37}$$

where $C = 2n(2n+1)K^2$ and D is constant depending only on n, K and T.

Proof. Define the following truncated functions:

$$f_N(t,x) := \begin{cases} f(t,x) & \text{if } |x| \leq N \\ f(t, Nx/|x|) & \text{if } |x| > N \end{cases}$$

$$g_N(t,x) := \begin{cases} g(t,x) & \text{if } |x| \leq N \\ g(t, Nx/|x|) & \text{if } |x| > N \end{cases}$$

and

$$x_{N,0}(\omega) := \begin{cases} x_0(\omega) & \text{if } |x_0(\omega)| \leq N \\ Nx_0(\omega)/|x_0(\omega)| & \text{if } |x_0(\omega)| > N \end{cases}$$

The truncated process $\{x_N(0,\omega)\}_{t\geq 0}$, which is the solution of the stochastic initial value problem

$$\begin{aligned} dx_N(t,\omega) &= f_N(t, x_N(t,\omega))\,dt + g_N(t, x_N(t,\omega))\,dW_t(\omega) \\ x_N(0,\omega) &= x_{N,0}(\omega) \end{aligned} \tag{12.38}$$

converges uniformly on $[0, T]$ with probability 1 to $x(t, \omega)$ as $n \to \infty$. By applying the Itô formula (11.65) to $V(t, x) = |x|^{2n}$ when $x = x_N(t, \omega)$, one obtains

$$\begin{aligned} |x_N(t,\omega)|^{2n} = {} & \left|x_{N,0}(\omega)\right|^{2n} \\ & + \int_{\tau=0}^{t} 2n\,|x_N(\tau,\omega)|^{2n-2} x_N(\tau,\omega) f_N(\tau, x_N(\tau,\omega))\,d\tau \\ & + \int_{\tau=0}^{t} n\,|x_N(\tau,\omega)|^{2n-2} |g_N(\tau, x_N(\tau,\omega))|^2\,d\tau \\ & + \int_{\tau=0}^{t} 2n(n-1)\,|x_N(\tau,\omega)|^{2n-4} |x_N(\tau,\omega)\,g_N(\tau, x_N(\tau,\omega))|^2\,d\tau \\ & + \int_{\tau=0}^{t} 2n\,|x_N(\tau,\omega)|^{2n-2} x_N(\tau,\omega)\,g_N(\tau, x_N(\tau,\omega))\,dW_\tau(\omega) \end{aligned}$$

Notice that the expected value of the last integral is zero due to the boundedness of the integrand. So, applying the growth condition, we get

$$\begin{aligned}
\mathsf{E}\left\{|x_N(t,\omega)|^{2n}\right\} &= \mathsf{E}\left\{\left|x_{N,0}(\omega)\right|^{2n}\right\} \\
&\quad + \int_{\tau=0}^{t} \mathsf{E}\left\{2n\,|x_N(\tau,\omega)|^{2n-2}\, x_N(\tau,\omega)\, f_N(\tau, x_N(\tau,\omega))\right\} d\tau \\
&\quad + \int_{\tau=0}^{t} n\mathsf{E}\left\{|x_N(\tau,\omega)|^{2n-2}\, |g_N(\tau, x_N(\tau,\omega))|^2\right\} d\tau \\
&\quad + \int_{\tau=0}^{t} 2n(n-1)\,\mathsf{E}\left\{|x_N(\tau,\omega)|^{2n-4}\, |x_N(\tau,\omega)\, g_N(\tau, x_N(\tau,\omega))|^2\right\} d\tau \\
&\le \mathsf{E}\left\{\left|x_{N,0}(\omega)\right|^{2n}\right\} \\
&\quad + (2n+1)\,K^2 \int_{\tau=0}^{t} \mathsf{E}\left\{\left(1+|x_N(\tau,\omega)|^2\right) |x_N(\tau,\omega)|^{2n-2}\right\} d\tau
\end{aligned}$$

Now, making use of the inequality

$$\left(1+x^2\right) x^{2n-2} \le 1 + 2x^{2n}$$

we obtain from the last relation

$$\begin{aligned}
\mathsf{E}\left\{|x_N(t,\omega)|^{2n}\right\} &\le \mathsf{E}\left\{\left|x_{N,0}(\omega)\right|^{2n}\right\} + (2n+1)\,nK^2 t \\
&\quad + (2n+1)\,2nK^2 \int_{\tau=0}^{t} \mathsf{E}\left\{|x_N(\tau,\omega)|^{2n}\right\} d\tau
\end{aligned}$$

Applying again the Gronwall lemma we derive

$$\begin{aligned}
\mathsf{E}\left\{|x_N(t,\omega)|^{2n}\right\} &\le h(t) \\
&\quad + (2n+1)\,2nK^2 \int_{\tau=0}^{t} \exp\left\{(2n+1)\,2nK^2(t-\tau)\right\} h(\tau)\, d\tau
\end{aligned}$$

where

$$h(t) := \mathsf{E}\left\{\left|x_{N,0}(\omega)\right|^{2n}\right\} + (2n+1)\,nK^2 t$$

Carrying out the integration in the last inequality yields (12.37) for $x_N(t,\omega)$. Letting $N \to \infty$ obtains the desired result for $x(t,\omega)$.

The second inequality in (12.37) follows directly from the first one and the relation (12.3). Theorem is proven. □

12.2 Solutions as diffusion processes

Diffusion processes (see Definition 10.5) are Markov processes whose probability law is specified by the *drift* and *diffusion* coefficients (10.20) and (10.21) corresponding to the conditional 'infinitesimal' mean and variance of the process, respectively.

Here we will show that

- any solution of a stochastic differential equation (12.1) (if it exists) is a diffusion Markov process (Definition 10.5) with the drift coefficient $f(t, x(t, \omega))$ and the diffusion coefficient $g^2(t, x(t, \omega))$;
- inversely, under some conditions being specified, a diffusion process (Definition 10.5)

  $$dy(t, \omega) = a(t, y(t, \omega))\,dt + b(t, y(t, \omega))\,dW_t(\omega), \quad b(t, y) \geq 0$$

  shares the same probability law as the solution of the stochastic differential equation (12.1)

  $$dy(t, \omega) = a(t, y(t, \omega))\,dt + \sqrt{b(t, y(t, \omega))}d\tilde{W}_t(\omega)$$

  (where $\tilde{W}_t(\omega)$ is another Wiener process).

Consider again the stochastic differential equation (12.1)

$$dx(t, \omega) = f(t, x(t, \omega))\,dt + g(t, x(t, \omega))\,dW_t(\omega)$$

defined on the interval $[0, T]$ with a specific initial condition $x(0, \omega) = x_0(\omega)$.

12.2.1 General Markov property

Theorem 12.5. ***The solution*** $x(t, \omega)$ *of (12.1) with* $x(0, \omega) = x_0(\omega)$ ***is a Markov process*** *on the interval* $[0, T]$ *with the initial distribution*

$$\mathsf{P}\{x(0, \omega) \in A\} = \mathsf{P}_0\{A\}$$

and the transition probabilities given by

$$\mathsf{P}\{s, x, t, A\} = \mathsf{P}\{x(t, \omega) \in A \mid x(s, \omega) = x\} \quad \textit{for all} \quad 0 \leq s \leq t \leq T$$

Proof. Let, as before, $x(t, \omega) \in \mathbb{R}^n$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathsf{P})$ with state space $\mathbb{R}^n$ and the index set $:= [0, T]$. For any $0 \leq s \leq t \leq T$ define

$$\mathcal{W}_{[0,t]} := \sigma\{x(0, \omega), W_s(\omega), 0 \leq s \leq t\} \tag{12.39}$$

as a minimal sigma-algebra (sub-sigma-algebra of $\mathcal{F}$) generated by $x(0, \omega)$ and $W_s(\omega)$ $(0 \leq s \leq t)$. To verify the Markov property for the given process $\{x(t, \omega)\}_{t \in J}$ one needs to show (10.3), namely,

$$\mathsf{P}\left\{x(t, \omega) \in A \mid \mathcal{F}_{[0,s]}\right\} \overset{a.s.}{=} \mathsf{P}\{x(t, \omega) \in A \mid x(s, \omega)\} \tag{12.40}$$

where $\mathcal{F}_{[0,s]} := \sigma\{x(t, \omega), t_1 \leq t \leq t_2\}$ is defined by (12.39). Notice that since $x(t, \omega)$ is $\mathcal{W}_{[0,t]}$-measurable, it follows that $\mathcal{F}_{[0,s]} \subseteq \mathcal{W}_{[0,t]}$. Therefore, to prove (12.40) it is sufficient to establish the stronger condition

$$\mathsf{P}\left\{x(t,\omega)\in A \mid \mathcal{W}_{[0,s]}\right\} \overset{a.s.}{=} \mathsf{P}\left\{x(t,\omega)\in A \mid x(s,\omega)\right\} \tag{12.41}$$

In turn, to verify (12.41), it suffices to prove that for every bounded measurable function

$$h(x,\omega) := \sum_{i=1}^{N} h_i(x)\, H_i(\omega) \tag{12.42}$$

(where $H_i(\omega)$ is a random variable independent of $\mathcal{W}_{[0,s]}$) the following property holds:

$$\mathsf{E}\left\{h(x(s,\omega),\omega) \mid \mathcal{W}_{[0,s]}\right\} \overset{a.s.}{=} \mathsf{E}\left\{h(x(s,\omega),\omega) \mid x(s,\omega)\right\} \tag{12.43}$$

This would imply that the property (12.43) holds for the class of all bounded measurable functions $h(x,\omega)$ (not only given by (12.42)) since the subclass (12.42) is dense within this class. Then, in particular, taking

$$h(x,\omega) = \chi(x(s,\omega)\in A)$$

together with the semigroup property

$$x(t,\omega \mid x(0,\omega)) = x_0) = x(t,\omega \mid x(s,\omega \mid x(0,\omega) = x_0)))$$

this leads to (12.41). But, in view of independency $H_i(\omega)$ of $x(s,\omega)$, we have

$$\begin{aligned}
\mathsf{E}\left\{h(x(s,\omega),\omega) \mid \mathcal{W}_{[0,s]}\right\} &= \sum_{i=1}^{N} \mathsf{E}\left\{h_i(x(s,\omega))\, H_i(\omega) \mid \mathcal{W}_{[0,s]}\right\} \\
&\overset{a.s.}{=} \sum_{i=1}^{N} h_i(x(s,\omega))\, \mathsf{E}\left\{H_i(\omega) \mid \mathcal{W}_{[0,s]}\right\} \\
&= \sum_{i=1}^{N} h_i(x(s,\omega))\, \mathsf{E}\left\{H_i(\omega)\right\} \\
&= \mathsf{E}\left\{h(x(s,\omega),\omega) \mid x(s,\omega)\right\}
\end{aligned}$$

which completes the proof. □

12.2.2 *Solution as a diffusion process*

Theorem 12.6. *Assume that f and g in (12.1) satisfy the conditions of Theorem 12.1 guaranteeing the existence and the uniqueness of a solution $x(t,\omega)$ of (12.3) defined for all $t\in[0,T]$. Then any solution of (12.1) is a diffusion process on $[0,T]$ with drift coefficient $f(t,x)$ and diffusion coefficient $g^2(t,x)$.*

Proof. We need to verify the properties (10.22), namely,

$$\mathsf{P}\left\{|x(t+h,\omega)-x(t,\omega)|>\varepsilon \mid x(s,\omega)=x\right\} = o(h)$$

$$\begin{aligned}
&\mathsf{E}\{[x(t+h,\omega)-x(t,\omega)] \\
&\qquad \times \chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon) \mid x(s,\omega)=x\} \\
&\quad = a(s,x)\,h + o(h)
\end{aligned}$$

$$\begin{aligned}&\mathsf{E}\{[x(t+h,\omega)-x(t,\omega)][x(t+h,\omega)-x(t,\omega)]^{\mathsf{T}}\\&\qquad\chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon)\mid x(s,\omega)=x\}\\&\quad=B(s,x)h+o(h)\end{aligned}$$

To prove the first one notice that by the Markov inequality (4.8) and using the property (12.37) it follows (for $r=4$)

$$\begin{aligned}&\mathsf{P}\{|x(t+h,\omega)-x(t,\omega)|>\varepsilon\mid x(s,\omega)=x\}\\&\quad\le\varepsilon^{-4}\mathsf{E}\left\{|x(t+h,\omega)-x(t,\omega)|^4\right\}=\le\varepsilon^{-4}Ch^2\end{aligned}$$

where $C=C(T)$ is a constant. This inequality implies the desired result. The proof of the last two properties can be done similarly invoking the Cauchy–Schwartz inequality (4.16) using the Lipschitz condition. Indeed, we have

$$\begin{aligned}&\mathsf{E}\{[x(t+h,\omega)-x(t,\omega)]\\&\quad\times\chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon)\mid x(s,\omega)=x\}\\&=\int_{u=s}^{t}\mathsf{E}\{f(u,x(u,\omega))\chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon)\}du=\int_{u=s}^{t}f(u,x)du\\&\quad+\int_{u=s}^{t}[\mathsf{E}\{f(u,x(u,\omega))\chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon)\}-f(u,x)]du\end{aligned}$$

and

$$\begin{aligned}&\left|\int_{u=s}^{t}[\mathsf{E}\{f(u,x(u,\omega))\chi(|x(t+h,\omega)-x(t,\omega)|\le\varepsilon)\}-f(u,x)]du\right|\\&\le\int_{u=s}^{t}\mathsf{E}\{|f(u,x(u,\omega))-f(u,x)|\}du\\&\le(t-s)^{1/2}\left(\int_{u=s}^{t}\mathsf{E}\left\{|f(u,x(u,\omega))-f(u,x)|^2\right\}du\right)^{1/2}\le O\left((t-s)^{3/2}\right)\end{aligned}$$

By the continuity of $f(u,x)$ we also have

$$\begin{aligned}\int_{u=s}^{t}f(u,x)du&=f(s,x)(t-s)+\int_{u=s}^{t}[f(u,x)-f(s,x)]du\\&=f(s,x)(t-s)+o(t-s)\end{aligned}$$

which together with the relations above implies the second desired result. The third one can be proven similarly. Theorem is proven. □

Remark 12.4. *In the vector case the drift coefficient* $a(t,x)$ *of the process* $\{x(t,\omega)\}_{t\in[0,T]}$ *is a vector function* $f(t,x)$ *and the diffusion matrix* $B(t,x)$ *is* $G(t,x)\,G^{\mathsf{T}}(t,x)$*, that is,*

$$\boxed{\begin{array}{c} a(t,x) = f(t,x) \\ B(t,x) = G(t,x)\,G^{\mathsf{T}}(t,x) \end{array}} \tag{12.44}$$

Let us now show the inversion.

Theorem 12.7. *Suppose that* $\{y(t,\omega)\}_{t\in[0,T]}$ *is a diffusion process*

$$dy(t,\omega) = a(t,y(t,\omega))\,dt + b(t,y(t,\omega))\,dW_t(\omega), \quad b(t,y) \geq 0$$

with $a(t,y)$ *and* $b(t,y)$ *satisfying the conditions in Theorem 12.1. Then there exists a Wiener process* $\tilde{W}_t(\omega)$ *for which* $y(t,\omega)$ *solves the stochastic differential equation*

$$\boxed{dy(t,\omega) = a(t,y(t,\omega))\,dt + \sqrt{b(t,y(t,\omega))}\,d\tilde{W}_t(\omega)} \tag{12.45}$$

Proof. (the principal scheme) Let

$$z(t,\omega) := g(t,y(t,\omega)) \tag{12.46}$$

where

$$g(t,y) = \int\limits_{v=0}^{y} \frac{dv}{\sqrt{b(t,v)}} \tag{12.47}$$

Define the function

$$\bar{a}(t,z) := \left[\frac{\partial}{\partial t} g(t,y) + a(t,y)\,\frac{\partial}{\partial y} g(t,y) + \frac{1}{2} b(t,y)\,\frac{\partial^2}{\partial y^2} g(t,y)\right]_{y=g^{-1}(t,z)}$$

Then $z(t,\omega)$ is a diffusion process with the drift $\bar{a}(t,z)$ and the diffusion coefficient unity. Define then

$$\tilde{W}_t(\omega) = z(t,\omega) - z(0,\omega) - \int \bar{a}(u,z(u,\omega))\,du \tag{12.48}$$

which, in fact, is a Wiener process that can be shown using the accepted assumptions. Then (12.48) is equivalent to the stochastic differential equation

$$dz(t,\omega) = \bar{a}(t,z(t,\omega))\,dt + d\tilde{W}_t(\omega)$$

Next, applying the Itô rule (11.65) to (12.46) and (12.47), we find that $y(t,\omega)$ is a solution of (12.45). □

12.3 Reducing by change of variables

Here we will discuss the method which provides a technique for solving a class of stochastic differential equations by quadratures. This method is related to a change of variables which leads to a simplification (or reducing) of the stochastic differential equation initially given.

12.3.1 General description of the method

Consider, as usual, the vector stochastic differential equation

$$\begin{aligned} &dx(t,\omega) = f(t, x(t,\omega))\,dt + G(t, x(t,\omega))\,dW_t(\omega) \\ &x(t,\omega) \in \mathbb{R}^n, \quad W_t(\omega) \in \mathbb{R}^L, \quad x(0,\omega) = x_0(\omega) \end{aligned} \tag{12.49}$$

and a smooth vector valued function $H : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ transforming the state vector x into a new state vector y as

$$y = H(t, x) \tag{12.50}$$

and suppose that for all $t \geq 0$ and all $x \in \mathbb{R}^n$ the following condition holds:

$$\det\left[\frac{\partial}{\partial x} H(t,x)\right] \neq 0 \tag{12.51}$$

which leads to the existence of the inverse function $H^{-1} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ such that

$$x = H^{-1}(t, y) \tag{12.52}$$

By Itô formula (11.64) implementation it follows that

$$dy(t,\omega) = \tilde{f}(t, y(t,\omega))\,dt + \tilde{G}(t, y(t,\omega))\,dW_t(\omega) \tag{12.53}$$

where

$$\begin{aligned} \tilde{f}(t,y) &= \left[\frac{\partial}{\partial t} H(t,x) + \frac{\partial}{\partial x} H(t,x)^{\intercal} f(t,x) \right. \\ &\qquad \left. + \frac{1}{2}\operatorname{tr}\left(G(t,x)\,G(t,x)\,\frac{\partial^2}{\partial x^2} H(t,x)\right)\right]_{x=H^{-1}(t,y)} \\ \tilde{G}(t,y) &= \left(\left[\frac{\partial}{\partial x} H(t,x)\right] G(t,x)\right)\Big|_{x=H^{-1}(t,y)} \end{aligned} \tag{12.54}$$

The simplification or reducing of the initial stochastic differential equation will take place if the new drift vector $\tilde{f}(t,y)$ and the new diffusion matrix $\tilde{G}(t,y)$ are independent of y, namely, if

$$\tilde{f}(t,y) := \bar{f}(t), \quad \tilde{G}(t,y) = \bar{G}(t) \tag{12.55}$$

or equivalently, if for all $t \geq 0$

$$\boxed{\frac{\partial}{\partial y}\tilde{f}(t, y) = 0, \quad \frac{\partial}{\partial y}\tilde{G}(t, y) = 0} \tag{12.56}$$

In this case the change of variables (12.50) permits the explicit representation of the solution $x(t, \omega)$ as

$$\boxed{\begin{gathered} x(t, \omega) = H^{-1}(t, y(t, \omega)) \\ y(t, \omega) = H(0, x_0(\omega)) + \int_{s=0}^{t} \bar{f}(s)\, ds + \int_{s=0}^{t} \bar{G}(s) dW_s(\omega) \end{gathered}} \tag{12.57}$$

12.3.2 Scalar stochastic differential equations

In the scalar case the condition (12.56) can be easily verified leading to the following result.

Theorem 12.8. (on the reducibility) *The scalar* ($n = L = 1$) *stochastic differential equation (12.49) is reducible if and only if the functions* f *and* G *satisfy the equation*

$$\boxed{\frac{\partial}{\partial x}\left[G\left(G^{-2}\frac{\partial}{\partial t}G - \frac{\partial}{\partial x}\left[\frac{f}{G}\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}G\right)\right] = 0} \tag{12.58}$$

Proof. Evidently (12.56) is fulfilled if and only if

$$\frac{\partial}{\partial t}H(t, x) + f(t, x)\frac{\partial}{\partial x}H(t, x) + \frac{1}{2}G^2(t, x)\frac{\partial^2}{\partial x^2}H(t, x) = \varphi(t)$$

$$\left[\frac{\partial}{\partial x}H(t, x)\right]G(t, x) = g(t)$$

Differentiation of the first equation on x gives

$$\frac{\partial^2}{\partial x \partial t}H(t, x) + \frac{\partial}{\partial x}\left[f(t, x)\frac{\partial}{\partial x}H(t, x) + \frac{1}{2}G^2(t, x)\frac{\partial^2}{\partial x^2}H(t, x)\right] = 0 \tag{12.59}$$

From the second equation it follows that

$$\frac{\partial}{\partial x}H(t, x) = \frac{g(t)}{G(t, x)} \tag{12.60}$$

and

$$\frac{\partial^2}{\partial t \partial x} H(t,x) = \frac{g'(t)\, G(t,x) - g(t) \frac{\partial}{\partial t} G(t,x)}{G^2(t,x)}$$

$$\frac{\partial^2}{\partial x^2} H(t,x) = -\frac{g(t)}{G^2(t,x)} \frac{\partial}{\partial x} G(t,x)$$

Substitution of these relations into (12.59) implies

$$\frac{g'(t)}{G(t,x)} - g(t) \left[\frac{\frac{\partial}{\partial t} G(t,x)}{G^2(t,x)} - \frac{\partial}{\partial x} \left(\frac{f(t,x)}{G(t,x)} \right) + \frac{1}{2} \frac{\partial^2}{\partial x^2} G(t,x) \right] = 0$$

or equivalently,

$$\frac{g'(t)}{g(t)} = G(t,x) \left[\frac{\frac{\partial}{\partial t} G(t,x)}{G^2(t,x)} - \frac{\partial}{\partial x} \left(\frac{f(t,x)}{G(t,x)} \right) + \frac{1}{2} \frac{\partial^2}{\partial x^2} G(t,x) \right]$$

Since the left-hand side is independent of x the property (12.58) follows. Theorem is proven. □

Corollary 12.2. *For stationary (autonomous) scalar stochastic differential equations (12.49) with*

$$f = f(x), \quad G = G(x)$$

the condition (12.58) becomes

$$\boxed{G \left(\frac{\partial}{\partial x} \left[\frac{f}{G} \right] - \frac{1}{2} \frac{\partial^2}{\partial x^2} G \right) = \text{const}} \tag{12.61}$$

and (12.60) gives

$$\boxed{H(t,x) = g(t) \int_{v=a}^{x} \frac{dv}{G(v)}} \tag{12.62}$$

for arbitrary a.

Example 12.1. *Consider the autonomous equation*

$$\boxed{dx(t,\omega) = [f_0(t) + a_0 x(t,\omega)]\, dt + G dW_t(\omega)}$$

with the constant diffusion coefficient $G > 0$ *and a constant parameter* $a_0 = \text{const}$ *in the drift. Then the condition (12.61) holds, and therefore, by (12.62) with* $a = 0$ *and (12.54) it follows that*

$$H(t,x) = g(t)\frac{x}{G} = y$$
$$x(t,\omega) = \frac{G}{g(t)}y(t,\omega)$$
$$y(t,\omega) = H(0, x_0(\omega)) + \int_{s=0}^{t} \bar{f}(s)\,ds + \int_{s=0}^{t} \bar{G}(s)dW_s(\omega)$$

where (with $g(t) = e^{-a_0 t}$*)*

$$\bar{f}(t) = g'(t)\frac{x}{G} + \frac{g(t)}{G}\left[f_0(t) + a_0 x(t,\omega)\right] = \frac{e^{-a_0 t}}{G} f_0(t)$$

and

$$\bar{G}(t) = e^{-a_0 t}$$

So, finally,

$$\boxed{\begin{aligned} x(t,\omega) = e^{a_0 t}x(0,\omega) + \int_{s=0}^{t} e^{a_0(t-s)} f_0(s)ds \\ + \int_{s=0}^{t} e^{a_0(t-s)} G dW_s(\omega) \end{aligned}} \tag{12.63}$$

which practically coincides with the solution of the usual (non-stochastic) differential equation.

Example 12.2. *Consider the scalar non-autonomouse linear stochastic equation*

$$\boxed{\begin{aligned} d\mathring{x}(t,\omega) = a(t)\mathring{x}(t,\omega)\,dt + g_1(t)\mathring{x}(t,\omega)\,dW(\omega) \\ \mathring{x}(0,\omega) = \mathring{x}_0(\omega) \end{aligned}} \tag{12.64}$$

The reducibility condition (12.58) is fulfilled automatically since in this case

$$f = a(t)\mathring{x}, \quad G = g_1(t)\mathring{x}$$

So, by (12.60)

$$H(t,\mathring{x}) = g(t)\int_{v=a}^{\mathring{x}} \frac{dv}{G(t,v)} = \frac{g(t)}{g_1(t)}\int_{v=a}^{\mathring{x}} \frac{dv}{v}$$
$$= \frac{g(t)}{g_1(t)} \ln \mathring{x}/\mathring{x}_0 = \ln \mathring{x}/\mathring{x}_0$$

with $a = \mathring{x}_0$, $g(t) = g_1(t)$ *and for* $\mathring{x}/\mathring{x}_0 > 0$. *Therefore, the explicit solution (12.57) becomes*

$$\mathring{x}(t,\omega) = \mathring{x}_0(\omega)\exp\{y(t,\omega)\}$$

$$y(t,\omega) = \int_{s=0}^{t} \bar{f}(s)\,ds + \int_{s=0}^{t} \bar{G}(s)dW_s(\omega)$$

$$\begin{aligned}\tilde{f}(t,y) &= \left[\frac{\partial}{\partial t}H(t,x) + \frac{\partial}{\partial x}H(t,\mathring{x})^{\intercal} f(t,x)\right.\\ &\quad \left. + \frac{1}{2}\mathrm{tr}\left(G(t,x)\,G(t,x)\,\frac{\partial^2}{\partial x^2}H(t,x)\right)\right]_{x=H^{-1}(t,y)}\\ &= a(t) - \frac{1}{2}g_1^2(t) = \bar{f}(t)\end{aligned}$$

$$\tilde{G}(t,y) = \left(\left[\frac{\partial}{\partial x}H(t,x)\right]G(t,x)\right)|_{x=H^{-1}(t,y)} = g_1(t) = \bar{G}(t)$$

and finally,

$$\boxed{\begin{aligned}&\mathring{x}(t,\omega) = \mathring{x}_0(\omega)\\ &\qquad\cdot\exp\left\{\int_{s=0}^{t}\left[a(s) - \frac{1}{2}g_1^2(s)\right]ds + \int_{s=0}^{t} g_1(s)\,dW_s(\omega)\right\}\end{aligned}} \tag{12.65}$$

12.4 Linear stochastic differential equations

12.4.1 Fundamental matrix

Consider here the general vector linear stochastic differential equation given by

$$\boxed{\begin{aligned}&dx(t,\omega) = [a_0(t) + A(t)x(t,\omega)]\,dt\\ &\qquad + \sum_{i=1}^{m}\left[g_{0,i}(t) + G_{1,i}(t)x(t,\omega)\right]dW_{i,t}(\omega)\end{aligned}} \tag{12.66}$$

with $x(0,\omega) = x_0(\omega)$ where

$\{x(t,\omega)\}_{t\in[0,T]}$ is a random n-vector process,

$W_{i,t}(\omega)$ is a standard Wiener process ($W_{i,t}(\omega) \in \mathbb{R}$, $i = 1,\ldots,m$), i.e., for

$$W_t(\omega) := \left(W_{1,t}(\omega),\ldots,W_{1,m}(\omega)\right)^{\intercal}$$

one has

$$\mathsf{E}\left\{W_t(\omega)\,W_t^{\intercal}(\omega)\right\} = t\cdot I_{m\times m}$$

$A(t)$ and $g_{0,i}(t)$ are $n\times n$ matrix function and n-vector of $t\in[0,T]$, and $a_0(t)$ and $G_{1,i}(t)$ are n-vector and $n\times n$-matrix functions, respectively.

A1. *Suppose that the functions $a_0(t)$, $A(t)$ and $g_{0,i}(t)$, $G_{1,i}(t)$ $(i = 1, \ldots, m)$ are measurable and bounded on* $[0, T]$.

If the the assumption A1 is fulfilled then by Theorem 12.1 there exists a unique solution $x(t, \omega)$ satisfying (12.66).

Definition 12.2. *The matrix $\Phi_\omega(t, s) \in \mathbb{R}^{n\times n}$ (which is stochastic, in general) is called the* ***fundamental matrix*** *of the homogeneous version of the linear stochastic equation*

$$\boxed{\begin{gathered} d\mathring{x}(t,\omega) = A(t)\mathring{x}(t,\omega)\,dt + \sum_{i=1}^{m} G_{1,i}(t)\mathring{x}(t,\omega)\,dW_{i,t}(\omega) \\ \mathring{x}(0,\omega) = \mathring{x}_0(\omega) \end{gathered}} \tag{12.67}$$

if $\mathring{x}(t,\omega)$, satisfying (12.67), for any $s \in [0, t]$ and any $\mathring{x}_s(\omega) = \mathring{x}(s,\omega)$ can be represented as

$$\boxed{\mathring{x}(t,\omega) \overset{a.s.}{=} \Phi_\omega(t,s)\,\mathring{x}(s,\omega)} \tag{12.68}$$

The basic properties of the fundamental matrix $\Phi_\omega(t, s)$ are presented in the following lemma.

Lemma 12.1. *For the fundamental matrix $\Phi_\omega(t, s)$, associated with the homogeneous linear stochastic equation (12.67) the following '**semigroup**' properties hold:*

1. *for any* $0 \le t_0 \le s \le t \le T$

$$\boxed{\begin{aligned} \Phi_\omega(t,t_0) &\overset{a.s.}{=} \Phi_\omega(t,s)\,\Phi_\omega(s,t_0) \\ \Phi_\omega(t,s) &= \Phi_\omega(t,t_0)\,\Phi_\omega^{-1}(s,t_0) \end{aligned}} \tag{12.69}$$

2.

$$\boxed{\begin{aligned} \Phi_\omega(t,s) \overset{a.s.}{=} I &+ \int_{\tau=s}^{t} A(\tau)\Phi_\omega(\tau,s)\,d\tau \\ &+ \int_{\tau=s}^{t} \sum_{i=1}^{m} G_{1,i}(\tau)\Phi_\omega(\tau,s)\,dW_{i,\tau}(\omega) \end{aligned}} \tag{12.70}$$

or equivalently, taking $s = 0$

$$\boxed{\begin{aligned} d\Phi_\omega(t,0) &= A(t)\Phi_\omega(t,0)\,dt \\ &+ \sum_{i=1}^{m} G_{1,i}(t)\Phi_\omega(t,0)\,dW_{i,t}(\omega)\,, \quad \Phi_\omega(0,0) = I \end{aligned}} \tag{12.71}$$

Proof.

1. The property (12.69) results directly from (12.68) since

$$\mathring{x}(t,\omega) \overset{a.s.}{=} \Phi_\omega(t,t_0)\,\mathring{x}(t_0,\omega) \overset{a.s.}{=} [\Phi_\omega(t,t_0)\,\Phi_\omega(t_0,s)]\,\mathring{x}(s,\omega)$$

The inverse matrix $\Phi_\omega^{-1}(s,t_0)$ always exists since for $t = t_0$ one has

$$\det \Phi_\omega(t_0,t_0) = 1 \overset{a.s.}{=} \det \Phi_\omega(t_0,s) \det \Phi_\omega(s,t_0)$$

and therefore, $\det \Phi_\omega(t_0,s) \neq 0$ and

$$\det \Phi_\omega(s,t_0) = 1/\det \Phi_\omega(t_0,s) \neq 0$$

2. The relations (12.67) and (12.68) imply

$$\begin{aligned}
&\mathring{x}(t,\omega) - \mathring{x}(s,\omega) \\
&\quad \overset{a.s}{=} [\Phi_\omega(t,s) - I]\,\mathring{x}(s,\omega) \\
&\quad \overset{a.s.}{=} \int_{\tau=s}^{t} A(\tau)\mathring{x}(\tau,\omega)\,d\tau + \int_{\tau=s}^{t} \sum_{i=1}^{m} G_{1,i}(\tau)\mathring{x}(\tau,\omega)\,dW_{i,\tau}(\omega) \\
&\quad = \int_{\tau=s}^{t} A(\tau)\Phi_\omega(\tau,s)\,\mathring{x}(s,\omega)\,d\tau + \int_{\tau=s}^{t} \sum_{i=1}^{m} G_{1,i}(\tau)\Phi_\omega(\tau,s)\,\mathring{x}(s,\omega)\,dW_{i,\tau}(\omega) \\
&\quad = \left[\int_{\tau=s}^{t} A(\tau)\Phi_\omega(\tau,s)\,d\tau + \int_{\tau=s}^{t} \sum_{i=1}^{m} G_{1,i}(\tau)\Phi_\omega(\tau,s)\,dW_{i,\tau}(\omega)\right]\mathring{x}(s,\omega)
\end{aligned}$$

and since $\mathring{x}(s,\omega)$ is any n-vector, the relation (12.70) follows. Lemma is proven. □

Remark 12.5. *In general, an explicit expression for the transition matrix $\Phi_\omega(t,s)$ cannot be given. However, if $G_{1,i}(t) = 0$ for all $t \in [0,T]$ and $A(\tau) = A = \text{const}$ then, as follows from (12.70),*

$$\Phi_\omega(t,s) = I + \int_{\tau=s}^{t} A(\tau)\Phi_\omega(\tau,s)\,d\tau$$

and hence,

$$\frac{\partial}{\partial t}\Phi_\omega(t,s) = A(t)\Phi_\omega(t,s), \quad \Phi_\omega(t,t) = I$$

or, equivalently,

$$\boxed{\Phi_\omega(t,s) = e^{A(t-s)}} \tag{12.72}$$

12.4.2 General solution

12.4.2.1 Scalar case

Consider the stochastic linear differential equation

$$dx(t,\omega) = [a_0(t) + a_1(t)x(t,\omega)]\,dt + [g_0(t) + g_1(t)x(t,\omega)]\,dW_t(\omega) \tag{12.73}$$

whose homogeneous version is

$$\begin{aligned} &d\mathring{x}(t,\omega) = a_1(t)\mathring{x}(t,\omega)\,dt + g_1(t)\mathring{x}(t,\omega)\,dW(\omega) \\ &\mathring{x}(0,\omega) = \mathring{x}_0(\omega) \end{aligned} \tag{12.74}$$

According to (12.65) its solution is

$$\mathring{x}(t,\omega) = \mathring{x}_0(\omega)\exp\left\{\int_{s=0}^{t}\left[a_1(s) - \frac{1}{2}g_1^2(s)\right]ds + \int_{s=0}^{t} g_1(s)\,dW_s(\omega)\right\}$$

In terms of fundamental solution $\Phi_\omega(t,s)$ (which in this case coincides with $\mathring{x}(t,\omega)$ under the condition that $\mathring{x}_0(\omega) = 1$) it can be represented as

$$\begin{aligned} &\mathring{x}(t,\omega) = \Phi_\omega(t,0)\,\mathring{x}_0(\omega) \\ &\Phi_\omega(t,0) = \exp\left\{\int_{s=0}^{t}\left[a_1(s) - \frac{1}{2}g_1^2(s)\right]ds + \int_{s=0}^{t} g_1(s)\,dW_s(\omega)\right\} \end{aligned} \tag{12.75}$$

where the differential of $\Phi_\omega(t,0)$ is

$$d\Phi_\omega(t,0) = \Phi_\omega(t,0)\,[a_1(t)dt + g_1(t)W(\omega)], \quad \Phi_\omega(0,0) = I$$

Let us try to find the solution to (12.73), using the classical '*variation of parameters technique*', as

$$x(t,\omega) = \Phi_\omega(t,0)\,y(t,\omega)$$

so that the problem is to determine

$$y(t,\omega) = \Phi_\omega^{-1}(t,0)\,x(t,\omega), \quad y(0,\omega) = x(0,\omega)$$

where

$$\begin{aligned} &\Phi_\omega^{-1}(t,0) = \exp\left\{-\int_{s=0}^{t}\left[a_1(s) - \frac{1}{2}g_1^2(s)\right]ds - \int_{s=0}^{t} g_1(s)\,dW_s(\omega)\right\} = e^{z(t,\omega)} \\ &z(t,\omega) = -\int_{s=0}^{t}\left[a_1(s) - \frac{1}{2}g_1^2(s)\right]ds - \int_{s=0}^{t} g_1(s)\,dW_s(\omega) \\ &dz(t,\omega) = -\left[a_1(t) - \frac{1}{2}g_1^2(t)\right]dt - g_1(t)\,dW_t(\omega) \end{aligned}$$

with its differential (by the Itô formula)

$$
\begin{aligned}
d\Phi_{\omega}^{-1}(t,0) &= \left[\frac{\partial}{\partial z}e^{z(t,\omega)}\right]dz(t,\omega)+\frac{1}{2}g_1^2(t)\left[\frac{\partial^2}{\partial z^2}e^{z(t,\omega)}\right] \\
&= e^{z(t,\omega)}\left(\left[-a_1(t)+g_1^2(t)\right]dt-g_1(t)\,dW_t(\omega)\right) \\
&= \Phi_{\omega}^{-1}(t,0)\left(\left[-a_1(t)+g_1^2(t)\right]dt-g_1(t)\,dW_t(\omega)\right)
\end{aligned}
$$

and applying the Itô formula (in fact, (11.77)) we get

$$
\begin{aligned}
dy(t,\omega) &= d\Phi_{\omega}^{-1}(t,0)\,x(t,\omega)+\Phi_{\omega}^{-1}(t,0)\,dx(t,\omega) \\
&\quad -\left[g_0(t)+g_1(t)x(t,\omega)\right]\Phi_{\omega}^{-1}(t,0)\,g_1(t)dt \\
&= \Phi_{\omega}^{-1}(t,0)\left(\left[-a_1(t)+g_1^2(t)\right]dt-g_1(t)\,dW_t(\omega)\right)x(t,\omega) \\
&\quad +\Phi_{\omega}^{-1}(t,0)\left[a_0(t)+a_1(t)x(t,\omega)\right]dt \\
&\quad +\Phi_{\omega}^{-1}(t,0)\left[g_0(t)+g_1(t)x(t,\omega)\right]dW_t(\omega) \\
&\quad -\left[g_0(t)+g_1(t)x(t,\omega)\right]\Phi_{\omega}^{-1}(t,0)\,g_1(t)dt \\
&= \Phi_{\omega}^{-1}(t,0)\left[a_0(t)-g_0(t)g_1(t)\right]dt+\Phi_{\omega}^{-1}(t,0)\,g_0(t)dW_t(\omega)
\end{aligned}
\tag{12.76}
$$

Therefore

$$
\begin{aligned}
y(t,\omega) &= x(0,\omega)+\int_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)\left[a_0(s)-g_0(s)g_1(s)\right]ds \\
&\quad +\int_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)\,g_0(s)dW_s(\omega)\,ds
\end{aligned}
$$

which establishes the following result.

Theorem 12.9. *The solution $x(t,\omega)$ of the linear stochastic differential equation (12.73) is*

$$
\boxed{
\begin{aligned}
x(t,\omega) = \Phi_{\omega}(t,0)\Bigg[& x(0,\omega) \\
& +\int_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)\left[a_0(s)-g_0(s)g_1(s)\right]ds \\
& +\int_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)\,g_0(s)dW_s(\omega)\,ds\Bigg]
\end{aligned}
}
\tag{12.77}
$$

where $\Phi_{\omega}(t,0)$ is defined by (12.75).

12.4.2.2 Vector case

The following theorem covers the preceding linear vector system case and is similar to the scalar case given above. That's why we give it without the proof.

Theorem 12.10. *The solution of (12.66) with the fundamental matrix $\Phi_\omega(t,s)$, associated with the homogeneous linear stochastic equation (12.67), can be represented as*

$$\boxed{\begin{aligned} x(t,\omega) \overset{a.s.}{=} \Phi_\omega(t,0)\Bigg[& x_0(\omega) \\ & + \int_{s=0}^{t} \Phi_\omega^{-1}(s,0)\left[a_0(s) - \sum_{i=1}^{m} G_{1,i}(s) g_{0,i}(s)\right] ds \\ & + \int_{s=0}^{t} \Phi_\omega^{-1}(s,0) \sum_{i=1}^{m} g_{0,i}(s) dW_{i,s}(\omega)\Bigg] \end{aligned}} \tag{12.78}$$

where $\Phi_\omega(t,0)$ satisfies (12.71).

The next result is based on the direct use of formula (12.78) and concerns the differential equations for the first and the second moments of process (12.66) denoted by

$$\boxed{\begin{aligned} m(t) &:= \mathsf{E}\{x(t,\omega)\} \\ Q(t) &:= \mathsf{E}\{x(t,\omega)\, x^{\mathsf{T}}(t,\omega)\} \end{aligned}} \tag{12.79}$$

Theorem 12.11. **(on the first and second moments)** *The first two moments (12.79) of the solution $x(t,\omega)$ for (12.66)*

$$\boxed{\begin{aligned} dx(t,\omega) = {} & [a_0(t) + A(t)x(t,\omega)]\, dt \\ & + \sum_{i=1}^{m} \left[g_{0,i}(t) + G_{1,i}(t)x(t,\omega)\right] dW_{i,t}(\omega) \end{aligned}}$$

assuming that $\operatorname{tr}\{\mathsf{Q}(0)\} < \infty$, *are the unique solutions to the following initial value problems:*

(a)

$$\boxed{\begin{aligned} \dot m(t) &= A(t)m(t) + a_0(t) \\ m(0) &= \mathsf{E}\{x(0,\omega)\} \end{aligned}} \tag{12.80}$$

(b)

$$\boxed{\begin{aligned}\dot{Q}(t) &= A(t)Q(t) + Q(t)A^{\mathsf{T}}(t) + \sum_{i=1}^{m} G_{1,i}(t)Q(t)G_{1,i}^{\mathsf{T}}(t) \\ &\quad + a_0(t)m^{\mathsf{T}}(t) + m(t)a_0^{\mathsf{T}}(t) \\ &\quad + \sum_{i=1}^{m}\left[G_{1,i}(t)m(t)g_{0,i}^{\mathsf{T}}(t) + g_{0,i}(t)m^{\mathsf{T}}(t)G_{1,i}^{\mathsf{T}}(t) + g_{0,i}(t)g_{0,i}^{\mathsf{T}}(t)\right] \\ &\qquad Q(0) = E\{x(0,\omega)x^{\mathsf{T}}(0,\omega)\}\end{aligned}} \tag{12.81}$$

Proof. The assertion (*a*) can be verified directly taking the expected value of the integral form of (12.66). To provide the assertion (*b*) it is sufficient to use Itô's rule for the product, namely,

$$\begin{aligned}d\left[x(t,\omega)x^{\mathsf{T}}(t,\omega)\right] &= x(t,\omega)dx^{\mathsf{T}}(t,\omega) + [dx(t,\omega)]x^{\mathsf{T}}(t,\omega) \\ &\quad + \sum_{i=1}^{m}\left[g_{0,i}(t) + G_{1,i}(t)x(t,\omega)\right] \\ &\quad \times\left[g_{0,i}(t) + G_{1,i}(t)x(t,\omega)\right]^{\mathsf{T}} dt\end{aligned}$$

Substituting the integral forms for $d[x(t,\omega)x^{\mathsf{T}}(t,\omega)]$, $dx(t,\omega)$ and $dx^{\mathsf{T}}(t,\omega)$ into the equation above and taking the expected value leads to the integral equation equivalent (12.81). Theorem is proven. □

Corollary 12.3. ($a_0(t) = 0$, $G_{1,i}(t) = 0$ $(i,\ldots,m)$) *If in (12.66)*

$$a_0(t) = 0, \quad G_{1,i}(t) = 0\,(i,\ldots,m)$$

namely, if

$$dx(t,\omega) = A(t)x(t,\omega)dt + \sum_{i=1}^{m} g_{0,i}(t)dW_{i,t}(\omega) \tag{12.82}$$

then

$$x(t,\omega) \overset{a.s.}{=} \Phi_\omega(t,0)x_0(\omega) + \int_{s=0}^{t} \Phi_\omega^{-1}(s,0)\sum_{i=1}^{m} g_{0,i}(s)dW_{i,s}(\omega) \tag{12.83}$$

and

$$\begin{aligned}\dot{Q}(t) &= A(t)Q(t) + Q(t)A^{\mathsf{T}}(t) + \sum_{i=1}^{m} g_{0,i}(t)g_{0,i}^{\mathsf{T}}(t) \\ Q(0) &= E\{x(0,\omega)x^{\mathsf{T}}(0,\omega)\}\end{aligned} \tag{12.84}$$

Corollary 12.4. (a special case) *For the vector linear stochastic differential equation*

$$\boxed{\begin{array}{c} dx\,(t,\omega) = [A(t)x\,(t,\omega) + b(t)]\,dt + C\,(t)\,dW_t\,(\omega) \\ W_t(\omega) \text{ – is } m\text{-dimensional standard Wiener process} \\ C\,(t) \in \mathbb{R}^{n\times m} \end{array}} \tag{12.85}$$

its solution is given by

$$\boxed{\begin{array}{l} x\,(t,\omega) \overset{a.s.}{=} \Phi_\omega\,(t,0)\,x_0\,(\omega) \\ \qquad + \Phi_\omega\,(t,0)\left[\int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,C\,(s)\,dW_s\,(\omega) + \int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,b(s)ds\right] \\ \qquad \Phi_\omega\,(t,0) : \dfrac{d}{dt}\Phi_\omega\,(t,0) = A(t)\Phi_\omega\,(t,0)\,, \quad \Phi_\omega\,(0,0) = I \end{array}} \tag{12.86}$$

and its first two moments are the solutions to the following initial value problems:

$$\boxed{\dot{m}\,(t) = A(t)m\,(t) + b(t), \quad m\,(0) = \mathsf{E}\,\{x\,(0,\omega)\}} \tag{12.87}$$

and

$$\boxed{\begin{array}{l} \dot{Q}\,(t) = A(t)Q\,(t) + Q\,(t)\,A^{\intercal}(t) \\ \qquad + b(t)m^{\intercal}\,(t) + m\,(t)\,b^{\intercal}(t) + C\,(t)\,C^{\intercal}\,(t) \\ \quad Q\,(0) = \mathsf{E}\,\{x\,(0,\omega)\,x^{\intercal}\,(0,\omega)\} \end{array}} \tag{12.88}$$

Proof. Formula (12.86) may be checked by direct differentiation. Equation (12.87) follows directly from (12.86) taking into account that

$$\begin{aligned} m(t) &= \mathsf{E}\,\{x\,(t,\omega)\} \\ &= \mathsf{E}\left\{\Phi_\omega\,(t,0)\,x_0\,(\omega) + \Phi_\omega\,(t,0)\int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,b(s)ds\right\} \\ &= \Phi_\omega\,(t,0)\left[m\,(0) + \int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,b(s)ds\right] \\ \dot{m}(t) &= \frac{d}{dt}\Phi_\omega\,(t,0)\left[m\,(0) + \int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,b(s)ds\right] + b(t) \\ &= A(t)\Phi_\omega\,(t,0)\left[m\,(0) + \int\limits_{s=0}^{t} \Phi_\omega^{-1}\,(s,0)\,b(s)ds\right] + b(t) \\ &= A(t)m(t) + b(t) \end{aligned}$$

Equation (12.88) results from the following representation

$$
\begin{aligned}
Q(t) := \mathsf{E}\{x(t,\omega)x^{\intercal}(t,\omega)\} &= \mathsf{E}\left\{\Phi_{\omega}(t,0)x_0(\omega)x_0^{\intercal}(\omega)\Phi_{\omega}^{\intercal}(t,0)\right\} \\
&+ \mathsf{E}\left\{\Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\int\limits_{\tau=0}^{t}\Phi_{\omega}^{-1}(s,0)C(s)\,dW_s(\omega)\right. \\
&\left.\cdot\, dW_{\tau}^{\intercal}(\omega)C^{\intercal}(\tau)\left[\Phi_{\omega}^{-1}(\tau,0)\right]^{\intercal}\Phi_{\omega}^{\intercal}(t,0)\right\} \\
&+ \Phi_{\omega}(t,0)x_0(\omega)\left[\Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)b(s)ds\right]^{\intercal} \\
&+ \Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)b(s)ds\left[\Phi_{\omega}(t,0)x_0(\omega)\right]^{\intercal} \\
&= \Phi_{\omega}(t,0)Q(0)\Phi_{\omega}^{\intercal}(t,0) \\
&+ \Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)C(s)C^{\intercal}(s)\left[\Phi_{\omega}^{-1}(s,0)\right]^{\intercal}ds\Phi_{\omega}^{\intercal}(t,0) \\
&+ \Phi_{\omega}(t,0)x_0(\omega)\left[\Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)b(s)ds\right]^{\intercal} \\
&+ \Phi_{\omega}(t,0)\int\limits_{s=0}^{t}\Phi_{\omega}^{-1}(s,0)b(s)ds\left[\Phi_{\omega}(t,0)x_0(\omega)\right]^{\intercal}
\end{aligned}
$$

with the following direct calculation of $\dot{Q}(t)$ leading to (12.88). Corollary is proven. □

PART IV

Applications

This page intentionally left blank

13 Parametric Identification

Contents

This chapter contains four parts dealing with the principal identification processes of aspects of different dynamic and static stochastic models. The *first part* includes the introduction to the problem, its formulation, consideration of different particular models and the description of the least square method (LSM) derived in nonrecurrent and recurrent form. The *second part* discusses the convergence analysis of LSM. There it is shown that a direct application of LSM to the identification of autoregression processes with correlated noises leads to incorrect results. Some other versions dealing with 'whitening' or 'instrumental variables' (IV) are required. The analysis of IV-methodology is also presented. The information bounds of these methods characterizing the 'rate of estimation' are derived in the *third part*. They are based on the Cramér–Rao inequalities. The asymptotically optimal as well as some of their robust versions are considered in the *fourth part*.

13.1 Introduction

13.1.1 Parameters estimation as a component of identification theory

Modern *identification theory* (Ljung, 1987, 1999) basically deals with the problem of the efficient extraction of signal and some dynamic properties of systems based on available data or measurements.

Dynamic system identification is traditionally concerned with two issues:

- estimation of parameters based on direct and complete state space measurements;
- state space estimation (filtering or observation) of completely known nonlinear dynamics.

In this chapter we will consider the first option dealing with parameters identification. The next chapter will be concerned with state estimation processes.

Parameters identification for different classes of dynamic systems has been extensively studied during the last three decades. Basically, the class of dynamic systems whose dynamics depends linearly on the unknown parameters was considered, and external noise was assumed to be of a stochastic nature (see, for example, Ljung and Soderstrom (1983), Chen and Guo (1991)). In an earlier paper (Poznyak, 1980) the convergence properties of a matrix version of LSM (discrete time procedure) in the presence of stochastic noise and a nonlinear state-space coordinate transformation were studied. In Caines (1988) a comprehensive survey of different identification procedures is given including a family of observer-based parameter identifiers which exploit some parameter relations to improve an identification performance.

Remark 13.1. *A general feature of the publications mentioned above is that exact state space vector measurements are assumed to be available.*

13.1.2 Problem formulation

Let $(\Omega, \mathcal{F}, P)$ be a probability space, i.e.,

- Ω be a space of elementary events ω;
- $\mathcal{F}$ be a sigma-algebra (or a collection of all possible events generated by Ω);
- P be a probabilistic measure on Ω.

Consider the sequence $\{x_n\}_{n=0,1,2,\ldots}$ of random vectors $x_n = x_n(\omega) \in R^N$ defined on $(\Omega, \mathcal{F}, P)$ and related by the recursion

$$\boxed{\begin{array}{c} x_{n+1} = Ax_n + Bu_n + D\zeta_n \\ x_0 \text{ is a given random variable} \end{array}} \tag{13.1}$$

where

$x_n \in \mathbb{R}^N$ is the state random vector available (measurable) at time n;
$A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times K}$ and $D \in \mathbb{R}^{N\times M}$ are constant matrices,
$u_n \in \mathbb{R}^K$ is an input signal (stochastic or deterministic) available (measurable) at time n;
$\zeta_n \in \mathbb{R}^M$ is an external random perturbation (or noise) which is not available during the process.

The main problem dealing with the identification of the parameters in (13.1) may be formulated as follows:

Problem 13.1. *Based on the available information* $\{x_{t+1}, x_t, u_t\}_{t=0,\ldots,n}$ *construct estimates* $\left\{\hat{A}_t, \hat{B}_t\right\}_{t=0,\ldots,n}$ *and, maybe,* $\left\{\hat{D}_t\right\}_{t=0,\ldots,n}$ *if possible, which are*

(a) ***asymptotically consistent****, that is, in some probabilistic sense*

$$\begin{array}{l} \left\{\hat{A}_t, \hat{B}_t\right\}_{t=0,\ldots,n} \underset{n\to\infty}{\to} \{A, B\} \\ \left\{\hat{D}_t\right\}_{t=0,\ldots,n} \underset{n\to\infty}{\to} \{D\} \end{array} \tag{13.2}$$

(b) ***asymptotically (or locally) optimal****, which means that there do not exist some other estimates which are better (in some strict probabilistic sense) than these ones.*

13.2 Some models of dynamic processes

Here we will present several partial models which are commonly used in the mathematical and engineering literature and which can be represented in the form (13.1).

13.2.1 Autoregression (AR) model

The so-called scalar *autoregression* (AR) model is given by

$$\boxed{\begin{array}{c} y_{n+1} = a_0 y_n + a_1 y_{n-1} + \ldots + a_{L_a} y_{n-L_a} + \xi_n \\ y_0, y_{-1}, \ldots, y_{-L_a} \text{ are given} \end{array}} \tag{13.3}$$

where $\{\xi_n\}$ is the sequence of statistically independent random variables defined on $(\Omega, \mathcal{F}, P)$ with zero means and bounded second moments, i.e.,

$$\begin{aligned} &\mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\left\{\xi_n^2\right\} = \sigma_n^2 \\ &\mathsf{P}\left\{\xi_n \in \Xi, \xi_t \in \Xi'\right\} = \mathsf{P}\{\xi_n \in \Xi\}\,\mathsf{P}\left\{\xi_t \in \Xi'\right\} \quad \text{for any } n \neq t \\ &\mathsf{E}\{\eta\} := \int\limits_{\omega \in \Omega} \eta(\omega)\, dP\{\omega\} \text{ is the Lebesgue integral} \end{aligned} \tag{13.4}$$

It can be rewritten in the form (13.1) with

$$\begin{aligned} &x_n := \left(y_n, y_{n-1}, \ldots y_{n-L_a}\right)^{\mathsf{T}} \in \mathbb{R}^N, \quad N = L_a + 1 \\ &A = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \\ &B = 0, \quad D = I, \quad \zeta_n = \begin{pmatrix} \xi_n & 0 & \cdots & 0 \end{pmatrix}^{\mathsf{T}} \end{aligned} \tag{13.5}$$

13.2.2 Regression (R) model

The so-called scalar *regression* (R) model is given by

$$\boxed{y_{n+1} = b_0 v_n + b_1 v_{n-1} + \ldots + b_{L_b} v_{n-L_b} + \xi_n} \tag{13.6}$$

where $\{\xi_n\}$ is the sequence of statistically independent random variables defined on $(\Omega, \mathcal{F}, P)$ with zero means and bounded second moments, and $\{v_n\}$ is a sequence of measurable (available) scalar inputs which may be deterministic or stochastic either in nature.

In the last case the input sequence is supposed to be independent of the noise sequence $\{\xi_n\}$.

It can also be rewritten in the form (13.1) with

$$\begin{aligned} &x_n := y_n \in \mathbb{R}, \quad N = 1 \\ &A = 0, \quad B = \begin{bmatrix} b_0 & b_1 & \cdots & b_{L_b} \end{bmatrix}, \quad D = 1, \quad \zeta_n = \xi_n \\ &u_n := \left(v_n, v_{n-1}, \ldots v_{n-L_b}\right)^{\intercal} \in \mathbb{R}^K, \quad K = L_b + 1 \end{aligned} \tag{13.7}$$

13.2.3 Regression–autoregression (RAR) model

The so-called scalar *regression–autoregression* (RAR) model represents the combination of AR (13.3) and R (13.6) models and is given by

$$\boxed{\begin{aligned} y_{n+1} &= a_0 y_n + a_1 y_{n-1} + \ldots + a_{L_a} y_{n-L_a} \\ &\quad + b_0 v_n + b_1 v_{n-1} + \ldots + b_{L_b} v_{n-L_b} + \xi_n \\ &= \sum_{s=0}^{L_a} a_s y_{n-s} + \sum_{l=0}^{L_b} b_l v_{n-l} + \xi_n \\ &\qquad y_0, y_{-1}, \ldots, y_{-L_a} \text{ are given} \end{aligned}} \tag{13.8}$$

In the general matrix form (13.1) it can be also presented as follows:

$$\begin{aligned} &x_n := \left(y_n, y_{n-1}, \ldots y_{n-L_a}\right)^{\intercal} \in \mathbb{R}^N, \quad N = L_a + 1 \\ &u_n := \left(v_n, v_{n-1}, \ldots v_{n-L_b}\right)^{\intercal} \in \mathbb{R}^K, \quad K = L_b + 1 \\ &A = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad D = I, \quad \zeta_n = \begin{pmatrix} \xi_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &B = \begin{bmatrix} b_0 & b_1 & \cdots & \cdots & b_{L_b} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \end{aligned} \tag{13.9}$$

13.2.4 Autoregression–moving average (ARMA) model

The scalar *autoregression–moving average* (ARMA) model is given by

$$\boxed{\begin{aligned} y_{n+1} &= a_0 y_n + a_1 y_{n-1} + \ldots + a_{L_a} y_{n-L_a} \\ &\quad + d_0 \xi_n + \ldots d_{L_d} \xi_{n-L_d} = \sum_{s=0}^{L_a} a_s y_{n-s} + \sum_{l=0}^{L_d} d_l \xi_{n-l} \\ &\qquad y_0, y_{-1}, \ldots, y_{-L_a} \text{ are given} \end{aligned}} \tag{13.10}$$

where $\{\xi_n\}$ is the sequence of statistically independent random variables defined on $(\Omega, \mathcal{F}, P)$ with zero means and bounded second moments.

In the general matrix form (13.1) it can be presented as follows:

$$
\begin{aligned}
& x_n := \left(y_n, y_{n-1}, \dots y_{n-L_a}\right)^{\mathsf{T}} \in \mathbb{R}^N, \quad N = L_a + 1 \\
& \zeta_n := \left(\xi_n, \xi_{n-1}, \dots \xi_{n-L_d}\right)^{\mathsf{T}} \in \mathbb{R}^M, \quad M = L_d + 1 \\
& A = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad B = 0 \\
& D = \begin{bmatrix} d_0 & d_1 & \cdots & \cdots & d_{L_d} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \zeta_n = \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix}
\end{aligned}
\tag{13.11}
$$

Remark 13.2. *Notice that in ARMA models, written in the general matrix form (13.11), the generalized noise vector sequence $\{\zeta_n\}$ is not independent, but is correlated since for $s : 1 \le s \le L_d$ and $\sigma_n^2 \ne 0$ (for all $n = 0, 1, \dots$)*

$$
\begin{aligned}
\mathsf{E}\left\{\zeta_n \zeta_{n-s}^{\mathsf{T}}\right\} &= \mathsf{E}\left\{ \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix} \left(\xi_{n-s} \quad \xi_{n-1-s} \quad \cdots \quad \xi_{n-L_d-s}\right) \right\} \\
&= \begin{bmatrix} 0 & 0 & \cdots & 0 & \sigma_{n-s}^2 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \sigma_{n-s-1}^2 & 0 & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \sigma_{n-L_d}^2 \end{bmatrix} \ne 0
\end{aligned}
$$

In the uncorrelated case, there should be found

$$
\mathsf{E}\left\{\zeta_n \zeta_{n-s}^{\mathsf{T}}\right\} = 0 \quad \textit{if} \quad s \ne 0
$$

13.2.5 Regression–autoregression–moving average (RARMA or ARMAX) model

The so-called scalar *regression–autoregression–moving average* (RARMA or ARMAX) model represents the combination of ARMA (13.10) and R (13.6) models and is given by

$$
\boxed{\begin{aligned}
y_{n+1} &= a_0 y_n + a_1 y_{n-1} + \ldots + a_{L_a} y_{n-L_a} \\
&\quad + b_0 v_n + b_1 v_{n-1} + \ldots + b_{L_b} v_{n-L_b} \\
&\quad + d_0 \xi_n + \ldots d_{L_d} \xi_{n-L_d} \\
&= \sum_{s=0}^{L_a} a_s y_{n-s} + \sum_{k=0}^{L_b} b_k v_{n-k} + \sum_{l=0}^{L_d} d_l \xi_{n-l} \\
&\quad y_0, y_{-1}, \ldots, y_{-L_a} \text{ are given}
\end{aligned}} \tag{13.12}
$$

Remark 13.3. *Here the measurable inputs $\{v_n\}$ being independent of $\{\xi_n\}$ are interpreted as 'exogenous' inputs, which explains the appearance of the capital 'X' in the abbreviation ARMAX.*

In the general matrix form (13.1) it can be presented as follows:

$$
\begin{aligned}
& x_n := \left(y_n, y_{n-1}, \ldots y_{n-L_a}\right)^{\intercal} \in R^N, \quad N = L_a + 1 \\
& u_n := \left(v_n, v_{n-1}, \ldots v_{n-L_b}\right)^{\intercal} \in R^K, \quad K = L_b + 1 \\
& \zeta_n := \left(\xi_n, \xi_{n-1}, \ldots \xi_{n-L_d}\right)^{\intercal} \in R^M, \quad M = L_d + 1 \\
& A = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} b_0 & b_1 & \cdots & \cdots & b_{L_b} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \\
& D = \begin{bmatrix} d_0 & d_1 & \cdots & \cdots & d_{L_d} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \zeta_n = \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix}
\end{aligned} \tag{13.13}
$$

13.2.6 Nonlinear regression–autoregression–moving average (NRARMAX) model

The nonlinear *regression–autoregression–moving average* (NRARMA or NARMAX) model represents the following nonlinear modification of the ARMAX (13.12) model:

$$
\boxed{\begin{aligned}
y_{n+1} &= a_0 \varphi_0 (y_n) + a_1 \varphi_1 (y_{n-1}) + \ldots + a_{L_a} \varphi_{L_a} \left(y_{n-L_a}\right) \\
&\quad + b_0 v_n + b_1 v_{n-1} + \ldots + b_{L_b} v_{n-L_b} \\
&\quad + d_0 \xi_n + \ldots d_{L_d} \xi_{n-L_d} \\
&= \sum_{s=0}^{L_a} a_s \varphi_s (y_{n-s}) + \sum_{k=0}^{L_b} b_k v_{n-k} + \sum_{l=0}^{L_d} d_l \xi_{n-l} \\
&\quad y_0, y_{-1}, \ldots, y_{-L_a} \text{ are given}
\end{aligned}} \tag{13.14}
$$

Here in the right-hand side of the recursion (13.14) the nonlinear functions

$$\varphi_s : \mathbb{R} \to \mathbb{R}$$

participate. Usually they should be selected (if we construct a model) in such a way that the L_m*-stability* (the boundedness of $\mathsf{E}\left\{|x_n|^m\right\} < \infty$) would be guaranteed.

13.3 LSM estimating

13.3.1 LSM deriving

Rewrite the general matrix recursion (13.1)

$$x_{n+1} = Ax_n + Bu_n + D\zeta_n$$

in the *extended* form

$$\boxed{\begin{array}{c} x_{n+1} = Cz_n + D\zeta_n \\ C := \begin{bmatrix} A & B \end{bmatrix} \in \mathbb{R}^{N\times(N+K)}, \quad z_n := \begin{bmatrix} x_n \\ u_n \end{bmatrix} \end{array}} \tag{13.15}$$

Here we will treat the vector $z_n \in \mathbb{R}^{N+K}$ as the '*generalized measurable (available) input*' at time n.

Definition 13.1. *The matrix $C_n \in \mathbb{R}^{N\times(N+K)}$ is said to be the* ***LSM-estimate of the matrix*** *C in (13.15) at time n if*

$$\boxed{C_n := \arg \min_{C\in\mathbb{R}^{N\times(N+K)}} \sum_{t=0}^{n} \|x_{t+1} - Cz_t\|^2} \tag{13.16}$$

Theorem 13.1. *If there exists a time $n_0 \geq 0$ such that*

$$\sum_{t=0}^{n_0} z_t z_t^{\intercal} > 0 \tag{13.17}$$

then the LSM-estimate C_n of the matrix C in (13.15) for all $n \geq n_0$ is ***uniquely defined*** *and is* ***given*** *by*

$$\boxed{\begin{array}{c} C_n = V_n^{\intercal} Z_n^{-1} \\ V_n := \sum_{t=0}^{n} z_t x_{t+1}^{\intercal}, \quad Z_n := \sum_{t=0}^{n} z_t z_t^{\intercal} = Z_n^{\intercal} > 0 \end{array}} \tag{13.18}$$

Proof. Using the identities

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$
$$\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$$

one has

$$\begin{aligned}
J_n(C) &:= \sum_{t=0}^{n} \|x_{t+1} - Cz_t\|^2 = \sum_{t=0}^{n} \left(\|x_{t+1}\|^2 - 2x_{t+1}^{\top} Cz_t + \|Cz_t\|^2 \right) \\
&= \sum_{t=0}^{n} \|x_{t+1}\|^2 - 2\sum_{t=0}^{n} \operatorname{tr}\left(x_{t+1}^{\top} Cz_t \right) + \sum_{t=0}^{n} \operatorname{tr}\left((Cz_t)^{\top} Cz_t \right) \\
&= \sum_{t=0}^{n} \|x_{t+1}\|^2 - 2\sum_{t=0}^{n} \operatorname{tr}\left(Cz_t x_{t+1}^{\top} \right) + \sum_{t=0}^{n} \operatorname{tr}\left(z_t^{\top} C^{\top} Cz_t \right) \\
&= \sum_{t=0}^{n} \|x_{t+1}\|^2 - 2\operatorname{tr}\left(C \sum_{t=0}^{n} z_t x_{t+1}^{\top} \right) + \operatorname{tr}\left(C^{\top} C \sum_{t=0}^{n} z_t z_t^{\top} \right) \\
&= \sum_{t=0}^{n} \|x_{t+1}\|^2 - 2\operatorname{tr}(CV_n) + \operatorname{tr}\left(CZ_n C^{\top} \right)
\end{aligned}$$

To calculate $\min_{C \in \mathbb{R}^{N \times (N+K)}} J_n(C)$ let us use the formulas (see (16.47) in Poznyak (2008))

$$\frac{\partial}{\partial A} \operatorname{tr}(BAC) = B^{\top} C^{\top}, \quad \frac{\partial}{\partial A} \operatorname{tr}\left(ABA^{\top} \right) = AB^{\top} + AB$$

that gives

$$\frac{\partial}{\partial C} J_n(C) = -2V_n^{\top} + 2CZ_n = 0$$

and hence,

$$C_n = V_n^{\top} Z_n^{-1}$$

which is exactly (13.18). The uniqueness results from the uniqueness of the representation of the last formula. □

Corollary 13.1. *If $N = 1$, that is, $C = c^{\top}$ then*

$$C_n = c_n^{\top} = \arg \min_{C \in \mathbb{R}^{1 \times (1+K)}} \sum_{t=0}^{n} \left\| x_{t+1} - c^{\top} z_t \right\|^2 \tag{13.19}$$

and

$$c_n = \left(c_n^{\top} \right)^{\top} = Z_n^{-1} V_n \in \mathbb{R}^{(1+K)} \tag{13.20}$$

Corollary 13.2. *The LSM-estimate (13.18) can be derived using the so-called system of the '**normal equation**' obtained by the multiplication of the right-hand side by $z_n^{\top}$ each equation (13.15), that is,*

$$x_{n+1} = Cz_n + D\zeta_n : z_n^{\top}$$

$$x_n = Cz_{n-1} + D\zeta_{n-1} : z_{n-1}^{\intercal}$$
$$\ldots$$
$$x_{n_0+1} = Cz_{n_0} + D\zeta_{n_0} : z_{n_0}^{\intercal}$$

which after the summation gives

$$\sum_{t=n_0}^{n} x_{t+1} z_t^{\intercal} = C \sum_{t=n_0}^{n} z_t z_t^{\intercal} + D \sum_{t=n_0}^{n} \zeta_t z_t^{\intercal}$$

Then the LSM-estimate (13.18) can be defined as the matrix C_n satisfying the identity called the ***'normal form equation'****:*

$$\sum_{t=n_0}^{n} x_{t+1} z_t^{\intercal} = C_n \sum_{t=n_0}^{n} z_t z_t^{\intercal} \tag{13.21}$$

that leads to (13.18) since (13.21) is evidently equivalent to

$$V_n^{\intercal} = C_n Z_n \tag{13.22}$$

or, if $Z_n > 0$,

$$C_n = V_n^{\intercal} Z_n^{-1} \tag{13.23}$$

Also the following relations hold:

$$V_n^{\intercal} = CZ_n + D \sum_{t=n_0}^{n} \zeta_t z_t^{\intercal}$$

or

$$C = V_n^{\intercal} Z_n^{-1} + D \left[\sum_{t=n_0}^{n} \zeta_t z_t^{\intercal} \right] Z_n^{-1} \tag{13.24}$$

13.3.2 Recurrent matrix version of LSM

Lemma 13.1. *The LSM-estimate C_n (13.18) can be represented recursively by*

$$\boxed{\begin{gathered} C_{n+1} = C_n + (x_{n+2} - C_n z_{n+1})\, z_{n+1}^{\intercal} \Gamma_{n+1} \\ \Gamma_{n+1} = \Gamma_n - \frac{\Gamma_n z_{n+1} z_{n+1}^{\intercal} \Gamma_n}{1 + z_{n+1}^{\intercal} \Gamma_n z_{n+1}}, \quad n \geq n_0 \\ C_{n_0} = V_{n_0}^{\intercal} Z_{n_0}^{-1}, \quad \Gamma_{n_0} = \left(\sum_{t=0}^{n_0} z_t z_t^{\intercal} \right)^{-1} = Z_{n_0}^{-1} \end{gathered}} \tag{13.25}$$

Proof. Indeed, by definition (13.18)

$$
\begin{aligned}
C_{n+1} &= V_{n+1}^{\intercal} Z_{n+1}^{-1} = \left(V_n + z_{n+1} x_{n+2}^{\intercal}\right)^{\intercal} Z_{n+1}^{-1} \\
&= V_n^{\intercal} Z_{n+1}^{-1} + \left(z_{n+1} x_{n+2}^{\intercal}\right)^{\intercal} Z_{n+1}^{-1} \\
&= \left(V_n^{\intercal} Z_n^{-1}\right) Z_n Z_{n+1}^{-1} + x_{n+2} z_{n+1}^{\intercal} Z_{n+1}^{-1} \\
&= C_n Z_n Z_{n+1}^{-1} + x_{n+2} z_{n+1}^{\intercal} Z_{n+1}^{-1}
\end{aligned} \tag{13.26}
$$

Notice also that

$$
I = Z_{n+1} Z_{n+1}^{-1} = \left(Z_n + z_{n+1} z_{n+1}^{\intercal}\right) Z_{n+1}^{-1} = Z_n Z_{n+1}^{-1} + z_{n+1} z_{n+1}^{\intercal} Z_{n+1}^{-1}
$$

which implies

$$
Z_n Z_{n+1}^{-1} = I - z_{n+1} z_{n+1}^{\intercal} Z_{n+1}^{-1}
$$

Substitution of the last identity into (13.26) gives

$$
\begin{aligned}
C_{n+1} &= C_n Z_n Z_{n+1}^{-1} + x_{n+2} z_{n+1}^{\intercal} Z_{n+1}^{-1} = C_n \left(I - z_{n+1} z_{n+1}^{\intercal} Z_{n+1}^{-1}\right) \\
&\quad + x_{n+2} z_{n+1}^{\intercal} Z_{n+1}^{-1} = C_n + \left(x_{n+2} - C_n z_{n+1}\right) z_{n+1}^{\intercal} Z_{n+1}^{-1}
\end{aligned} \tag{13.27}
$$

Defining $\Gamma_n := Z_n^{-1}$ and using the *Sherman–Morrison* formula (see (2.8) in Poznyak (2008))

$$
\left(A + u v^{\intercal}\right)^{-1} = A^{-1} - \frac{A^{-1} u v^{\intercal} A^{-1}}{1 + v^{\intercal} A^{-1} u}, \quad v^{\intercal} A^{-1} u \neq -1 \tag{13.28}
$$

we get

$$
\begin{aligned}
\Gamma_{n+1} &= Z_{n+1}^{-1} = \left(Z_n + z_{n+1} z_{n+1}^{\intercal}\right)^{-1} \\
&= Z_n^{-1} - \frac{Z_n^{-1} z_{n+1} z_{n+1}^{\intercal} Z_n^{-1}}{1 + z_{n+1}^{\intercal} Z_n^{-1} z_{n+1}} = \Gamma_n - \frac{\Gamma_n z_{n+1} z_{n+1}^{\intercal} \Gamma_n}{1 + z_{n+1}^{\intercal} \Gamma_n z_{n+1}}
\end{aligned}
$$

that together with (13.27) leads to (13.25). □

Corollary 13.3. *For $N = 1$ the recurrent version of the LSM procedure (13.25) is as follows:*

$$
\boxed{
\begin{aligned}
c_{n+1} &= c_n + \Gamma_{n+1} z_{n+1} \left(x_{n+2} - c_n z_{n+1}\right) \\
\Gamma_{n+1} &= \Gamma_n - \frac{\Gamma_n z_{n+1} z_{n+1}^{\intercal} \Gamma_n}{1 + z_{n+1}^{\intercal} \Gamma_n z_{n+1}}, \quad n \geq n_0 \\
c_{n_0} &= Z_{n_0}^{-1} V_{n_0}, \quad \Gamma_{n_0} = \left(\sum_{t=0}^{n_0} z_t z_t^{\intercal}\right)^{-1} = Z_{n_0}^{-1}
\end{aligned}
} \tag{13.29}
$$

Corollary 13.4. *The algorithm (13.25) is asymptotically equivalent to the following one:*

$$\boxed{\begin{array}{c} c_{n+1} = c_n + \Gamma_{n+1} z_{n+1} \left(x_{n+2} - c_n z_{n+1} \right) \\ \Gamma_{n+1} = \Gamma_n - \dfrac{\Gamma_n z_{n+1} z_{n+1}^{\intercal} \Gamma_n}{1 + z_{n+1}^{\intercal} \Gamma_n z_{n+1}}, \quad n = 0, 1, \ldots \\ \Gamma_0 := \rho^{-1} I, \quad 0 < \rho \ll 1 \text{ (small enough)} \end{array}} \tag{13.30}$$

since for large enough n,

$$n\Gamma_n = \left(\frac{1}{n} \sum_{t=n_0}^{n} z_t z_t^{\intercal} \right)^{-1} \simeq \left(\frac{1}{n} \left[\sum_{t=1}^{n} z_t z_t^{\intercal} + \rho I \right] \right)^{-1}$$

13.4 Convergence analysis

13.4.1 Unbiased estimates

Definition 13.2. *A matrix function $C_n = C_n(x_1, z_1, \ldots, x_n, z_n, x_{n+1})$, depending on the available information and treated as an estimate of a matrix C at time n, is said to be*

- ***unbiased** at time n if*

$$\boxed{\mathsf{E}\{C_n\} = C} \tag{13.31}$$

- ***asymptotically unbiased** if*

$$\boxed{\mathsf{E}\{C_n\} - C \to 0} \tag{13.32}$$

  whereas $n \to \infty$.

Lemma 13.2. *The LSM-estimate (13.18) is unbiased if and only if*

$$\boxed{D\mathsf{E}\left\{ \left(\sum_{t=n_0}^{n} \zeta_t z_t^{\intercal} \right) Z_n^{-1} \right\} = 0} \tag{13.33}$$

Proof. It follows directly from (13.23) and (13.24) since

$$C = V_n^{\intercal} Z_n^{-1} + D \left[\sum_{t=n_0}^{n} \zeta_t z_t^{\intercal} \right] Z_n^{-1} = C_n + D \left[\sum_{t=n_0}^{n} \zeta_t z_t^{\intercal} \right] Z_n^{-1} \tag{13.34}$$

Lemma is proven. □

Corollary 13.5. *If for the regression **R-model** (13.7) the 'input vectors' z_t are statistically independent of the 'noise-vector' ζ_t which has a zero-mean value (i.e., $\mathsf{E}\{\zeta_t\} = 0$), then the corresponding LSM-estimate (13.18) is **unbiased**.*

Proof. Indeed,

$$D\mathsf{E}\left\{\left[\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}}\right] Z_n^{-1}\right\} = D\mathsf{E}\left\{\left[\sum_{t=n_0}^{n} \zeta_t \left(z_t^{\mathsf{T}} Z_n^{-1}\right)\right]\right\}$$
$$= D\sum_{t=n_0}^{n} \mathsf{E}\left\{\zeta_t \left(z_t^{\mathsf{T}} Z_n^{-1}\right)\right\} = D\sum_{t=n_0}^{n} \mathsf{E}\{\zeta_t\}\,\mathsf{E}\left\{z_t^{\mathsf{T}} Z_n^{-1}\right\} = 0$$

which proves the desired result. □

13.4.2 Asymptotic consistency

13.4.2.1 Strong LNL for dynamic models

Lemma 13.3. *If the dynamic model (13.15) is **stable**, that is, if*

$$\|A\| := \sqrt{\lambda_{\max}\left(A^{\mathsf{T}} A\right)} < 1 \tag{13.35}$$

and the 4th moments $\mathsf{E}\{\|\zeta_n\|^4\}$, $\mathsf{E}\{\|u_n\|^4\}$ and $\mathsf{E}\{\|x_0\|^4\}$ are uniformly (in n) bounded, namely,

$$\limsup_{n\to\infty} \mathsf{E}\{\|\zeta_n\|^4 + \|u_n\|^4\} + \mathsf{E}\{\|x_0\|^4\} < \infty \tag{13.36}$$

then for this model the strong large number law (see Corollary 8.10) holds:

$$\boxed{\begin{gathered}\left\|\frac{1}{n}\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}} - \frac{1}{n}\sum_{t=n_0}^{n} \mathsf{E}\left\{\zeta_t z_t^{\mathsf{T}}\right\}\right\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0 \\ \left\|\frac{1}{n} Z_n - \frac{1}{n}\mathsf{E}\{Z_n\}\right\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0\end{gathered}} \tag{13.37}$$

or equivalently,

$$\boxed{\begin{gathered}\frac{1}{n}\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}} \overset{a.s.}{=} \frac{1}{n}\sum_{t=n_0}^{n} \mathsf{E}\left\{\zeta_t z_t^{\mathsf{T}}\right\} + o_\omega(1) \\ \frac{1}{n} Z_n \overset{a.s.}{=} \frac{1}{n}\mathsf{E}\{Z_n\} + o_\omega(1) \\ o_\omega(1) \underset{n\to\infty}{\overset{a.s.}{\to}} 0\end{gathered}} \tag{13.38}$$

and, hence, by (13.34)

$$
\begin{aligned}
C - C_n &= D\left[\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}}\right] Z_n^{-1} = D\left[\frac{1}{n}\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}}\right]\left(\frac{1}{n} Z_n\right)^{-1} \\
&\overset{a.s.}{=} D\left[\frac{1}{n}\mathsf{E}\left\{\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}}\right\} + o_\omega(1)\right]\left(\frac{1}{n}\mathsf{E}\{Z_n\} + o_\omega(1)\right)^{-1} \\
&\overset{a.s.}{=} D\left[\frac{1}{n}\mathsf{E}\left\{\sum_{t=n_0}^{n} \zeta_t z_t^{\mathsf{T}}\right\}\right]\left(\frac{1}{n}\mathsf{E}\{Z_n\}\right)^{-1} + o_\omega(1)
\end{aligned}
\tag{13.39}
$$

Proof. It follows from the fact that for any stable model the existence of a moment for the input sequences implies the existence of the corresponding moment for the output sequence. So, by the Cauchy–Schwartz inequality,

$$
\begin{aligned}
\sigma^2_{\zeta_t z_t^{\mathsf{T}}} &:= \mathsf{E}\left\{\left\|\left(\zeta_t z_t^{\mathsf{T}}\right)^{\mathsf{T}}\left(\zeta_t z_t^{\mathsf{T}}\right)\right\|\right\} \le \sqrt{\mathsf{E}\left\{\left\|\left(\zeta_t z_t^{\mathsf{T}}\right)^{\mathsf{T}}\right\|^2\right\}}\sqrt{\mathsf{E}\left\{\left\|\left(\zeta_t z_t^{\mathsf{T}}\right)\right\|^2\right\}} \\
&= \mathsf{E}\left\{\left\|\left(\zeta_t z_t^{\mathsf{T}}\right)^{\mathsf{T}}\right\|^2\right\} \le \mathsf{E}\left\{\|\zeta_t\|^2\|z_t\|^2\right\} \le \sqrt{\mathsf{E}\left\{\|\zeta_t\|^4\right\}}\sqrt{\mathsf{E}\left\{\|z_t\|^4\right\}} \\
&\le \limsup_{n\to\infty}\sqrt{\mathsf{E}\left\{\|\zeta_t\|^4\right\}}\limsup_{n\to\infty}\sqrt{\mathsf{E}\left\{\|z_t\|^4\right\}} < \infty
\end{aligned}
$$

And hence, in this case,

$$
\begin{aligned}
&\sum_{n\in N^+}\frac{1}{n(n-1)}\sigma^2_{\zeta_t z_t^{\mathsf{T}}}\sqrt{\sum_{r=0}^{n-1}\sigma^2_{\zeta_r z_r^{\mathsf{T}}}} \le \text{Const}\sum_{n\in N^+}\frac{\sqrt{n}}{n(n-1)} \\
&= \text{Const}\sum_{n\in N^+}\frac{1}{\sqrt{n}(n-1)} = O\left(\sum_{n\in N^+}\frac{1}{n^{3/2}}\right) < \infty
\end{aligned}
$$

Therefore, Corollary 8.10 holds. The relation (13.39) results directly from (13.34). Lemma is proven. □

Definition 13.3. *An estimate* $C_n = C_n(x_0; x_1, z_1; \ldots; x_n, z_n, x_{n+1})$ *of the matrix* C *in (13.15) is said to be* ***asymptotic consistent****, or simply,* ***consistent***

- ***with probability one*** *(or, almost sure) if*

$$
C_n \underset{n\to\infty}{\overset{a.s.}{\to}} C \tag{13.40}
$$

  or equivalently, if

$$
\mathsf{P}\{\omega\in\Omega : C_n \to C \textit{ whereas } n\to\infty\} = 1
$$

- ***in probability*** *if for any* $\varepsilon > 0$

$$
\mathsf{P}\{\omega\in\Omega : \|C_n - C\| > \varepsilon\} \to 0 \quad \textit{whereas} \quad n\to\infty \tag{13.41}
$$

- ***in mean-square*** *if*

$$\mathsf{E}\left\{\|C_n - C\|^2\right\} \to 0 \quad \textit{whereas} \quad n \to \infty \tag{13.42}$$

The matrix norm above is intended to be as in (13.35), namely,

$$\|C_n - C\| := \sqrt{\lambda_{\max}\left((C_n - C)^{\mathsf{T}}(C_n - C)\right)} \tag{13.43}$$

Remark 13.4. *The mean-square consistency implies consistency in probability since by the Chebyshev inequality (4.10) it follows that*

$$\mathsf{P}\{\omega \in \Omega : \|C_n - C\| > \varepsilon\} \le \varepsilon^{-2}\mathsf{E}\left\{\|C_n - C\|^2\right\}$$

and, if the right-hand side of this inequality tends to zero, then the left-hand side tends to zero too.

13.4.2.2 Convergence analysis for static models

For the regression (static) **R-models** (13.7)

$$x_n = Bu_n + \xi_n \tag{13.44}$$

where the noise sequence is independent of the regression inputs $\{u_n\}$, the identification error (13.39) with $C = B$ becomes

$$\begin{aligned} C - C_n &\overset{a.s.}{=} D\left[\mathsf{E}\left\{\frac{1}{n}\sum_{t=n_0}^{n}\zeta_t z_t^{\mathsf{T}}\right\}\right]\left(\mathsf{E}\left\{\frac{1}{n}Z_n\right\}\right)^{-1} + o_\omega(1) \\ &= \left[\mathsf{E}\left\{\frac{1}{n}\sum_{t=n_0}^{n}\xi_t u_t^{\mathsf{T}}\right\}\right]\left(\mathsf{E}\left\{\frac{1}{n}\sum_{t=n_0}^{n}u_t u_t^{\mathsf{T}}\right\}\right)^{-1} + o_\omega(1) \\ &= o_\omega(1) \underset{n\to\infty}{\overset{a.s.}{\to}} 0 \end{aligned} \tag{13.45}$$

since

$$\mathsf{E}\left\{\frac{1}{n}\sum_{t=n_0}^{n}\zeta_t z_t^{\mathsf{T}}\right\} = \frac{1}{n}\sum_{t=n_0}^{n}\mathsf{E}\left\{\zeta_t z_t^{\mathsf{T}}\right\} = \frac{1}{n}\sum_{t=n_0}^{n}\underbrace{\mathsf{E}\{\zeta_t\}}_{0}\mathsf{E}\left\{z_t^{\mathsf{T}}\right\} = 0$$

We also have

$$\begin{aligned} (C - C_n)^{\mathsf{T}}(C - C_n) &= \left(Z_n^{-1}\right)^{\mathsf{T}}\left[\sum_{t=n_0}^{n} z_t\xi_t\right]\left[\sum_{t=n_0}^{n}\xi_t z_t^{\mathsf{T}}\right]Z_n^{-1} \\ &= \left(Z_n^{-1}\right)^{\mathsf{T}}\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\xi_t\xi_s\left(z_t z_s^{\mathsf{T}}\right)\right]Z_n^{-1} \end{aligned}$$

and hence,

$$\begin{aligned}
\mathsf{E}\left\{\|C_n - C\|^2\right\} &:= \mathsf{E}\left\{\lambda_{\max}\left((C - C_n)^{\intercal}(C - C_n)\right)\right\} \\
&\leq \mathsf{E}\left\{\operatorname{tr}\left((C - C_n)^{\intercal}(C - C_n)\right)\right\} \\
&= \mathsf{E}\left\{\operatorname{tr}\left(\left(Z_n^{-1}\right)^{\intercal}\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\xi_t\xi_s\left(z_t z_s^{\intercal}\right)\right]Z_n^{-1}\right)\right\} \\
&= \mathsf{E}\left\{\operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\xi_t\xi_s\left(z_t z_s^{\intercal}\right)\right]Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right)\right\} \\
&= \mathsf{E}\left\{\operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\xi_t\xi_s\left[\left(z_t z_s^{\intercal}\right)Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right]\right]\right)\right\} \\
&= \operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\mathsf{E}\left\{\xi_t\xi_s\right\}\mathsf{E}\left\{\left(z_t z_s^{\intercal}\right)Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right)
\end{aligned}$$

Since $\{\xi_n\}$ is a sequence of independent variables, then

$$\mathsf{E}\left\{\xi_t\xi_s\right\} = \sigma_t^2\delta_{t,s}$$

$$\delta_{t,s} := \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases} \text{ is the Kronecker symbol}$$

and hence,

$$\begin{aligned}
\mathsf{E}\left\{\|C_n - C\|^2\right\} &= \operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sum_{s=n_0}^{n}\sigma_t^2\delta_{t,s}\mathsf{E}\left\{\left(z_t z_s^{\intercal}\right)Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right) \\
&= \operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sigma_t^2\mathsf{E}\left\{\left(z_t z_t^{\intercal}\right)Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right)
\end{aligned}$$

For stationary noises, when $\sigma_t^2 = \sigma^2$, it follows that

$$\begin{aligned}
\mathsf{E}\left\{\|C_n - C\|^2\right\} &= \operatorname{tr}\left(\left[\sum_{t=n_0}^{n}\sigma_t^2\mathsf{E}\left\{\left(z_t z_t^{\intercal}\right)Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right) \\
&= \sigma^2\operatorname{tr}\left(\left[\mathsf{E}\left\{\underbrace{\left[\sum_{t=n_0}^{n}\left(z_t z_t^{\intercal}\right)\right]}_{Z_n}Z_n^{-1}\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right) \\
&= \sigma^2\operatorname{tr}\left(\left[\mathsf{E}\left\{\left(Z_n^{-1}\right)^{\intercal}\right\}\right]\right) = \sigma^2\operatorname{tr}\left(\mathsf{E}\left\{Z_n^{-1}\right\}\right) \\
&= \frac{\sigma^2}{n}\operatorname{tr}\left(\left[\mathsf{E}\left\{\left(\frac{1}{n}Z_n\right)^{-1}\right\}\right]\right)
\end{aligned} \tag{13.46}$$

The considerations above lead to the following statement.

Claim 13.1. *For the **R-model** (13.44) with independent noise sequence $\{\xi_n\}$ the LSM-estimates (13.25) are*

- ***consistent with probability one**, which results from (13.45) if the regression inputs u_n as well as the noise ξ_n have 4th bounded moments;*
- ***consistent in the mean-square sense**, which results from (13.46) if the so-called **persistent excitation condition** holds, namely, if*

$$\boxed{\limsup_{n\to\infty} \operatorname{tr}\left(\mathsf{E}\left\{\left(\frac{1}{n}Z_n\right)^{-1}\right\}\right) < \infty} \tag{13.47}$$

which is equivalent to the following inequality:

$$\boxed{\liminf_{n\to\infty} \lambda_{\min}\left(\frac{1}{n}\mathsf{E}\{Z_n\}\right) > 0} \tag{13.48}$$

since

$$\begin{aligned}
\mathsf{E}\left\{\|C_n - C\|^2\right\} &= \frac{\sigma^2}{n}\operatorname{tr}\left(\left[\mathsf{E}\left\{\left(\frac{1}{n}Z_n\right)^{-1}\right\}\right]\right) \\
&\le \frac{\sigma^2}{n}N\mathsf{E}\left\{\lambda_{\max}\left(\frac{1}{n}Z_n\right)^{-1}\right\} \le \frac{\sigma^2}{n}N\mathsf{E}\left\{\lambda_{\min}^{-1}\left(\frac{1}{n}Z_n\right)\right\} \\
&= \frac{\sigma^2}{n}N\mathsf{E}\left\{\lambda_{\min}^{-1}\left(\frac{1}{n}\mathsf{E}\{Z_n\} + o_\omega(1)\right)\right\} \\
&\simeq \frac{\sigma^2}{n}N\left[\lambda_{\min}^{-1}\left(\frac{1}{n}\mathsf{E}\{Z_n\} + \mathsf{E}\{o_\omega(1)\}\right)\right] \\
&\simeq \frac{\sigma^2}{n}N\left[\liminf_{n\to\infty}\lambda_{\min}\left(\frac{1}{n}\mathsf{E}\{Z_n\}\right) + \mathsf{E}\{o_\omega(1)\}\right]^{-1} \underset{n\to\infty}{\to} 0
\end{aligned}$$

13.4.2.3 Why LSM does not work for dynamic models with correlated noises

Consider, first, the ARMA model (13.10) given by

$$x_{n+1} = Ax_n + D\zeta_n \tag{13.49}$$

where

$$x_n := \begin{pmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{n-L_a} \end{pmatrix} \in \mathbb{R}^N, \quad A = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\zeta_n = \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix}, \quad D = \begin{bmatrix} d_0 & d_1 & \cdots & \cdots & d_{L_d} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

The LSM-formula (13.18) becomes as follows:

$$\begin{aligned} & C_n = V_n^{\intercal} Z_n^{-1} \\ & V_n := \sum_{t=0}^{n} x_t x_{t+1}^{\intercal}, \quad Z_n := \sum_{t=0}^{n} x_t x_t^{\intercal} = Z_n^{\intercal} > 0 \end{aligned} \tag{13.50}$$

and the corresponding estimation error (13.45) is

$$\begin{aligned} C - C_n & \overset{a.s.}{=} \left[\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ \zeta_t x_t^{\intercal} \right\} \right] \left(\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t x_t^{\intercal} \right\} \right)^{-1} + o_{\omega}(1) \\ & = \frac{1}{n} \sum_{t=n_0}^{n} \begin{bmatrix} \mathsf{E}\{\xi_t y_t\} & \mathsf{E}\{\xi_t y_{t-1}\} & \cdots & \mathsf{E}\left\{\xi_t y_{t-L_a}\right\} \\ \mathsf{E}\{\xi_{t-1} y_t\} & \mathsf{E}\{\xi_{t-1} y_{t-1}\} & \cdots & \mathsf{E}\left\{\xi_{t-1} y_{t-L_a}\right\} \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{E}\left\{\xi_{t-L_d} y_t\right\} & \mathsf{E}\left\{\xi_{t-L_d} y_{t-1}\right\} & \cdots & \mathsf{E}\left\{\xi_{t-L_d} y_{t-L_a}\right\} \end{bmatrix} \cdot \\ & \quad \cdot \left(\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t x_t^{\intercal} \right\} \right)^{-1} + o_{\omega}(1) \\ & = \frac{1}{n} \sum_{t=n_0}^{n} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \mathsf{E}\{\xi_{t-1} y_t\} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{E}\left\{\xi_{t-L_d} y_t\right\} & \mathsf{E}\left\{\xi_{t-L_d} y_{t-1}\right\} & \cdots & \mathsf{E}\left\{\xi_{t-L_d} y_{t-L_a}\right\} \end{bmatrix} \cdot \\ & \quad \cdot \left(\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t x_t^{\intercal} \right\} \right)^{-1} + o_{\omega}(1) \overset{a.s.}{\nrightarrow} 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

Claim 13.2. *For dynamic models (in particular, for ARMA models) the LS method* ***does not work properly****; namely, it leads to biased estimates!*

Example 13.1. *Consider a simplest stable ARMA model with 1-step correlated noise given by*

$$\begin{aligned} & y_{n+1} = a y_n + \xi_n + d\xi_{n-1}, \quad y_0 \text{ is given} \\ & |a| < 1, \quad d \in \mathbb{R}, \quad \mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\left\{\xi_n^2\right\} = \sigma^2 > 0 \end{aligned} \tag{13.51}$$

Then the LS estimate (13.50) is

$$a_n = \left[\frac{1}{n}\sum_{t=1}^{n} y_t y_{t+1}\right]\left[\frac{1}{n}\sum_{t=1}^{n} y_t^2\right]^{-1}$$

and under the conditions of LNL it becomes

$$a_n \overset{a.s.}{=} \frac{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\{y_t y_{t+1}\}}{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\left\{y_t^2\right\}} + o_\omega(1)$$

or, equivalently,

$$a_n \overset{a.s.}{=} \frac{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\{y_t (a y_t + \xi_t + d\xi_{t-1})\}}{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\left\{y_t^2\right\}} + o_\omega(1)$$

$$= a\mathsf{E}\frac{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\{y_t (\xi_t + d\xi_{t-1})\}}{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\left\{y_t^2\right\}} + o_\omega(1) = a + d\frac{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\{y_t \xi_{t-1}\}}{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\left\{y_t^2\right\}} + o_\omega(1)$$

So, the corresponding identification error becomes

$$a_n - a \overset{a.s.}{=} d\frac{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\{y_t \xi_{t-1}\}}{\frac{1}{n}\sum\limits_{t=1}^{n} \mathsf{E}\left\{y_t^2\right\}} + o_\omega(1)$$

For stable models with $|a| < 1$ there exist limits

$$\lim_{n\to\infty} \mathsf{E}\{y_n \xi_{n-1}\} \quad \textit{and} \quad \lim_{n\to\infty} \mathsf{E}\left\{y_n^2\right\}$$

and hence, by the Kronecker lemma

$$a_n - a \overset{a.s.}{=} d\frac{\lim\limits_{n\to\infty} \mathsf{E}\{y_n \xi_{n-1}\}}{\lim\limits_{n\to\infty} \mathsf{E}\left\{y_n^2\right\}} + o_\omega(1) \tag{13.52}$$

Let us calculate these limits. From (13.51) it follows

$$\mathsf{E}\{y_{n+1}\xi_n\} = a\mathsf{E}\{y_n\xi_n\} + \mathsf{E}\left\{\xi_n^2\right\} + d\mathsf{E}\{\xi_{n-1}\xi_n\} = \sigma^2 \tag{13.53}$$

$$\begin{aligned}
\mathsf{E}\left\{y_{n+1}^2\right\} &= a^2\mathsf{E}\left\{y_n^2\right\} + \mathsf{E}\left\{\xi_n^2\right\} + d^2\mathsf{E}\left\{\xi_{n-1}^2\right\} \\
&\quad + 2a\mathsf{E}\{y_n\xi_n\} + 2ad\mathsf{E}\{y_n\xi_{n-1}\} + 2d\mathsf{E}\{\xi_{n-1}\xi_n\} \\
&= a^2\mathsf{E}\left\{y_n^2\right\} + \left(1 + d^2\right)\sigma^2 + 2ad\mathsf{E}\{y_n\xi_{n-1}\} \\
&= a^2\mathsf{E}\left\{y_n^2\right\} + \left(1 + d^2\right)\sigma^2 + 2ad\sigma^2
\end{aligned} \tag{13.54}$$

Since, for the stable linear recursion

$$z_{n+1} = \bar{a}z_n + c, \quad |\bar{a}| < 1$$

we have

$$\begin{aligned} z_{n+1} &= \bar{a}z_n + c = \bar{a}\left(\bar{a}z_{n-1} + c\right) + c \\ &= \bar{a}^2 z_{n-1} + c + \bar{a}c = \ldots = \bar{a}^n z_1 + c + \bar{a}c + \bar{a}^2 c + \ldots + \bar{a}^n c \\ &= \bar{a}^n z_1 + c\left(\frac{1-\bar{a}^{n+1}}{1-\bar{a}}\right) \underset{n\to\infty}{\rightarrow} \frac{c}{1-\bar{a}} \end{aligned}$$

then, for (13.54), we get

$$\begin{aligned} &\mathsf{E}\left\{y_n^2\right\} \to \frac{\left(1+d^2\right)+2ad}{1-a^2}\sigma^2 \\ &= \frac{\left(1-a^2\right)+\left(a^2+2ad+d^2\right)}{1-a^2}\sigma^2 = \left[1+\frac{(a+d)^2}{1-a^2}\right]\sigma^2 \end{aligned} \tag{13.55}$$

Substitution the obtained limits (13.53) and (13.55) in to (13.52) implies

$$\boxed{\begin{aligned} a_n - a &\overset{a.s.}{=} d\frac{\sigma^2}{\left[1+\frac{(a+d)^2}{1-a^2}\right]\sigma^2} + o_\omega(1) \\ &= d\frac{1}{1+\frac{(a+d)^2}{1-a^2}} + o_\omega(1) \underset{n\to\infty}{\overset{a.s.}{\rightarrow}} d\frac{1}{1+\frac{(a+d)^2}{1-a^2}} \end{aligned}}$$

The derivative calculation of the limit value with respect to d gives

$$\left(d\frac{1}{1+\frac{(a+d)^2}{1-a^2}}\right)' = \left(a^2-1\right)\frac{d^2-1}{\left(d^2+2ad+1\right)^2}$$

So, the extremal points are $d = \pm 1$ and hence

$$\left(d\frac{1}{1+\frac{(a+d)^2}{1-a^2}}\right)_{d=1} = \frac{1}{2} - \frac{1}{2}a, \quad \left(d\frac{1}{1+\frac{(a+d)^2}{1-a^2}}\right)_{d=-1} = -\frac{1}{2}a - \frac{1}{2}$$

These relations imply the following conclusion: the maximum bias of the LSM estimate is

$$\max_d \lim_{n\to\infty} |a_n - a| = \frac{1}{2}\max\{|1-a|\,;\,|1+a|\}$$

The illustrative graphics ($x := d$, $y := |a_n - a|$ for $a = 0.5$) are shown in Fig. 13.1

Conclusion 13.1. ***Be careful: the LS method does not work*** *for identification of parameters of dynamic models with correlated noises!*

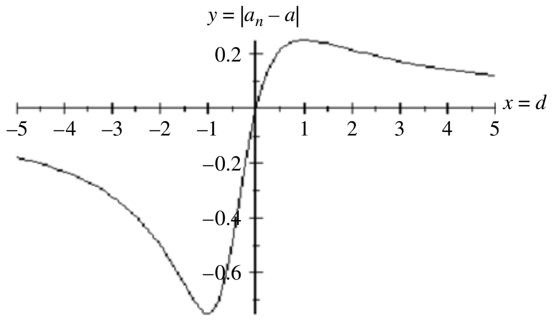

Fig. 13.1. The bias dependence on the correlation coefficient d.

The next section presents the technique (see, for example, Soderstrom and Stoica (1980), Poznyak and Tsypkin (1989) and Kaz'min and Poznyak (1992)) which corrects the LSM in this situation.

13.4.2.4 Instrumental variables method as a corrected version of LS method

Consider again the ARMA model (13.10) given by

$$x_{n+1} = Ax_n + D\zeta_n$$

with the noise component $D\zeta_n$ as

$$D = \begin{bmatrix} & d_1 & \cdots & \cdots & d_{L_d} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \zeta_n = \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix}$$

$$D\zeta_n = \begin{pmatrix} \sum_{s=0}^{L_d} d_s \xi_{n-s} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Define the new auxiliary sequence $\{v_n\}$ of random variables.

Definition 13.4. *We call the sequence $\{v_n\}$ of random vectors $v_n \in \mathbb{R}^N$* ***a sequence of instrumental variables (IV)*** *if it is obtained as the output of a stable forming filter $H(q)$ with the input sequence $\{x_n\}$, that is,*

$$v_n = H(q)\, x_n \tag{13.56}$$

where

$$\begin{aligned}
&H(q) = \frac{P(q)}{Q(q)} \\
&P(q) := \sum_{s=0}^{k_P} f_s q^s, \quad Q(q) := \sum_{s=0}^{k_Q} g_s q^s \\
&Q(q) \text{ is stable polynomial} \\
&q : q z_n = z_{n-1} \text{ is the delay operator}
\end{aligned} \tag{13.57}$$

Construct the estimating procedure, analogously to (13.50), as follows:

$$\begin{aligned}
&C_n = A_n = V_n^{\mathsf{T}} Z_n^{-1} \\
&V_n := \sum_{t=0}^{n} v_t x_{t+1}^{\mathsf{T}}, \quad Z_n := \sum_{t=0}^{n} x_t v_t^{\mathsf{T}}
\end{aligned} \tag{13.58}$$

with the instrumental variable v_n formed as

$$v_n = x_{n-L_d} \tag{13.59}$$

that corresponds to the forming filter

$$H(q) = q^{L_d} \tag{13.60}$$

Lemma 13.4. *The instrumental variable v_n is uncorrelated with the noise ζ_n.*

Proof. Indeed, since the components ξ_{n-s} $(s = 0, \ldots, L_d)$ of the vector ζ_n are independent random variables it follows that

$$\mathsf{E}\left\{\zeta_n v_n^{\mathsf{T}}\right\} = \mathsf{E}\left\{\zeta_n x_{n-L_d}^{\mathsf{T}}\right\} = \begin{bmatrix} \mathsf{E}\left\{\xi_n x_{n-L_d}^{\mathsf{T}}\right\} \\ \vdots \\ \mathsf{E}\left\{\xi_{n-L_d} x_{n-L_d}^{\mathsf{T}}\right\} \end{bmatrix} = 0 \qquad \square$$

For the instrumental variable estimates (13.58) the identification error is as follows:

$$\begin{aligned}
A - A_n &\overset{a.s.}{=} A - \left[\sum_{t=n_0}^{n} x_{t+1} v_t^{\mathsf{T}}\right]\left(\sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}}\right)^{-1} \\
&= A - \left[\sum_{t=n_0}^{n} (A x_t + \zeta_t) v_t^{\mathsf{T}}\right]\left(\sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}}\right)^{-1} \\
&= A - A\left[\sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}}\right]\left(\sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}}\right)^{-1} + \left[\sum_{t=n_0}^{n} \zeta_t v_t^{\mathsf{T}}\right]\left(\sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}}\right)^{-1}
\end{aligned}$$

$$= \left[\frac{1}{n} \sum_{t=n_0}^{n} \zeta_t v_t^{\mathsf{T}} \right] \left(\frac{1}{n} \sum_{t=n_0}^{n} x_t v_t^{\mathsf{T}} \right)^{-1}$$

$$= \left[\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ \zeta_t v_t^{\mathsf{T}} \right\} \right] \left(\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t v_t^{\mathsf{T}} \right\} \right)^{-1} + o_{\omega}(1) \tag{13.61}$$

which implies the following proposition.

Proposition 13.1. *The method of IV variables is consistent with probability one if and only if*

1. *the analog of the '**persistent excitation**' condition is fulfilled, i.e.,*

$$\boxed{\liminf_{n\to\infty} \left| \det \left(\frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t v_t^{\mathsf{T}} \right\} \right) \right| > 0} \tag{13.62}$$

(for example, if no signals, that is, $x_t \equiv 0$, the matrix $\sum_{t=n_0}^{n} \mathsf{E}\left\{ x_t v_t^{\mathsf{T}} \right\}$ is not invertible);

2. *the instrumental variables are **asymptotically uncorrelated with the noise**, i.e.,*

$$\boxed{\lim_{n\to\infty} \frac{1}{n} \sum_{t=n_0}^{n} \mathsf{E}\left\{ \zeta_t v_t^{\mathsf{T}} \right\} = 0} \tag{13.63}$$

Proof. It follows directly from (13.61). □

Example 13.2. *Consider the same example as in (13.51), namely,*

$$y_{n+1} = a y_n + \xi_n + d\xi_{n-1}, \quad y_0 \text{ is given}$$
$$|a| < 1, \quad d \in R, \quad \mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\left\{\xi_n^2\right\} = \sigma^2 > 0$$

but with the parameter estimations obtained by the IV method:

$$a_n = \frac{\sum_{t=1}^{n} y_{t-1} y_{t+1}}{\sum_{t=1}^{n} y_t y_{t-1}} = \frac{\frac{1}{n} \sum_{t=1}^{n} y_{t-1} y_{t+1}}{\frac{1}{n} \sum_{t=1}^{n} y_t y_{t-1}}$$

Obviously

$$a_n = \frac{\frac{1}{n} \sum_{t=1}^{n} y_{t-1} \left(a y_t + \xi_t + d\xi_{t-1} \right)}{\frac{1}{n} \sum_{t=1}^{n} y_{t-1} y_t}$$

$$= a + \frac{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}\xi_t}{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}y_t} + d\frac{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}\xi_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}y_t}$$

$$= a + \frac{\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}\xi_t\}}{\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}y_t\}} + d\frac{\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}\xi_{t-1}\}}{\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}y_t\}} + o_\omega(1) \tag{13.64}$$

Since

$$\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}y_t\} = \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}(ay_{t-1} + \xi_{t-1} + d\xi_{t-2})\}$$
$$= a\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\left\{y_{t-1}^2\right\} + d\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}\xi_{t-2}\} = a\frac{\sigma_\xi^2}{1-a^2} + d^2\sigma_\xi^2 + o(1)$$

then if

$$\left(\frac{a}{1-a^2} + d^2\right)\sigma_\xi^2 \neq 0$$

it follows that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}y_t\} \neq 0$$

But we also have

$$\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}\xi_{t-1}\} = \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\{y_{t-1}\xi_t\} = 0$$

Therefore, by (13.64), it follows that

$$a_n = a + o_\omega(1) \underset{n\to\infty}{\overset{a.s.}{\to}} a$$

Corollary 13.6. *This exactly means that the **corrected version of LSM-estimating with use of the instrumental variables (as the delayed state variables) guarantees the consistency property with probability one**.*

13.4.2.5 'Whitening' process

The general dynamic NARMAX model (13.14)

$$x_{n+1} = Cz_n + D\zeta_n$$

can be represented also as follows:

$$x_{n+1} = Cz_n + D\begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-L_d} \end{pmatrix} = Cz_n + \begin{pmatrix} d_0 \\ d_1 q \\ \vdots \\ d_{L_d} q^{L_d} \end{pmatrix} \xi_n$$

and for the scalar state-case it becomes

$$\boxed{\begin{aligned} x_{n+1} &= c^{\intercal} z_n + H(q)\xi_n \\ H(q) &= \sum_{s=0}^{L_d} d_s q^s \end{aligned}} \tag{13.65}$$

Supposing that the polynomial $H(q)$ is Hurwitz (stable), we may apply to both sides of (13.65) the transformation $H^{-1}(q)$ that implies

$$\tilde{x}_{n+1} = c^{\intercal} \tilde{z}_n + \xi_n + o_\omega(1) \tag{13.66}$$

where the sequences $\{\tilde{x}_n\}$ and $\{\tilde{z}_n\}$ are defined as

$$\boxed{\tilde{x}_n := H^{-1}(q)x_n, \quad \tilde{z}_n = H^{-1}(q)z_n} \tag{13.67}$$

or, equivalently,

$$H(q)\tilde{x}_n := x_n, \quad H(q)\tilde{z}_n = H^{-1}(q)z_n$$

This means that $\{\tilde{x}_n\}$ and $\{\tilde{z}_n\}$ are the outputs of the following regression models

$$\boxed{\begin{aligned} d_0\tilde{x}_n + d_1\tilde{x}_{n-1} + \ldots + d_{L_d}\tilde{x}_{n-L_d} &= x_n \\ d_0\tilde{z}_n + d_1\tilde{z}_{n-1} + \ldots + d_{L_d}\tilde{z}_{n-L_d} &= z_n \end{aligned}} \tag{13.68}$$

(with zero's initial conditions). Evidently $\tilde{z}_n$ and ξ_n are non-correlated since we have asymptotically a standard '*white noise*' case, and hence, the classical LSM-algorithm (13.25) can be directly applied to (13.66).

Summary 13.1. *To identify the parameters c in the scalar state-variable NARMAX model (13.65) one might apply the following procedure, called* ***LSM with whitening****, consisting of two parallel processes:*

- *Form the auxiliary sequences $\{\tilde{x}_n\}$ and $\{\tilde{z}_n\}$ using (13.68);*
- *Apply the classical LSM procedure (13.25)*

$$\boxed{\begin{gathered} C_n = \tilde{V}_n^{\intercal} \tilde{Z}_n^{-1} \\ \tilde{V}_n := \sum_{t=0}^{n} \tilde{x}_t \tilde{x}_{t+1}^{\intercal}, \quad \tilde{Z}_n := \sum_{t=0}^{n} \tilde{x}_t \tilde{x}_t^{\intercal} = \tilde{Z}_n^{\intercal} > 0 \end{gathered}} \tag{13.69}$$

or, its recurrent version

$$\boxed{\begin{aligned} C_{n+1} &= C_n + (\tilde{x}_{n+2} - C_n\tilde{z}_{n+1})\,\tilde{z}_{n+1}^{\intercal}\tilde{\Gamma}_{n+1} \\ \tilde{\Gamma}_{n+1} &= \tilde{\Gamma}_n - \frac{\tilde{\Gamma}_n\tilde{z}_{n+1}\tilde{z}_{n+1}^{\intercal}\tilde{\Gamma}_n}{1+\tilde{z}_{n+1}^{\intercal}\Gamma_n\tilde{z}_{n+1}}, \quad n \geq n_0 \\ C_{n_0} &= \tilde{V}_{n_0}^{\intercal}\tilde{Z}_{n_0}^{-1}, \quad \Gamma_{n_0} = \left(\sum_{t=0}^{n_0}\tilde{z}_t\tilde{z}_t^{\intercal}\right)^{-1} = Z_{n_0}^{-1} \end{aligned}} \tag{13.70}$$

13.5 Information bounds for identification methods

Here we will present the approach to the analysis of the rates of convergence for the parameter estimating procedures and pay especial attention to the question:
'*Which identification algorithm is better and in which situations?*'

13.5.1 Cauchy–Schwartz inequality for stochastic matrices

Here we present the relation which serves as the basic instrument for deriving the so-called *information bounds* characterizing the limiting possibilities of an estimating procedure under a given information situation.

13.5.1.1 Stochastic Hölder and Cauchy–Schwartz inequalities

We start with the *stochastic Hölder inequality 4.4* which says that for any positive p and q such that

$$p > 1, \quad q > 1, \quad p^{-1} + q^{-1} = 1 \tag{13.71}$$

and any scalar random variables ξ, η defined on $(\Omega, \mathcal{F}, P)$ and having p and q finite absolute moments, i.e.,

$$\mathsf{E}\left\{|\xi|^p\right\} < \infty, \quad \mathsf{E}\left\{|\eta|^q\right\} < \infty \tag{13.72}$$

the following inequality holds:

$$\boxed{\mathsf{E}\{|\xi\eta|\} \leq \left(\mathsf{E}\left\{|\xi|^p\right\}\right)^{1/p}\left(\mathsf{E}\left\{|\eta|^q\right\}\right)^{1/q}} \tag{13.73}$$

The stochastic Cauchy–Schwartz inequality (4.16) results from (13.73) letting $p = q = 2$, which leads to

$$\boxed{(\mathsf{E}\{\xi\eta\})^2 \leq (\mathsf{E}\{|\xi\eta|\})^2 \leq \mathsf{E}\left\{|\xi|^2\right\}\mathsf{E}\left\{|\eta|^2\right\}} \tag{13.74}$$

and if additionally $E\left\{|\eta|^2\right\} > 0$, then

$$\boxed{\begin{aligned} \mathsf{E}\left\{|\xi|^2\right\} &\geq \mathsf{E}\left\{|\xi\eta|\right\}\left[\mathsf{E}\left\{|\eta|^2\right\}\right]^{-1}\mathsf{E}\left\{|\xi\eta|\right\} \\ &\geq \mathsf{E}\left\{\xi\eta\right\}\left[\mathsf{E}\left\{|\eta|^2\right\}\right]^{-1}\mathsf{E}\left\{\xi\eta\right\} \end{aligned}} \tag{13.75}$$

Below we shall show that the last inequality (13.75) admits the generalization for the vector case.

13.5.1.2 Nonnegative definiteness of a symmetric block-matrix

Lemma 13.5. *Let S be a symmetric nonnegative definite block-matrix such that*

$$\boxed{\begin{gathered} S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \geq 0 \\ 0 \leq S_{11} = S_{11}^{\mathsf{T}} \in \mathbb{R}^{N\times N} \\ 0 < S_{22} = S_{22}^{\mathsf{T}} \in \mathbb{R}^{M\times M}, \quad S_{12} = S_{21}^{\mathsf{T}} \end{gathered}}$$

Then obligatorily

$$\boxed{S_{11} \geq S_{12}S_{22}^{-1}S_{21}^{\mathsf{T}}} \tag{13.76}$$

Proof. Since $S \geq 0$, it follows that for any $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$

$$\begin{aligned} 0 &\leq \begin{pmatrix} x \\ y \end{pmatrix}^{\mathsf{T}} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= x^{\mathsf{T}}S_{11}x + x^{\mathsf{T}}S_{12}y + y^{\mathsf{T}}S_{21}x + y^{\mathsf{T}}S_{22}y \end{aligned}$$

Using the presentation

$$S_{22} = S_{22}^{1/2}S_{22}^{1/2}\left(\left(S_{22}^{1/2}\right)^{\mathsf{T}} = S_{22}^{1/2} > 0\right)$$

and letting

$$y = -S_{22}^{-1}S_{21}^{\mathsf{T}}x$$

one gets

$$\begin{aligned} 0 \leq\ & x^{\mathsf{T}}S_{11}x - x^{\mathsf{T}}S_{12}S_{22}^{-1}S_{21}^{\mathsf{T}}x - x^{\mathsf{T}}S_{21}S_{22}^{-1}S_{21}x \\ &+ x^{\mathsf{T}}S_{21}S_{22}^{-1}S_{22}S_{22}^{-1}S_{21}^{\mathsf{T}}x = x^{\mathsf{T}}S_{11}x - x^{\mathsf{T}}S_{12}S_{22}^{-1}S_{21}^{\mathsf{T}}x - z^{\mathsf{T}}S_{21}S_{22}^{-1}S_{21}x \\ &+ z^{\mathsf{T}}S_{21}S_{22}^{-1}S_{21}^{\mathsf{T}}x = x^{\mathsf{T}}S_{11}x - x^{\mathsf{T}}S_{12}S_{22}^{-1}S_{21}^{\mathsf{T}}x = x^{\mathsf{T}}\left[S_{11} - S_{12}S_{22}^{-1}S_{21}^{\mathsf{T}}\right]x \end{aligned}$$

Since this inequality is valid for any $x \in \mathbb{R}^N$ (13.76) follows. Lemma is proven. □

13.5.1.3 Main inequality for stochastic vectors

Theorem 13.2. *Let $\xi \in \mathbb{R}^N$, $\eta \in \mathbb{R}^M$ be random vectors defined on $(\Omega, \mathcal{F}, P)$ such that their second moments exists, i.e.,*

$$\boxed{\mathsf{E}\left\{\|\xi\|^2\right\} < \infty, \quad \mathsf{E}\left\{\|\eta\|^2\right\} < \infty} \tag{13.77}$$

and additionally, the covariation matrix $\mathsf{E}\{\eta\eta^\intercal\}$ is non-singular, namely,

$$\boxed{\mathsf{E}\left\{\eta\eta^\intercal\right\} > 0} \tag{13.78}$$

Then the following matrix inequality holds:

$$\boxed{\mathsf{E}\left\{\xi\xi^\intercal\right\} \geq \mathsf{E}\left\{\xi\eta^\intercal\right\}\left[\mathsf{E}\left\{\eta\eta^\intercal\right\}\right]^{-1}\mathsf{E}\left\{\eta\xi^\intercal\right\}} \tag{13.79}$$

Proof. Since

$$0 \leq \begin{pmatrix} \xi \\ \eta \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix}^\intercal = \begin{pmatrix} \xi\xi^\intercal & \xi\eta^\intercal \\ \eta\xi^\intercal & \eta\eta^\intercal \end{pmatrix}$$

and applying the mathematical expectation to both sides of this inequality, we derive

$$0 \leq \begin{pmatrix} \mathsf{E}\{\xi\xi^\intercal\} & \mathsf{E}\{\xi\eta^\intercal\} \\ \mathsf{E}\{\eta\xi^\intercal\} & \mathsf{E}\{\eta\eta^\intercal\} \end{pmatrix}$$

The desired result (13.79) follows from this inequality and Lemma 13.5. □

Remark 13.5. *The inequality (13.75) results from (13.79) if we consider the case $N = M = 1$.*

Now we are ready to formulate the main result of this section.

13.5.2 Fisher information matrix

Definition 13.5. *The collection $y^n := \{y_1, \ldots, y_n\}$ of random vectors $y \in R^N$ defined on $(\Omega, \mathcal{F}, P)$ is called* ***regular with respect to a vector parameter*** *$c \in C \subseteq R^K$ if the following* ***regularity conditions*** *hold: there exists the joint distribution density function*

$$p_{y^n}(v \mid c) := p_{y^n}(v_1, \ldots, v_n \mid c)$$

depending on the vector parameter c such that

1.

$$\boxed{\sup_{c \in C} \mathsf{E}\left\{\left\|\nabla_c \ln p_{y^n}\left(y^n \mid c\right)\right\|^2\right\} < \infty} \tag{13.80}$$

2. *for any* $c \in C$ *the* ***Fisher information matrix*** $\mathbb{I}_F(c,n)$ *defined by*

$$\boxed{\mathbb{I}_F(c,n) := \mathsf{E}\left\{\nabla_c \ln p_{y^n}\left(y^n \mid c\right) \nabla_c^{\intercal} \ln p_{y^n}\left(y^n \mid c\right)\right\}} \tag{13.81}$$

is ***strictly positive****, that is,*

$$\boxed{\mathbb{I}_F(c,n) > 0} \tag{13.82}$$

The next lemma combines several important properties of $\mathbb{I}_F(c,n)$.

Lemma 13.6. (on some properties of $\mathbb{I}_F(c,n)$)

1. *If random vectors* $y_1, \ldots, y_n$ *are independent then*

$$\boxed{\mathbb{I}_F(c,n) = \sum_{t=1}^{n} \mathbb{I}_F^t(c)} \tag{13.83}$$

where $\mathbb{I}_F^t(c)$ *is the Fisher information matrix of the random vector* y_t *and is given by*

$$\boxed{\begin{aligned} \mathbb{I}_F^t(c) &= \mathsf{E}\left\{\nabla_c \ln p_{y_t}(y_t \mid c)\, \nabla_c^{\intercal} \ln p_{y_t}(y_t \mid c)\right\} \\ &= \int\limits_{v \in R^N} \left[\nabla_c \ln p_{y_t}(v \mid c)\, \nabla_c^{\intercal} \ln p_{y_t}(v \mid c)\right] p_{y_t}(v \mid c)\, dv \\ &= \int\limits_{v \in R^N:\ p_{y_t}(v|c)>0} \frac{\nabla_c p_{y_t}(v \mid c)\, \nabla_c^{\intercal} p_{y_t}(v \mid c)}{p_{y_t}(v \mid c)} dv \end{aligned}} \tag{13.84}$$

2. *If random vectors* $y_1, \ldots, y_n$ *are independent and identically distributed then*

$$\boxed{\mathbb{I}_F(c,n) = n\, \mathbb{I}_F^1(c)} \tag{13.85}$$

Proof.

1. Since for a collection y^n of independent random vectors $y_1, \ldots, y_n$ satisfying

$$p_{y^n}(v_1, \ldots, v_n \mid c) = \prod_{t=1}^{n} p_{y_t}(v_t \mid c)$$

one has

$$\ln p_{y^n}(v_1, \ldots, v_n \mid c) = \sum_{t=1}^{n} \ln p_{y_t}(v_t \mid c)$$

and therefore,

$$\nabla_c \ln p_{y^n}\left(y^n \mid c\right) = \sum_{t=1}^{n} \nabla_c \ln p_{y_t}\left(v_t \mid c\right)$$

By (13.81) it follows that

$$\begin{aligned}
\mathbb{I}_F(c,n) &:= \mathsf{E}\left\{\nabla_c \ln p_{y^n}\left(y^n \mid c\right) \nabla_c^{\intercal} \ln p_{y^n}\left(y^n \mid c\right)\right\} \\
&= \mathsf{E}\left\{\sum_{t=1}^{n} \nabla_c \ln p_{y_t}\left(y_t \mid c\right) \sum_{t=1}^{n} \nabla_c^{\intercal} \ln p_{y_t}\left(y_t \mid c\right)\right\} \\
&= \mathsf{E}\left\{\sum_{t=1}^{n}\sum_{s=1}^{n} \nabla_c \ln p_{y_t}\left(y_t \mid c\right) \nabla_c^{\intercal} \ln p_{y_s}\left(y_s \mid c\right)\right\} \\
&= \sum_{t=1}^{n}\sum_{s=1}^{n} \mathsf{E}\left\{\nabla_c \ln p_{y_t}\left(y_t \mid c\right) \nabla_c^{\intercal} \ln p_{y_s}\left(y_s \mid c\right)\right\} \\
&= \sum_{t=1}^{n}\sum_{\substack{s=1 \\ s\neq t}}^{n} \mathsf{E}\left\{\nabla_c \ln p_{y_t}\left(y_t \mid c\right)\right\} \mathsf{E}\left\{\nabla_c^{\intercal} \ln p_{y_s}\left(y_s \mid c\right)\right\} \\
&\quad + \sum_{t=1}^{n} \mathsf{E}\left\{\nabla_c \ln p_{y_t}\left(y_t \mid c\right) \nabla_c^{\intercal} \ln p_{y_t}\left(y_t \mid c\right)\right\}
\end{aligned} \tag{13.86}$$

Notice that

$$\begin{aligned}
\mathsf{E}\left\{\nabla_c \ln p_{y_t}\left(y_t \mid c\right)\right\} &= \int_{v\in R^N} \nabla_c \ln p_{y_t}\left(v \mid c\right) p_{y_t}\left(v \mid c\right) dv \\
&= \mathsf{E} \int_{v\in R^N} \nabla_c p_{y_t}\left(v \mid c\right) dv = 0
\end{aligned}$$

which results by differentiating on c the identity

$$\int_{v\in R^N} p_{y_t}\left(v \mid c\right) dv = 1$$

Hence (13.86) is

$$\mathbb{I}_F(c,n) = \sum_{t=1}^{n} \mathbb{I}_F^t(c)$$

2. The formula (13.85) is the evident sequence of (13.83).
Lemma is proven. □

Example 13.3. *Consider a* ***Gaussian random variable*** $y \in \mathbb{R}$ *having the distribution density function*

$$p_y\left(v \mid c\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(v-c)^2}{2\sigma^2}\right\}$$

Then

$$\begin{aligned}
\mathbb{I}_F^1(c) &= \int\limits_{v\in\mathbb{R}:p_y(v|c)>0} \frac{\left(\frac{d}{dc}p_y(v\mid c)\right)^2}{p_y(v\mid c)}dv \\
&= \mathbb{I}_F^1(c) = \int\limits_{v\in\mathbb{R}:p_y(v|c)>0} \frac{\left(\frac{d}{dv}p_y(v\mid c)\right)^2}{p_y(v\mid c)}dv \\
&= \int\limits_{v\in\mathbb{R}:p_{y_t}(v|c)>0} \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(v-c)^2}{2\sigma^2}\right\}\left(\frac{(v-c)}{\sigma^2}\right)\right)^2}{p_y(v\mid c)}dv \\
&= \int\limits_{v\in\mathbb{R}} (v-c)^2\frac{1}{\sigma^4}p_y(v\mid c)\,dv = \sigma^{-2}
\end{aligned} \tag{13.87}$$

Example 13.4. *For a **Laplace random variable** $y \in \mathbb{R}$ having the distribution density function*

$$p_y(v\mid c) = \frac{1}{2a}\exp\left\{-\frac{|v-c|}{a}\right\}$$

one has

$$\begin{aligned}
\mathbb{I}_F^1(c) &= \int\limits_{v\in\mathbb{R}:p_y(v|c)>0} \frac{\left(\frac{d}{dv}p_y(v\mid c)\right)^2}{p_y(v\mid c)}dv \\
&= \int\limits_{v\in\mathbb{R}:p_{y_t}(v|c)>0} \frac{\left(\frac{1}{2a}\exp\left\{-\frac{|v-c|}{a}\right\}\left(\frac{\operatorname{sign}(v-c)}{a}\right)\right)^2}{p_y(v\mid c)}dv \\
&= \frac{1}{a^2}\int\limits_{v\in\mathbb{R}} p_y(v\mid c)\,dv = a^{-2}
\end{aligned} \tag{13.88}$$

Example 13.5. *For a **random variable** $y \in \mathbb{R}$ having the distribution density function*

$$p_y(v\mid c) = \left\{\begin{matrix} \frac{\pi}{a}\cos^2\left(\frac{\pi}{2a}(v-c)\right) & \text{for } |v-c|\le a \\ 0 & \text{for } |v-c|>a \end{matrix}\right]$$

one has

$$\mathbb{I}_F^1(c) = \int\limits_{v\in\mathbb{R}:p_y(v|c)>0} \frac{\left(\frac{d}{dv}p_y(v\mid c)\right)^2}{p_y(v\mid c)}dv$$

$$= \int\limits_{v \in \mathbb{R}: p_{y_t}(v|c)>0} \frac{\left(\frac{\pi}{a}\cos\left(\frac{\pi}{2a}(v-c)\right)\sin\left(\frac{\pi}{2a}(v-c)\right)2\frac{\pi}{2a}\right)^2}{p_y(v \mid c)} dv$$

$$= \left(\frac{\pi}{a}\right)^2 \int\limits_{v:|v-c|\le a} \sin^2\left(\frac{\pi}{2a}(v-c)\right) p_y(v \mid c)\, dv$$

$$= \left(\frac{\pi}{a}\right)^2 \int\limits_{v:|v-c|\le a} \left[1-\cos^2\left(\frac{\pi}{2a}(v-c)\right)\right] p_y(v \mid c)\, dv$$

$$= \left(\frac{\pi}{a}\right)^2 - \left(\frac{\pi}{a}\right)^2 \int\limits_{v:|v-c|\le a} \cos^3\left(\frac{\pi}{2a}(v-c)\right) dv = \left(\frac{\pi}{a}\right)^2$$

Exercise 13.1. *Show that for any absolutely continuous and quadratically integrable random variable ζ with a symmetric distribution $p_\zeta(v)$ the following inequality holds:*

$$\boxed{I_F\left(p_\zeta\right)\sigma_\zeta^2 \ge 1} \tag{13.89}$$

where

$$\boxed{\begin{array}{c} \mathbb{I}_F\left(p_\zeta\right) := \displaystyle\int\limits_{v\in\mathbb{R}:\, p_\zeta(v)>0} \left[\frac{d}{dv}\ln p_\zeta(v)\right]^2 p_\zeta(v)\, dv \\ \sigma_\zeta^2 := \mathsf{E}\left\{(\zeta - \mathsf{E}\{\zeta\})^2\right\} \end{array}}$$

Indeed, letting in (13.74)

$$\xi := \frac{d}{dv}\ln p_\zeta(v)\,|_{v=\zeta}, \quad \eta := \zeta - \mathsf{E}\{\zeta\}$$

we obtain

$$\mathsf{E}\left\{|\xi|^2\right\} = \mathbb{I}_F\left(p_\zeta\right), \quad \mathsf{E}\left\{|\eta|^2\right\} = \sigma_\zeta^2$$

Since

$$\int\limits_{v\in\mathbb{R}} \frac{d}{dv} p_\zeta(v)\, dv = \int\limits_{v\in\mathbb{R}} dp_\zeta(v) = p_\zeta(\infty) - p_\zeta(-\infty) = 0 - 0 = 0$$

and, in view of the symmetricity property, after differentiating by c both parts of the identity

$$\int\limits_{v\in\mathbb{R}} v p_\zeta(v-c)\, dv = \int\limits_{v\in\mathbb{R}} (v-c)\, p_\zeta(v-c)\, dv + c = c$$

we get

$$\int\limits_{v\in\mathbb{R}} v\frac{d}{dv} p_\zeta(v)\, dv = \int\limits_{v\in\mathbb{R}} v\frac{d}{dc} p_\zeta(v-c)\,|_{c=0}\, dv = 1$$

Therefore

$$\begin{aligned}\mathsf{E}\{\xi\eta\} &= \int\limits_{v\in\mathbb{R}:\ p_\zeta(v)>0} (v-\mathsf{E}\{\zeta\})\frac{d}{dv}\ln p_\zeta(v)\,p_\zeta(v)\,dv \\ &= \int\limits_{v\in R}(v-\mathsf{E}\{\zeta\})\frac{d}{dv}p_\zeta(v)\,dv \\ &= \int\limits_{v\in\mathbb{R}} v\frac{d}{dv}p_\zeta(v)\,dv + \mathsf{E}\{\zeta\}\int\limits_{v\in\mathbb{R}}\frac{d}{dv}p_\zeta(v)\,dv = 1\end{aligned}$$

which implies (13.89).

Notice that, in view of (13.87),

$$\boxed{I_F\left(p_\zeta\right)\sigma_\zeta^2 = 1}$$

only for Gaussian random variables.

13.5.3 Cramér–Rao inequality

The main theorem of this section is as follows (see Cramér (1957)).

Theorem 13.3. (Cramer–Rao, 1946) *For any parameter estimate (a function of available data) $c_n = c_n(y_1, y_2, \ldots, y_n)$ which is* ***unbiased*** *and* ***quadratically integrable****, i.e., for any $c \in C \subseteq \mathbb{R}^K$*

$$\mathsf{E}\{c_n\} = c, \quad \sup_{c\in C}\mathsf{E}\left\{\|c_n - c\|^2\right\} < \infty \tag{13.90}$$

under ***regular observations*** *y^n the following information bound holds:*

$$\boxed{\mathsf{E}\left\{(c_n - c)(c_n - c)^\mathsf{T}\right\} \ge \mathbb{I}_F^{-1}(c, n)} \tag{13.91}$$

that exactly means that ***under regular observations it is impossible to reach the quality of estimating*** $\mathsf{E}\{(c_n - c)(c_n - c)^\mathsf{T}\}$ ***better than*** $\mathbb{I}_F^{-1}(c, n)$.

Proof. Let us put in (13.79)

$$\xi := c_n - c, \quad \eta := \nabla_c \ln p_{y^n}\left(y^n \mid c\right)$$

By the unbiased property

$$0 = \mathsf{E}\{c_n - c\} = \int\limits_{v\in R^N}(c_n - c)\,p_{y^n}(v \mid c)\,dv = 0$$

Differentiating this identity implies

$$0 = -I \int\limits_{v \in R^N} p_{y^n}(v \mid c)\, dv + \int\limits_{v \in R^N} (c_n - c)\, \nabla_c^{\intercal} p_{y^n}(v \mid c)\, dv$$

or equivalently,

$$\mathsf{E}\left\{\xi\eta^{\intercal}\right\} = \mathsf{E}\left\{(c_n - c)\, \nabla_c^{\intercal} \ln p_{y^n}\left(y^n \mid c\right)\right\} = I$$

Analogously

$$\mathsf{E}\left\{\eta\xi^{\intercal}\right\} = \mathsf{E}\left\{\nabla_c \ln p_{y^n}\left(y^n \mid c\right)(c_n - c)^{\intercal}\right\} = I$$

Noticing that

$$\mathsf{E}\left\{\eta\eta^{\intercal}\right\} = \mathsf{E}\left\{\nabla_c \ln p_{y^n}\left(y^n \mid c\right) \nabla_c^{\intercal} \ln p_{y^n}\left(y^n \mid c\right)\right\} = \mathbb{I}_F(c, n)$$

and using (13.79) leads to (13.91). Theorem is proven. □

13.6 Efficient estimates

13.6.1 Efficient and asymptotic efficient estimates

13.6.1.1 The main definition

Based on the *Cramer–Rao inequality* (13.91) we may introduce the following important definitions.

Definition 13.6. *An unbiased and quadratically integrable parameter estimate* $c_n = c_n(y_1, y_2, \ldots, y_n)$*, constructed based on regular observations, is said to be*

(a) ***efficient****, if it verifies the exact equality in (13.91);*
(b) ***asymptotically efficient****, if it ensures the equality*

$$\boxed{\lim_{n\to\infty} n\, \mathsf{E}\left\{(c_n - c)(c_n - c)^{\intercal}\right\} = \lim_{n\to\infty} n\mathbb{I}_F^{-1}(c, n)} \tag{13.92}$$

(provided that both above limits exist).

13.6.1.2 Information bound in the general scalar-output case

Here we will calculate the information bound

$$\boxed{\mathcal{I} := \lim_{n\to\infty} n\mathbb{I}_F^{-1}(c, n)} \tag{13.93}$$

for the measuring scheme (13.15) with *white noise*, i.e.,

$$\boxed{x_{n+1} = Cz_n + \xi_n} \tag{13.94}$$

generating the data set $(x_1, z_0; x_2, z_1; \ldots; x_{n+1}, z_n)$ where $x_n \in \mathbb{R}$.

Theorem 13.4. *If in (13.94) with i.i.d. centered 'noise' sequence* $\{\xi_n\}$ *the generalized inputs* $\{z_n\}$ *satisfy the following conditions:*

1. *'strong law of large number' (SLNL) for* $\{z_n\}$

$$\boxed{\left\| \frac{1}{n}\sum_{t=0}^{n}\left\{z_t z_t^{\mathsf{T}}\right\} - \frac{1}{n}\sum_{t=0}^{n}\mathsf{E}\left\{z_t z_t^{\mathsf{T}}\right\}\right\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{13.95}$$

2. *the convergence of 'averaged' inputs covariation*

$$\boxed{\frac{1}{n}\sum_{t=0}^{n}\mathsf{E}\left\{z_t z_t^{\mathsf{T}}\right\} \underset{n\to\infty}{\to} \mathcal{R} > 0} \tag{13.96}$$

3. z_n *(under a fixed prehistory* $z^{n-1} := (z_1, \ldots, z_{n-1})$ *and fixed* x_n*) does not depend on* c*, that is,*

$$\nabla_c \ln p_{z_n}\left(v_n^z \mid x_n, z^{t-1}, c\right) = 0 \quad (n = 1, \ldots)$$

then ***the information bound*** *under the regular data* $y_n := \left(x_{n+1}^{\mathsf{T}}, z_n^{\mathsf{T}}\right)^{\mathsf{T}} \in R^{1+K}$ *is*

$$\boxed{\mathcal{I} = \lim_{n\to\infty} n\mathbb{I}_F^{-1}(c, n) = \mathcal{R}^{-1} I_F^{-1}\left(p_{\xi_1}\right)} \tag{13.97}$$

where

$$\boxed{I_F\left(p_{\xi_1}\right) := \mathsf{E}\left\{\left[\left(\ln p_{\xi_1}(\xi_1)\right)'\right]^2\right\}} \tag{13.98}$$

Proof. By the Bayes formula (1.69) and by definition (13.81) we have

$$\begin{aligned}
p_{y^n}(v_1, \ldots, v_n \mid c) &= p_{y_n}\left(v_n \mid y^{n-1}, c\right) p_{y^{n-1}}(v_1, \ldots, v_{n-1} \mid c) \\
&= \cdots = \left[\prod_{t=1}^{n} p_{y_t}\left(v_t \mid y^{t-1}, c\right)\right] p_{y_1}(v_1 \mid y_0, c) \\
&= \prod_{t=0}^{n} p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right) p_{z_t}\left(v_t^z \mid y^{t-1}, c\right)
\end{aligned}$$

and hence,

$$\begin{aligned}
\mathbb{I}_F(c, n) &= \mathsf{E}\left\{\nabla_c \ln p_{y^n}\left(y^n \mid c\right) \nabla_c^{\mathsf{T}} \ln p_{y^n}\left(y^n \mid c\right)\right\} \\
&= \mathsf{E}\left\{\sum_{t=1}^{n} \nabla_c \ln p_{y_t}\left(v_t \mid y^{t-1}, c\right) \sum_{t=1}^{n} \nabla_c^{\mathsf{T}} \ln p_{y_t}\left(v_t \mid y^{t-1}, c\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n}\sum_{s=1}^{n} \mathsf{E}\left\{\nabla_c \ln p_{y_t}\left(v_t \mid y^{t-1}, c\right) \nabla_c^{\intercal} \ln p_{y_s}\left(v_s \mid y^{s-1}, c\right)\right\} \\
&= \sum_{t=1}^{n}\sum_{s=1}^{n} \mathsf{E}\left\{\left[\nabla_c \ln p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right) + \nabla_c \ln p_{z_t}\left(v_t^z \mid y^{t-1}, c\right)\right]\right. \\
&\quad \left. \cdot \left[\nabla_c\left(\ln p_{x_{s+1}}\left(v_s^x \mid z_s, y^{s-1}, c\right) + \nabla_c \ln p_{z_s}\left(v_s^z \mid y^{s-1}, c\right)\right)\right]^{\intercal}\right\}
\end{aligned}
$$

Taking into account that

$$
\begin{aligned}
&p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right) = p_{\xi_t}\left(v_t^x - c^{\intercal} z_t \mid z_t, y^{t-1}, c\right) = p_{\xi_t}\left(v_t^x - c^{\intercal} z_t\right) \\
&\nabla_c \ln p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right) = -\frac{z_t}{p_{\xi_t}\left(v_t^x - c^{\intercal} z_t\right)} p'_{\xi_t}\left(v_t^x - c^{\intercal} z_t\right) \\
&\nabla_c \ln p_{z_t}\left(v_t^z \mid y^{t-1}, c\right) = \nabla_c \ln p_{z_t}\left(v_t^z \mid x_t, z^{t-1}, c\right) = 0 \ (t = 1, \ldots)
\end{aligned}
$$

and since $\{\xi_t\}_{t\geq 1}$ is i.i.d. centered sequence, it follows that for $t \neq s$

$$
\begin{aligned}
&\mathsf{E}\left\{\nabla_c \ln p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right) \nabla_c^{\intercal} \ln p_{x_{s+1}}\left(v_s^x \mid z_s, y^{s-1}, c\right)\right\} \\
&\quad = \mathsf{E}\left\{\nabla_c \ln p_{x_{t+1}}\left(v_t^x \mid z_t, y^{t-1}, c\right)\right\} \mathsf{E}\left\{\nabla_c^{\intercal} \ln p_{x_{s+1}}\left(v_s^x \mid z_s, y^{s-1}, c\right)\right\} = 0
\end{aligned}
$$

and, as a result,

$$
\mathbb{I}_F\left(c, n\right) = \sum_{t=1}^{n} \mathsf{E}\left\{z_t z_t^{\intercal}\left[\left(\ln p_{\xi_t}\left(\xi_t\right)\right)'\right]^2\right\} = \mathsf{E}\left\{\sum_{t=1}^{n} z_t z_t^{\intercal}\right\} I_F\left(p_{\xi_t}\right)
$$

which implies (13.97). Theorem is proven. □

Below, based on the definitions and statements above, we will analyze some partial models and corresponding asymptotically efficient algorithms.

13.6.1.3 LSM for R-models under Gaussian noises

Consider the regression model (13.6)

$$
\begin{aligned}
&y_{n+1} = \sum_{t=0}^{L_b} b_t v_{n-t} + \xi_n = (\mathbf{b}, \mathbf{v}_n) + \xi_n \\
&\mathbf{b}^{\intercal} = \left(b_0, \ldots, b_{L_b}\right), \quad \mathbf{v}_n^{\intercal} = \left(v_n, v_{n-1}, \ldots, v_{n-L_b}\right)
\end{aligned}
$$

with stationary white-noise terms ξ_n having Gaussian distribution $p_{\xi}(x) = \mathcal{N}\left(0, \sigma^2\right)$. Applying the LSM algorithm (13.20) for the identification of the vector $\mathbf{b}$ we obtain

$$
\begin{aligned}
&\mathbf{b}_n = Z_n^{-1} V_n \\
&V_n := \sum_{t=0}^{n} \mathbf{v}_t y_{t+1}, \quad Z_n := \sum_{t=0}^{n} \mathbf{v}_t \mathbf{v}_t^{\intercal} = Z_n^{\intercal} > 0
\end{aligned}
$$

Then we can get the expression (13.45) for the identification error in the following form:

$$\Delta_n := \mathbf{b}_n - \mathbf{b} = Z_n^{-1} V_n - \mathbf{b} = Z_n^{-1} \left(\sum_{t=0}^{n} \mathbf{v}_t y_{t+1} \right) - \mathbf{b}$$
$$= Z_n^{-1} \left(\sum_{t=0}^{n} \mathbf{v}_t \left[\mathbf{v}_t^{\mathsf{T}} \mathbf{b} + \xi_t \right] \right) - \mathbf{b} = Z_n^{-1} \sum_{t=0}^{n} \mathbf{v}_t \xi_t$$

Supposing that $\{\mathbf{v}_t\}$ is a deterministic regressor, we derive the following expression for the *error covariance matrix*:

$$\mathsf{E}\left\{ (\mathbf{b}_n - \mathbf{b})(\mathbf{b}_n - \mathbf{b})^{\mathsf{T}} \right\} = \mathsf{E}\left\{ \Delta_n \Delta_n^{\mathsf{T}} \right\} = \mathsf{E}\left\{ \left(Z_n^{-1} \sum_{t=0}^{n} \mathbf{v}_t \xi_t \right) \left(\sum_{t=0}^{n} \xi_t \mathbf{v}_t^{\mathsf{T}} Z_n^{-1} \right) \right\}$$
$$= Z_n^{-1} \mathsf{E}\left\{ \sum_{t=0}^{n} \sum_{s=0}^{n} \mathbf{v}_t \mathbf{v}_s^{\mathsf{T}} \xi_t \xi_s \right\} Z_n^{-1} = Z_n^{-1} \sum_{t=0}^{n} \sum_{s=0}^{n} \mathbf{v}_t \mathbf{v}_s^{\mathsf{T}} \mathsf{E}\left\{ \xi_t \xi_s \right\} Z_n^{-1}$$
$$= Z_n^{-1} \left(\sum_{t=0}^{n} \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} \right) Z_n^{-1} \sigma^2 = Z_n^{-1} \sigma^2 = \sigma^2 \left(\sum_{t=0}^{n} \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} \right)^{-1} \tag{13.99}$$

Using the definition of the Fisher information matrix (13.81) and taking into account that the noise sequence is independent and stationary (13.83), one can easily realize the following calculation:

$$\mathbb{I}_F(\mathbf{b}, n) = \sum_{t=1}^{n} \mathbb{I}_F^t(\mathbf{b})$$
$$= \sum_{t=1}^{n} \int_{v \in R^N} \frac{\nabla_{\mathbf{b}} p_{y_{t+1}}(v \mid \mathbf{b}) \nabla_{\mathbf{b}}^{\mathsf{T}} p_{y_{t+1}}(v \mid \mathbf{b})}{p_{y_{t+1}}(v \mid \mathbf{b})} dv$$
$$= \sum_{t=1}^{n} \int_{v \in R^N} \frac{\nabla_{\mathbf{b}} p_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right) \nabla_{\mathbf{b}}^{\mathsf{T}} p_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right)}{p_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right)} dv$$
$$= \sum_{t=1}^{n} \int_{v \in R^N} \frac{p'_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right) \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} p'_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right)}{p_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right)} dv$$
$$= \sum_{t=1}^{n} \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} \int_{v \in R^N} \frac{\left[p'_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right) \right]^2}{p_{\xi_t}\left(v - \mathbf{v}_t^{\mathsf{T}} \mathbf{b} \mid \mathbf{b}\right)} dv \overset{(13.87)}{=} \sigma^{-2} \sum_{t=1}^{n} \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} \tag{13.100}$$

Comparing (13.99) with (13.100) we conclude that

$$\boxed{\mathsf{E}\left\{ (\mathbf{b}_n - \mathbf{b})(\mathbf{b}_n - \mathbf{b})^{\mathsf{T}} \right\} = \sigma^2 \left(\sum_{t=0}^{n} \mathbf{v}_t \mathbf{v}_t^{\mathsf{T}} \right)^{-1} = \mathbb{I}_F^{-1}(\mathbf{b}, n)}$$

which exactly means the following.

Claim 13.3. *The LSM estimate (13.20) being applied to an R-model identification with **deterministic regressors** and under Gaussian stationary noises is **effective**, that is, in this situation **there does not exist any other estimate better than LSM**. **Under noises with other distributions LSM estimates are non-effective**.*

Claim 13.4. *It is not difficult to check that the LSM estimate (13.20) being applied to an R-model identification with **stochastic regressors,** which are independent of Gaussian stationary noises, is **asymptotically effective**, that is, in this situation **there does not exist** (see inequality (13.89)) **any other estimate better (asymptotically) than LSM**. **Under noises with other distributions LSM estimates are non-asymptotically effective**.*

13.6.1.4 LSM for AR models under Gaussian noises

For the AR model (13.3), analogously to the previous example, we have

$$y_{n+1} = \sum_{s=0}^{L_a} a_s y_{n-s} + \xi_n = \mathbf{a}^\mathsf{T} \mathbf{v}_n + \xi_n$$
$$\mathbf{a}^\mathsf{T} = \left(a_0, \ldots, a_{L_a}\right), \quad \mathbf{v}_n^\mathsf{T} = \left(y_n, y_{n-1}, \ldots, y_{n-L_a}\right)$$

and

$$\begin{aligned} &\mathsf{E}\left\{(\mathbf{a}_n - \mathbf{a})(\mathbf{a}_n - \mathbf{a})^\mathsf{T}\right\} = \mathsf{E}\left\{\left(Z_n^{-1}\sum_{t=0}^{n}\mathbf{v}_t\xi_t\right)\left(\sum_{t=0}^{n}\xi_t\mathbf{v}_t^\mathsf{T} Z_n^{-1}\right)\right\} \\ &= \mathsf{E}\left\{Z_n^{-1}\sum_{t=0}^{n}\sum_{s=0}^{n}\mathbf{v}_t\mathbf{v}_s^\mathsf{T}\xi_t\xi_s Z_n^{-1}\right\} = \sum_{t=0}^{n}\sum_{s=0}^{n}\mathsf{E}\left\{Z_n^{-1}\mathbf{v}_t\mathbf{v}_s^\mathsf{T} E\left\{\xi_t\xi_s\right\} Z_n^{-1}\right\} \\ &= \sigma^2\sum_{t=0}^{n}\mathsf{E}\left\{Z_n^{-1}\mathbf{v}_t\mathbf{v}_t^\mathsf{T} Z_n^{-1}\right\} = \sigma^2\mathsf{E}\left\{Z_n^{-1}\left(\sum_{t=0}^{n}\mathbf{v}_t\mathbf{v}_t^\mathsf{T}\right)Z_n^{-1}\right\} \\ &= \sigma^2\mathsf{E}\left\{Z_n^{-1}\right\} \end{aligned} \tag{13.101}$$

or equivalently,

$$n\mathsf{E}\left\{(\mathbf{a}_n - \mathbf{a})(\mathbf{a}_n - \mathbf{a})^\mathsf{T}\right\} = \sigma^2\mathsf{E}\left\{\left(\frac{1}{n}\sum_{t=0}^{n}\mathbf{v}_t\mathbf{v}_t^\mathsf{T}\right)^{-1}\right\}$$

Also, by Bayes' rule, it follows that

$$p_{y^{n+1}}(v \mid \mathbf{a}) = p_{y_{n+1}}\left(v_{n+1} \mid y^n, \mathbf{a}\right) p_{y_n}\left(v_n \mid y^{n-1}, \mathbf{a}\right) \cdots p_{y_0}\left(v_0 \mid y^{-1}, \mathbf{a}\right)$$

and hence,

$$\ln p_{y^{n+1}}(v \mid \mathbf{a}) = \sum_{t=0}^{n} \ln p_{y_{t+1}}\left(v_{t+1} \mid y^t, \mathbf{a}\right)$$

Notice that $p_{y_0}\left(v_0 \mid y^{-1}, \mathbf{a}\right)$ does not depends on $\mathbf{a}$. So,

$$\begin{aligned}\nabla_{\mathbf{a}} \ln p_{y^{n+1}}(v \mid \mathbf{a}) &= \sum_{t=1}^{n} \nabla_{\mathbf{a}} \ln p_{y_{t+1}}\left(v_{t+1} \mid y^t, \mathbf{a}\right) \\ &= \sum_{t=0}^{n} \nabla_{\mathbf{a}} \ln p_{\xi_t}\left(v_{t+1} - \mathbf{v}_t^{\intercal}\mathbf{a} \mid y^t, \mathbf{a}\right)\end{aligned}$$

and therefore, by the noise independence property

$$\begin{aligned}\mathbb{I}_F(\mathbf{a}, n) &= E\left\{\sum_{t=1}^{n} \int_{v\in R^N} \frac{\nabla_{\mathbf{a}} p_{y_{t+1}}(v_{t+1} \mid \mathbf{a})\, \nabla_{\mathbf{a}}^{\intercal} p_{y_{t+1}}(v_{t+1} \mid \mathbf{a})}{p_{y_{t+1}}(v_{t+1} \mid \mathbf{a})} dv_{t+1}\right\} \\ &= \sum_{t=1}^{n} E\left\{\int_{v\in R^N} \frac{\nabla_{\mathbf{a}} p_{\xi_t}\left(v_{t+1} - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right) \nabla_{\mathbf{a}}^{\intercal} p_{\xi_t}\left(v_{t+1} - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)}{p_{\xi_t}\left(v_{t+1} - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)} dv_{t+1}\right\} \\ &= \sum_{t=1}^{n} E\left\{\int_{v\in R^N} \frac{p_{\xi_t}'\left(v - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right) \mathbf{v}_t \mathbf{v}_t^{\intercal} p_{\xi_t}'\left(v - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)}{p_{\xi_t}\left(v - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)} dv\right\} \\ &= \sum_{t=1}^{n} E\left\{\mathbf{v}_t \mathbf{v}_t^{\intercal} \int_{v\in R^N} \frac{\left[p_{\xi_t}'\left(v - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)\right]^2}{p_{\xi_t}\left(v - \mathbf{v}_t^{\intercal}\mathbf{a} \mid \mathbf{a}\right)} dv\right\} \overset{(13.87)}{=} \sigma^{-2} \sum_{t=1}^{n} E\left\{\mathbf{v}_t \mathbf{v}_t^{\intercal}\right\}\end{aligned}$$

By the strong large number law

$$\frac{1}{n} \sum_{t=0}^{n} \mathbf{v}_t \mathbf{v}_t^{\intercal} \overset{a.s.}{=} \frac{1}{n} \sum_{t=0}^{n} \mathsf{E}\left\{\mathbf{v}_t \mathbf{v}_t^{\intercal}\right\} + o_\omega(1)$$

For a stable auto-regression (when the polynomial $p_a(q) = \sum_{s=0}^{L_a} a_s q^s$ is Hurwitz, i.e., having all roots inside the united circle)

$$\begin{aligned}&\frac{1}{n} \sum_{t=0}^{n} \mathsf{E}\left\{\mathbf{v}_t \mathbf{v}_t^{\intercal}\right\} \to \mathcal{R} \\ &\mathcal{R} : \mathcal{R} = A\mathcal{R}A + \Xi\end{aligned} \tag{13.102}$$

where

$$A = \begin{Vmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ \mathbf{1} & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{1} & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & \mathbf{1} & 0 \end{Vmatrix}, \quad \Xi = \sigma^2 \begin{Vmatrix} \mathbf{1} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{Vmatrix}$$

Hence, for large enough n,

$$n\mathsf{E}\left\{(\mathbf{a}_n-\mathbf{a})(\mathbf{a}_n-\mathbf{a})^{\mathsf{T}}\right\}=\sigma^2\mathsf{E}\left\{\left(\frac{1}{n}\sum_{t=0}^{n}\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}}\right)^{-1}\right\}=\sigma^2\mathcal{R}^{-1}+o(1)$$

$$n\mathbb{I}_F^{-1}(\mathbf{a},n)\overset{(13.87)}{=}\sigma^2\left(\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}}\right\}\right)^{-1}=\sigma^2\mathcal{R}^{-1}+o(1)$$

which implies

$$\boxed{\lim_{n\to\infty}n\mathsf{E}\left\{(\mathbf{a}_n-\mathbf{a})(\mathbf{a}_n-\mathbf{a})^{\mathsf{T}}\right\}=\lim_{n\to\infty}n\mathbb{I}_F^{-1}(\mathbf{a},n)=\sigma^2\mathcal{R}^{-1}} \tag{13.103}$$

This identity exactly means the following.

Claim 13.5. *The LSM estimate (13.20) being applied to an AR-model identification with Gaussian stationary noises is* ***asymptotically effective****, that is, in this situation* ***there does not exist*** *(see inequality (13.89))* ***any other estimate better than LSM. Under noises with other distributions LSM estimates are non-asymptotically effective****.*

The natural question is: '*Which type of estimates are effective or asymptotically effective when noises are of a non-Gaussian nature?*'. The next section gives an answer to this question.

13.6.2 Recurrent LSM with a nonlinear residual transformation

13.6.2.1 Nonlinear residual transformation

Let us consider the LSM algorithm in its recurrent form (13.29) where instead of the scalar residual $(x_{n+2}-c_n z_{n+1})$ its *nonlinear transformation* $\varphi(x_{n+2}-c_n z_{n+1})$ is used, namely,

$$\boxed{\begin{aligned}c_{n+1}&=c_n+\Gamma_{n+1}z_{n+1}\varphi\left(x_{n+2}-c_n^{\mathsf{T}}z_{n+1}\right)\\ \Gamma_{n+1}&=\Gamma_n-\frac{\Gamma_n z_{n+1}z_{n+1}^{\mathsf{T}}\Gamma_n}{1+z_{n+1}^{\mathsf{T}}\Gamma_n z_{n+1}},\quad n\geq n_0\\ c_{n_0}&=Z_{n_0}^{-1}V_{n_0},\quad \Gamma_{n_0}=\left(\sum_{t=0}^{n_0}z_t z_t^{\mathsf{T}}\right)^{-1}=Z_{n_0}^{-1}\end{aligned}} \tag{13.104}$$

where $\varphi:R\to R$ is a function *to be selected to obtain a higher convergence rate* under a non-Gaussian regular noise. If a process (13.15) **under white noise** disturbances to be identified is **stable** (converges to a stationary process), then

$$n\Gamma_n:=nZ_n^{-1}=\left(\frac{1}{n}\sum_{t=0}^{n}\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}}\right)^{-1}\underset{n\to\infty}{\overset{a.s.}{\to}}$$

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=0}^{n}\mathsf{E}\left\{\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}}\right\}\right)^{-1}=\lim_{n\to\infty}\left(\mathsf{E}\left\{\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}}\right\}\right)^{-1}:=\mathcal{R}^{-1}$$

and hence,

$$\boxed{\Gamma_n \overset{a.s.}{\simeq} \frac{1}{n+1}\mathcal{R}^{-1}}$$

Claim 13.6. *Procedure (13.104) is* ***asymptotically equivalent*** *to the following one:*

$$\boxed{c_{n+1} = c_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\varphi\left(x_{n+2} - c_n^{\intercal} z_{n+1}\right)} \tag{13.105}$$

13.6.2.2 Convergence analysis and asymptotic normality

Theorem 13.5. (on an almost-sure convergence) *If*

1. *the conditions of Theorem 13.4 are fulfilled, and additionally,*

$$\boxed{\sup_n \mathsf{E}\left\{\|z_n\|^4\right\} < \infty} \tag{13.106}$$

2. *in the model (13.15) the scalar noise sequence $\{\xi_n\}_{n\geq 1}$ is i.i.d. and such that*

$$\boxed{\mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\left\{\xi_n^2\right\} = \sigma^2 > 0, \quad \mathsf{E}\left\{\xi_n^4\right\} = \mathsf{E}\left\{\xi_1^4\right\} < \infty} \tag{13.107}$$

3. *a nonlinear residual transformation $\varphi : \mathbb{R} \to \mathbb{R}$ provides for all $x \in \mathbb{R}$ the existence of the functions*

$$\boxed{\begin{array}{l} \psi(x) := \mathsf{E}\{\varphi(x+\xi_n)\} \\ S(x) := \mathsf{E}\left\{\varphi^2(x+\xi_n)\right\} \end{array}} \tag{13.108}$$

satisfying

$$\boxed{\begin{array}{c} x\psi(x) \geq \delta x^2, \quad \delta > 0 \\ S(x) \leq k_0 + k_1 x^2, \quad 0 \leq k_0, k_1 < \infty \end{array}} \tag{13.109}$$

then under any initial conditions c_0 the sequence $\{c_n\}_{n\geq 0}$, generated by the procedure (13.105), converges with probability one to the value $C = c$ participating in (13.15).

Proof. Define $\Delta_n := c_n - c$ for which we have

$$\Delta_{n+1} = \Delta_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\varphi\left(-\Delta_n^{\intercal} z_{n+1} + \xi_n\right) \tag{13.110}$$

and hence,

$$\|\Delta_{n+1}\|_{\mathcal{R}}^2 = \|\Delta_n\|_{\mathcal{R}}^2 + \frac{2}{n+1}\Delta_n^{\intercal} z_{n+1}\varphi\left(-\Delta_n^{\intercal} z_{n+1} + \xi_n\right) + \frac{1}{(n+1)^2}\left\|\mathcal{R}^{-1} z_{n+1}\right\|_{\mathcal{R}}^2 \varphi^2\left(-\Delta_n^{\intercal} z_{n+1} + \xi_n\right)$$

Taking conditional mathematical expectation $\mathsf{E}\{\cdot/\mathcal{F}_n\}$ of both sides of the last relation (with $\mathcal{F}_n := (\xi_0, z_0, x_1; \ldots; \xi_{n-1}, z_n, x_{n+1}; z_{n+1})$) we obtain

$$\mathsf{E}\left\{\|\Delta_{n+1}\|_{\mathcal{R}}^2 / \mathcal{F}_n\right\} \overset{a.s.}{=} \|\Delta_n\|_{\mathcal{R}}^2 + \frac{2}{n+1}\Delta_n^{\intercal} z_{n+1}\psi\left(-\Delta_n^{\intercal} z_{n+1}\right) + \frac{1}{(n+1)^2}\|z_{n+1}\|_{\mathcal{R}^{-1}}^2 S\left(-\Delta_n^{\intercal} z_{n+1}\right)$$

By the assumptions of this theorem the last relation implies

$$\begin{aligned}\mathsf{E}\left\{\|\Delta_{n+1}\|_{\mathcal{R}}^2 / \mathcal{F}_n\right\} &\overset{a.s.}{\le} \|\Delta_n\|_{\mathcal{R}}^2 - \frac{2\delta}{n+1}\left(\Delta_n^{\intercal} z_{n+1}\right)^2 \\ &\quad + \frac{1}{(n+1)^2}\|z_{n+1}\|_{\mathcal{R}^{-1}}^2\left[k_0 + k_1\left(\Delta_n^{\intercal} z_{n+1}\right)^2\right] \\ &\le \|\Delta_n\|_{\mathcal{R}}^2\left[1 + \frac{k_1}{(n+1)^2}\left\|\mathcal{R}^{-1}\right\|\|z_{n+1}\|^4\right] \\ &\quad - \frac{2\delta}{n+1}\left(\Delta_n^{\intercal} z_{n+1}\right)^2 + \frac{k_0}{(n+1)^2}\|z_{n+1}\|_{\mathcal{R}^{-1}}^2\end{aligned} \tag{13.111}$$

Therefore, the direct application of the Robbins–Siegmund theorem 7.11 guarantees that

$$\|\Delta_n\|_{\mathcal{R}}^2 \underset{n\to\infty}{\overset{a.s.}{\to}} w \tag{13.112}$$

and

$$\sum_{n=1}^{\infty}\frac{1}{n}\left(\Delta_{n-1}^{\intercal} z_n\right)^2 \overset{a.s.}{<} \infty \tag{13.113}$$

By the Abel identity (see Lemma 12.2 in Poznyak (2008))

$$\sum_{t=n_0}^{n} A_t B_t = A_n \sum_{t=n_0}^{n} B_t - \sum_{t=n_0}^{n}\left(A_t - A_{t-1}\right)\sum_{s=n_0}^{t-1} B_s$$

valid for any matrices A_t and B_t of the corresponding dimensions, we have

$$\begin{aligned}\sum_{n=n_0}^{N}\frac{1}{n}\left(\Delta_{n-1}^{\intercal} z_n\right)^2 &= \sum_{n=n_0}^{N}\frac{1}{n}\left(\Delta_{n-1}^{\intercal} z_n z_n^{\intercal}\Delta_{n-1}\right) \\ &= \sum_{n=n_0}^{N}\frac{1}{n}\mathrm{tr}\left\{\Delta_{n-1}^{\intercal} z_n z_n^{\intercal}\Delta_{n-1}\right\} = \mathrm{tr}\left\{\sum_{n=n_0}^{N}\left(\frac{1}{n}\Delta_{n-1}\Delta_{n-1}^{\intercal}\right)\left(z_n z_n^{\intercal}\right)\right\}\end{aligned}$$

$$\operatorname{tr}\left\{\Delta_{N-1}\Delta_{N-1}^{\intercal}\frac{1}{N}\sum_{n=n_0}^{N} z_n z_n^{\intercal} - \sum_{n=n_0}^{N}\left[\left(\frac{1}{n}\Delta_{n-1}\Delta_{n-1}^{\intercal}\right)-\left(\frac{1}{n-1}\Delta_{n-2}\Delta_{n-2}^{\intercal}\right)\right]\sum_{s=n_0}^{n-1}\left(z_s z_s^{\intercal}\right)\right\}$$

Taking into account that by the assumption *1*.

$$\frac{1}{N}\sum_{n=n_0}^{N} z_n z_n^{\intercal} \underset{N\to\infty}{\overset{a.s.}{\to}} \mathcal{R}$$

for large enough n_0 we get

$$\begin{aligned}
\sum_{n=n_0}^{N}\frac{1}{n}\left(\Delta_{n-1}^{\intercal}z_n\right)^2 &= \operatorname{tr}\left\{\Delta_{N-1}\Delta_{N-1}^{\intercal}\left(\mathcal{R}+o_\omega(1)\right) - \sum_{n=n_0}^{N}\left[\left(\frac{n-1}{n}\Delta_{n-1}\Delta_{n-1}^{\intercal}\right)-\left(\Delta_{n-2}\Delta_{n-2}^{\intercal}\right)\right]\frac{1}{n-1}\sum_{s=n_0}^{n-1}\left(z_s z_s^{\intercal}\right)\right\} \\
&= \operatorname{tr}\left\{\|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1) - \sum_{n=n_0}^{N}\left[\left(\frac{n-1}{n}\Delta_{n-1}\Delta_{n-1}^{\intercal}\right)-\left(\Delta_{n-2}\Delta_{n-2}^{\intercal}\right)\right]\left(\mathcal{R}+o_\omega(1)\right)\right\} \\
&= \|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1) - \sum_{n=n_0}^{N}\left[\frac{n-1}{n}\left(\|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1)\right)-\left(\|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1)\right)\right] \\
&= \|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1) + \sum_{n=n_0}^{N}\frac{1}{n}\left(\|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1)\right)
\end{aligned}$$

So, for $N\to\infty$ we have

$$\sum_{n=n_0}^{\infty}\frac{1}{n}\left(\Delta_{n-1}^{\intercal}z_n\right)^2 = w + \sum_{n=n_0}^{\infty}\frac{1}{n}\left(\|\Delta_n\|_{\mathcal{R}}^2 + o_\omega(1)\right) \overset{a.s.}{<} \infty \tag{13.114}$$

Since $\sum_{n=n_0}^{\infty}\frac{1}{n}=\infty$, from the last estimation it follows that there exists a subsequence $\left\{\|\Delta_{n_k}\|_{\mathcal{R}}^2\right\}_{k\geq 1}$ which converges to zero. But $\left\{\|\Delta_n\|_{\mathcal{R}}^2\right\}_{n\geq 1}$ converges itself, and therefore all its subsequences have the same limit which implies that $w=0$. Theorem is proven. □

One can represent recursion (13.110), corresponding to procedure (13.105), as follows:

$$\Delta_{n+1} = \Delta_n + \frac{1}{n+1}\mathcal{R}^{-1}z_{n+1}\left[\psi\left(-\Delta_n^{\intercal}z_{n+1}\right)+\zeta_n\right] \tag{13.115}$$

where

$$\zeta_n := \varphi\left(-\Delta_n^{\intercal} z_{n+1} + \xi_n\right) - \psi\left(-\Delta_n^{\intercal} z_{n+1}\right)$$

is a martingale-difference (see Definition 7.5) since

$$\mathsf{E}\left\{\zeta_n / \mathcal{F}_n\right\} \overset{a.s.}{=} 0$$

If, additionally, we suppose that $\psi'(x)$ is differentiable in $x = 0$ and $\psi'(0) > 0$, taking into account that $\Delta_n \underset{n\to\infty}{\overset{a.s.}{\rightarrow}} 0$, from (13.115) we get

$$\begin{aligned}\Delta_{n+1} &= \Delta_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\left[-\psi'(0)\,\Delta_n^{\intercal} z_{n+1} + o_\omega(1) + \zeta_n\right] \\ &= \Delta_n - \frac{\psi'(0)}{n+1}\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal}\Delta_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\left(o_\omega(1) + \zeta_n\right)\end{aligned}$$

The Lemma 13.7 below shows that two processes

$$\Delta_{n+1} = \Delta_n - \frac{\psi'(0)}{n+1}\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal}\Delta_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\left(o_\omega(1) + \zeta_n\right) \tag{13.116}$$

and

$$\tilde{\Delta}_{n+1} = \left(1 - \frac{\psi'(0)}{n+1}\right)\tilde{\Delta}_n + \frac{1}{n+1}\mathcal{R}^{-1} z_{n+1}\left(o_\omega(1) + \zeta_n\right) \tag{13.117}$$

are $\sqrt{n}$*-equivalent*, that is,

$$\sqrt{n}\left(\Delta_n - \tilde{\Delta}_n\right) \underset{n\to\infty}{\overset{a.s.}{\rightarrow}} 0 \tag{13.118}$$

Lemma 13.7. *Under the conditions of Theorem 13.5 and if $\psi'(x)$ is differentiable in $x = 0$ and $\psi'(0) > 0$, two processes (13.116) and (13.117) are $\sqrt{n}$-equivalent in the sense of (13.118).*

Proof. For $\Theta_n := \sqrt{n}\left(\Delta_n - \tilde{\Delta}_n\right)$ applying the representation

$$\sqrt{1 + \frac{1}{n}} = 1 + \frac{1 + o(1)}{2n}$$

one has

$$\begin{aligned}\Theta_{n+1} &= \sqrt{n+1}\left[\left(I - \frac{\psi'(0)}{n+1}\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal}\right)\Delta_n - \left(1 - \frac{\psi'(0)}{n+1}\right)\tilde{\Delta}_n\right] \\ &= \sqrt{n+1}\left[\left(\left(1 - \frac{\psi'(0)}{n+1}\right) - \frac{\psi'(0)}{n+1}\left[\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal} - I\right]\right)\Delta_n \right. \\ &\qquad \left. - \left(1 - \frac{\psi'(0)}{n+1}\right)\tilde{\Delta}_n\right]\end{aligned}$$

$$= \frac{\sqrt{n+1}}{\sqrt{n}} \left[\left(1 - \frac{\psi'(0)}{n+1} \right) \Theta_n - \frac{\psi'(0)}{n+1} \left[\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal} - I \right] \Delta_n \right]$$
$$= \left(1 - \frac{\psi'(0)}{\sqrt{n(n+1)}} \right) \Theta_n - \frac{\psi'(0)}{\sqrt{n(n+1)}} \left[\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal} - I \right] \Delta_n$$
$$= \left(1 - \frac{\psi'(0) + o(1)}{n} \right) \Theta_n - \frac{\psi'(0) + o(1)}{n} \left[\mathcal{R}^{-1} z_{n+1} z_{n+1}^{\intercal} - I \right] \Delta_n$$

This implies

$$\|\Theta_{n+1}\| \leq \left[1 - \frac{\psi'(0) + o(1)}{n} \right] \|\Theta_n\| - \frac{\psi'(0) + o(1)}{n} \left(\left\| \mathcal{R}^{-1} \right\| \|z_{n+1}\|^2 \Delta_n + 1 \right) \|\Delta_n\|$$

By property (13.114) the series of the second terms converges with probability one. So, by Lemma 7.9, it follows $\|\Theta_n\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. Lemma is proven. □

Now we are ready to formulate the main result concerning the rate of convergence using the asymptotic normality property for the normalized (by $\sqrt{n}$) identification error.

Theorem 13.6. (on asymptotic normality) *Let the conditions of Lemma 13.7 hold, and additionally,*

$$\boxed{\psi(0) = 0, \quad S(0) > 0 \quad \text{and} \quad 2\psi'(0) > 1} \tag{13.119}$$

Then

$$\boxed{\sqrt{n}(c_n - c) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, V)} \tag{13.120}$$

with

$$\boxed{V = \frac{S(0)}{2\psi'(0) - 1} \mathcal{R}^{-1}} \tag{13.121}$$

Proof. By Lemma 13.7 it is sufficient to prove

$$\sqrt{n}\tilde{\Delta}_n \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, V)$$

Rewrite (13.117) as

$$\tilde{\Delta}_n = \left(1 - \frac{\psi'(0)}{n} \right) \tilde{\Delta}_{n-1} + \frac{1}{n} \mathcal{R}^{-1} z_n (o_\omega(1) + \zeta_{n-1})$$
$$= \cdots = \pi_n \tilde{\Delta}_1 + \sum_{t=1}^{n} \pi_n \pi_t^{-1} \frac{1}{t} \mathcal{R}^{-1} z_t (o_\omega(1) + \zeta_{t-1})$$

with

$$\pi_n := \prod_{t=1}^{n} \left(1 - \frac{\psi'(0)}{t}\right) = \exp \sum_{t=1}^{n} \ln \left(1 - \frac{\psi'(0)}{t}\right)$$
$$\overset{1+x \le \exp x}{\le} \exp \sum_{t=1}^{n} \ln \exp \left(-\frac{\psi'(0)}{t}\right) = \exp \left(-\psi'(0) \sum_{t=1}^{n} \frac{1}{t}\right)$$
$$\le \exp\left(-\psi'(0) \ln n\right) = n^{-\psi'(0)}$$

We also have that

$$\sqrt{n}\pi_n \le n^{-\psi'(0)+1/2} \underset{n\to\infty}{\to} 0 \tag{13.122}$$

which says that $\left\{\tilde{\Delta}_n\right\}_{n\ge1}$ is $\sqrt{n}$-equivalent to the process $\{z_n\}_{n\ge1}$

$$z_n := \pi_n \sum_{t=1}^{n} \frac{1}{t} \mathcal{R}^{-1} z_t \left(o_\omega(1) + \zeta_{t-1}\right) = \frac{1}{n} A_n^{-1} \sum_{k=1}^{n} A_k \zeta_k$$

with

$$A_k^{-1} := k\pi_k, \quad \zeta_k := \mathcal{R}^{-1} z_k \left(o_\omega(1) + \zeta_{k-1}\right)$$

Applying then Theorem 8.15 for $y_n := \sqrt{n} z_n$ we get

$$y_n := \frac{1}{\sqrt{n}} A_n^{-1} \sum_{k=1}^{n} A_k \zeta_k \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, K)$$

with $V = \lim_{n\to\infty} \mathsf{E}\left\{y_n y_n^{\intercal}\right\}$ which (by Lemma 8.11) is the solution of the following Lyapunov matrix equation

$$\left(A - \frac{1}{2}I\right) V + V \left(A - \frac{1}{2}I\right)^{\intercal} = -R \tag{13.123}$$

where A and R are defined as

$$A := \lim_{n\to\infty} n\left(A_n^{-1}A_{n-1} - I\right) = \lim_{n\to\infty} n\left(\frac{n}{n-1}\pi_n\pi_{n-1}^{-1} - 1\right)$$
$$= \lim_{n\to\infty} n\left(\frac{n}{n-1}\left[1 - \frac{\psi'(0)}{n}\right] - 1\right)$$
$$= \lim_{n\to\infty} n\left(\left[1 + \frac{1+o(1)}{n}\right]\left[1 - \frac{\psi'(0)}{n}\right] - 1\right)$$
$$= \lim_{n\to\infty} n\left(\frac{1 - \psi'(0) + o(1)}{n}\right) = 1 - \psi'(0)$$

and

$$
\begin{aligned}
R &:= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}R_t = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t}\mathsf{E}\left\{\zeta_t\zeta_s^{\mathsf{T}}\right\} = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{\zeta_t\zeta_t^{\mathsf{T}}\right\} \\
&= \mathcal{R}^{-1}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{z_t\left(o_\omega(1)+\zeta_{t-1}\right)z_t^{\mathsf{T}}\left(o_\omega(1)+\zeta_{t-1}\right)\right\}\mathcal{R}^{-1} \\
&= \mathcal{R}^{-1}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{z_t z_t^{\mathsf{T}}\zeta_{t-1}^{2}\right\}\mathcal{R}^{-1} = \mathcal{R}^{-1}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{\zeta_{t-1}^{2}\right\} \\
&= \mathcal{R}^{-1}\mathsf{E}\left\{\varphi^2(\xi_1)\right\} = \mathcal{R}^{-1}S(0)
\end{aligned}
$$

since

$$
\begin{aligned}
&\zeta_n = \varphi\left(-\Delta_n^{\mathsf{T}}z_{n+1}+\xi_n\right)-\psi\left(-\Delta_n^{\mathsf{T}}z_{n+1}\right) \\
&\left|\zeta_n^2-\varphi^2(\xi_n)\right| \underset{n\to\infty}{\overset{a.s.}{\to}} 0
\end{aligned}
$$

Notice that both conditions of Lemma 8.11 are fulfilled since

$$
A-\frac{1}{2}I = 1-\psi'(0)-\frac{1}{2} = \frac{1}{2}-\psi'(0) < 0
$$

and

$$
R = \mathcal{R}^{-1}S(0) > 0
$$

Substitution these relations into (13.123) leads to

$$
\left(A-\frac{1}{2}I\right)V + V\left(A-\frac{1}{2}I\right)^{\mathsf{T}} = 2\left[\frac{1}{2}-\psi'(0)\right]V = -R = -\mathcal{R}^{-1}S(0)
$$

which proves (13.121). Theorem is proven. □

Remark 13.6. *Theorem 13.6 above shows that the procedure (13.104) with nonlinear residual transformation $\varphi(\cdot)$ provides the* ***asymptotic equivalence*** *(on distribution) of the normalized identification error $\sqrt{n}\Delta_n$ to a Gaussian centered vector with the covariance matrix V (13.121), i.e.,*

$$
\boxed{\sqrt{n}\Delta_n \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,V)}
$$

or equivalently,

$$
\boxed{\Delta_n \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0,n^{-1}V\right)} \tag{13.124}
$$

which shows that asymptotically the ***rate of convergence*** *of the identification procedure (13.104) is $n^{-1}V$.*

13.6.3 The best selection of nonlinear residual transformation

13.6.3.1 Low bound for the normalized error covariation matrix

Lemma 13.8. *Under any regular noise distribution p_ξ (when $I_F\left(p_\xi\right) > 0$) and for any nonlinear residual transformation φ, satisfying the conditions (13.119), the following inequalities hold:*

$$\boxed{\frac{S(0)}{2\psi'(0)-1} \geq \frac{S(0)}{[\psi'(0)]^2} \geq I_F^{-1}\left(p_\xi\right)} \tag{13.125}$$

Proof. The first inequality results from evident relations

$$0 \leq \left[1-\psi'(0)\right]^2 = 1-2\psi'(0)+\left[\psi'(0)\right]^2$$
$$\left[\psi'(0)\right]^2 \geq 2\psi'(0)-1$$

The second one follows directly from the Cauchy–Schwartz matrix stochastic inequality (13.79)

$$\mathsf{E}\left\{\xi\xi^{\mathsf{T}}\right\} \geq \mathsf{E}\left\{\xi\eta^{\mathsf{T}}\right\}\left[\mathsf{E}\left\{\eta\eta^{\mathsf{T}}\right\}\right]^{-1}\mathsf{E}\left\{\eta\xi^{\mathsf{T}}\right\}$$

if we put there

$$\xi := \varphi\left(\xi_1\right), \quad \eta := \left(\ln p_{\xi_1}\left(\xi_1\right)\right)'$$

and use the property

$$\begin{aligned}
\mathsf{E}\left\{\xi\eta^{\mathsf{T}}\right\} &= \mathsf{E}\left\{\varphi\left(\xi_1\right)\left(\ln p_{\xi_1}\left(\xi_1\right)\right)'\right\} \\
&= \int_{v\in\mathbb{R}} \varphi(v)\left(\ln p_{\xi_1}(v)\right)' p_{\xi_1}(v)\,dv = \int_{v\in\mathbb{R}} \varphi(v)\,p'_{\xi_1}(v)\,dx \\
&= -\int_{v\in\mathbb{R}} \varphi'(v)\,p_{\xi_1}(v)\,dx = -\left[\frac{d}{dx}\int_{v\in\mathbb{R}} \varphi(x+v)\,p_{\xi_1}(v)\,dx\right]_{x=0} \\
&= -\left[\frac{d}{dx}\psi(x)\right]_{x=0} = -\psi'(0)
\end{aligned}$$

Lemma is proven. □

Corollary 13.7. *Any admissible selection of nonlinear transformation φ in (13.104) can not provide the covariation V (13.121) of the normalized identification error less than $\left[I_F\left(p_\xi\right)\mathcal{R}\right]^{-1}$, that is,*

$$\boxed{V = \frac{S(0)}{2\psi'(0)-1}\mathcal{R}^{-1} \geq \left[I_F\left(p_\xi\right)\mathcal{R}\right]^{-1}} \tag{13.126}$$

13.6.3.2 Recurrent maximum likelihood algorithm

Select in (13.104) or in (13.105)

$$\boxed{\varphi(v) = \varphi^*(v) := -\mathbb{I}_F^{-1}\left(p_\xi\right)\frac{d}{dv}\ln p_\xi(v)} \tag{13.127}$$

(in this case $\mathbb{I}_F^{-1}\left(p_\xi\right) = I_F^{-1}\left(p_\xi\right)$ is a scalar). Then the algorithm (13.104) becomes

$$\boxed{\begin{aligned} c_n &= c_{n-1} - \Gamma_n I_F^{-1}\left(p_\xi\right) z_n \frac{d}{dv}\ln p_\xi\left(x_{n+1} - c_{n-1}^\top z_n\right) \\ \Gamma_{n+1} &= \Gamma_n - \frac{\Gamma_n z_{n+1} z_{n+1}^\top \Gamma_n}{1 + z_{n+1}^\top \Gamma_n z_{n+1}}, \quad n \geq n_0 \\ c_{n_0} &= Z_{n_0}^{-1} V_{n_0}, \quad \Gamma_{n_0} = \left(\sum_{t=0}^{n_0} z_t z_t^\top\right)^{-1} = Z_{n_0}^{-1} \end{aligned}} \tag{13.128}$$

It is called the *recurrent version* of the *maximum likelihood estimating procedure (see Hall and Heyde (1976) and Devyaterikov and Poznyak (1985)).*

Theorem 13.7. *If for the scalar model (13.15) the conditions of Theorem 13.4 hold, then procedure (13.128) is* ***asymptotically efficient*** *(****the best one****) under any regular (not only Gaussian)* ***i.i.d. noise*** *in the dynamics of the system.*

Proof. As we already mentioned above, this procedure is asymptotically equivalent to the following one:

$$c_n = c_{n-1} - \frac{1}{n}\mathcal{R}^{-1} I_F^{-1}\left(p_\xi\right) z_n \frac{d}{dv}\ln p_\xi\left(x_{n+1} - c_{n-1}^\top z_n\right) \tag{13.129}$$

for which we have

$$S(0) = \mathsf{E}\left\{\varphi^2(\xi_1)\right\} = \mathsf{E}\left\{\left[-I_F^{-1}\left(p_\xi\right)\left(\ln p_{\xi_1}(\xi_1)\right)'\right]^2\right\} = I_F^{-1}\left(p_\xi\right)$$

and

$$\begin{aligned} \psi'(0) &= \left[-\frac{d}{dx}\int_{v\in\mathbb{R}} I_F^{-1}\left(p_\xi\right)\varphi(x+v)\, p_{\xi_1}(v)\, dx\right]_{x=0} \\ &= -I_F^{-1}\left(p_\xi\right)\left[\frac{d}{dx}\int_{v\in\mathbb{R}}\left(\ln p_{\xi_1}(x+v)\right)' p_{\xi_1}(v)\, dx\right]_{x=0} \\ &= -I_F^{-1}\left(p_\xi\right)\left[-\frac{d}{dx}\int_{v\in\mathbb{R}}\left(\ln p_{\xi_1}(x+v)\right) p'_{\xi_1}(v)\, dx\right]_{x=0} \end{aligned}$$

$$= -I_F^{-1}\left(p_\xi\right)\left[-\int\limits_{v\in\mathbb{R}} \frac{p'_{\xi_1}(x+v)}{p_{\xi_1}(x+v)} p'_{\xi_1}(v)\,dx\right]_{x=0} = 1$$

Therefore

$$V = \frac{S(0)}{2\psi'(0)-1}\mathcal{R}^{-1} = \frac{I_F^{-1}\left(p_\xi\right)}{2\cdot 1-1}\mathcal{R}^{-1} = \left[I_F\left(p_\xi\right)\mathcal{R}\right]^{-1}$$

which proves the theorem. □

13.6.4 Asymptotically efficient procedure under correlated noise

If the model is given

$$\boxed{\begin{array}{c} x_{n+1} = Cz_n + D\zeta_n \\ C := \begin{bmatrix} A & B \end{bmatrix} \in R^{N\times(N+K)}, \quad z_n := \begin{bmatrix} x_n \\ u_n \end{bmatrix} \end{array}} \tag{13.130}$$

where ζ_n is the colored noise generated by a forming filter (13.57) as

$$\boxed{\begin{array}{c} \zeta_n = H(q)\,\xi_n = \dfrac{P(q)}{Q(q)}\xi_n \\ P(q) := \displaystyle\sum_{s=0}^{k_P} f_s q^s, \quad Q(q) := \sum_{s=0}^{k_Q} g_s q^s \\ Q(q) \text{ is stable, } qz_n = z_{n-1} \text{ is the delay operator} \end{array}} \tag{13.131}$$

then the following result holds.

Theorem 13.8. *For any stable scalar ($N = 1$) model (13.15) the recurrent LSM-procedure (13.29) with nonlinear residual transformation applied to the whited signals, that is,*

$$\boxed{\begin{array}{c} c_{n+1} = c_n - \Gamma_{n+1}\tilde{z}_{n+1}\mathbb{I}_F^{-1}\left(p_\xi\right)\dfrac{d}{dv}\ln p_\xi\left(\tilde{x}_{n+2} - c_n^\intercal \tilde{z}_{n+1}\right) \\ \Gamma_{n+1} = \Gamma_n - \dfrac{\Gamma_n \tilde{z}_{n+1}\tilde{z}_{n+1}^\intercal \Gamma_n}{1+\tilde{z}_{n+1}^\intercal \Gamma_n \tilde{z}_{n+1}}, \quad n \geq n_0 \\ c_{n_0} = \tilde{Z}_{n_0}^{-1} V_{n_0}, \quad \Gamma_{n_0} = \left(\displaystyle\sum_{t=0}^{n_0} \tilde{z}_t \tilde{z}_t^\intercal\right)^{-1} = \tilde{Z}_{n_0}^{-1} \\ \tilde{z}_n = H^{-1}(q)\,z_n, \quad \tilde{x}_n = H^{-1}(q)\,x_n \end{array}} \tag{13.132}$$

*is **asymptotically efficient** (**the best one**) under any regular (not only Gaussian) **colored noise** generated by the forming **stable** ($Q(q)$ is stable) and **minimal phase** ($P(q)$ is stable) filter (13.131).*

Proof. Applying the '*whitening operation*' to both sides of (13.130) for $N = 1$ one gets

$$H^{-1}(q) : x_{n+1} = c^{\mathsf{T}} z_n + H(q)\xi_n$$

or, equivalently $\left(\text{up to terms } O_{\omega}\left(e^{-\alpha n}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0\right)$,

$$H^{-1}(q)x_{n+1} = \tilde{x}_{n+1} = c^{\mathsf{T}} H^{-1}(q)z_n + \xi_n = c^{\mathsf{T}} \tilde{z}_n + \xi_n$$

Then the desired result follows directly from Theorem 13.7. □

Finally, we may conclude that the *joint procedure* (13.132) realizing the *recurrent maximum likelihood* algorithm together with parallel '*whitening*' is asymptotically effective (the best one) for correlated noise generated by a stable, minimum-phase filter under an i.i.d. input noise with the density function p_ξ.

13.7 Robustification of identification procedures

13.7.1 The Huber robustness

The numerical estimating procedure (13.128) is asymptotically efficient only if the exact information of a noise density distribution function p_ξ is available. This assumption practically never can be realized exactly since it is, in fact, a statistical one by its nature and can be ensured only with some '*level of admissibility*'. This exactly means that in practice we deal not with an exact noise density p_ξ, but with a class of possible noises and their densities; namely, we are dealing with a class of noises, the corresponding densities of which belong to some class $\mathcal{P}$, i.e.,

$$\boxed{p_\xi \in \mathcal{P}} \tag{13.133}$$

As has been mentioned above, the recurrent estimation procedure (13.104) with nonlinear residual transformation φ provides the convergence rate equal (in a 'distribution sense') to

$$\boxed{\begin{aligned} &\Delta_n \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, n^{-1}V\right) \\ &V = \frac{S(0)}{2\psi'(0) - 1}\mathcal{R}^{-1} \end{aligned}} \tag{13.134}$$

that is,

- the order of the convergence is $1/n$, and
- the constant (in fact, the matrix) of convergence is V.

As follows from (13.134), V depends on a real noise density distribution p_ξ (since $S(0)$, $\psi'(0)$ and $\mathcal{R}$ may be dependent on p_ξ) and on a nonlinear function φ (through $S(0)$ and $\psi'(0)$). That's why, to emphasize this dependence, we will use the notation

$$\boxed{V = V\left(p_\xi, \varphi\right)}$$

Following Huber (1975) and Tsypkin (1984), let us introduce the main definition of this section.

Definition 13.7. *The pair of functions* (p_ξ^*, φ^*) *defines the estimating procedure (13.128) with the nonlinear residual transformation* φ^* *which is* ***robust*** *with respect to a distribution* p_ξ*, belonging to a class* $\mathcal{P}$*, if for any admissible* φ*, satisfying the assumptions of Theorem 13.6, and any noise distribution* $p_\xi \in \mathcal{P}$ *the following 'saddle-point' inequalities hold:*

$$\boxed{V\left(p_\xi, \varphi^*\right) \leq V\left(p_\xi^*, \varphi^*\right) \leq V\left(\varphi, p_\xi^*\right)} \tag{13.135}$$

Here both inequalities should be treated in a 'matrix sense', that is,

$$A = A^\intercal \leq B = B^\intercal \quad \textit{if } B - A \geq 0$$

In other words,

- the distribution p_ξ^* is the '*worst*' within the class $\mathcal{P}$, and
- the nonlinear transformation φ^* is '*the best one*' oriented on the '*worst*' noise with the distribution p_ξ^*.

This can be expressed mathematically as follows:

$$\boxed{\begin{aligned} \varphi^* &:= \arg\inf_{\varphi} \sup_{p_\xi \in \mathcal{P}} V\left(p_\xi, \varphi\right) \\ p_\xi^* &:= \arg \sup_{p_\xi \in \mathcal{P}} \inf_{\varphi} V\left(p_\xi, \varphi\right) \end{aligned}} \tag{13.136}$$

so that

$$\boxed{\inf_{\varphi} \sup_{p_\xi \in \mathcal{P}} V\left(p_\xi, \varphi\right) = \sup_{p_\xi \in \mathcal{P}} \inf_{\varphi} V\left(p_\xi, \varphi\right) := V^*} \tag{13.137}$$

According to (13.126), for any fixed $p_\xi \in \mathcal{P}$

$$\inf_{\varphi} V\left(p_\xi, \varphi\right) = \left[I_F\left(p_\xi\right)\mathcal{R}\right]^{-1} \tag{13.138}$$

where $\inf\limits_{\varphi}$ is achieved for

$$\varphi(v) = \varphi^*(v) := -I_F^{-1}\left(p_\xi\right)\frac{d}{dv}\ln p_\xi(v) \tag{13.139}$$

So, finally, the *robust procedure designing* is reduced to the solution of the problem:

$$\boxed{\sup_{p_\xi \in \mathcal{P}} \inf_{\varphi} V\left(p_\xi, \varphi\right) = \sup_{p_\xi \in \mathcal{P}} \left[I_F\left(p_\xi\right)\mathcal{R}\right]^{-1}} \tag{13.140}$$

Below we will consider in more detail the robust procedure designing for two basic models: the static *regression* (R-model) and dynamic *autoregression* model (AR-model).

13.7.2 Robust identification of static (regression) models

For static models (R-models) where the generalized inputs are independent of the state of the system, the matrix $\mathcal{R}$ does not depend on p_ξ, and therefore, the problem (13.140) is reduced to the following one:

$$\boxed{\sup_{p_\xi \in \mathcal{P}} \left[I_F \left(p_\xi \right) \right]^{-1} \quad \text{or} \quad \inf_{p_\xi \in \mathcal{P}} I_F \left(p_\xi \right)} \tag{13.141}$$

Consider now several most significant classes $\mathcal{P}$ of *a priori* informative noise distributions and solutions of the problem (13.141) within these classes.

Class $\mathcal{P}_1$ (*of all symmetric distributions nonsingular in the point* $x = 0$):

$$\boxed{\mathcal{P}_1 := \left\{ p_\xi : p_\xi (0) \geq \frac{1}{2a} > 0 \right\}} \tag{13.142}$$

We deal with this class if there is no *a priori* information on noise distribution.

Lemma 13.9. (on class $\mathcal{P}_1$)

$$\boxed{p_\xi^* (x) = \arg \inf_{p_\xi \in \mathcal{P}_1} I_F \left(p_\xi \right) = \frac{1}{2a} \exp \left\{ -\frac{|x|}{a} \right\}} \tag{13.143}$$

that is, the value for $\mathcal{P}_1$ distribution density is the Laplace one given by (13.143).

Proof. By the Cauchy–Schwarz inequality

$$\left(\int_{\mathbb{R}} f \varphi p_\xi dx \right)^2 \leq \left(\int_{\mathbb{R}} f^2 p_\xi dx \right) \left(\int_{\mathbb{R}} \varphi^2 p_\xi dx \right) \tag{13.144}$$

valid for any f, φ, and any noise distribution density p_ξ (for which the integrals have a sense), for $f := p_\xi'(x)/p_\xi(x)$ after integrating by parts it follows that

$$I_F(p_\xi) \geq \left(\int_{\mathbb{R}} p_\xi (x) \, d\varphi (x) \right)^2 \Big/ \int_{\mathbb{R}} \varphi^2 (x) \, p_\xi (x) \, dx \tag{13.145}$$

where the equality is attained when

$$p_\xi'(x)/p_\xi(x) = \lambda \varphi (x), \quad \lambda \text{ is any constant}$$

Taking $\varphi (x) := \text{sign}(x)$ in (13.145) and using the identity

$$\left[\text{sign}(x) \right]' = 2\delta (x) \, p_\xi (0)$$

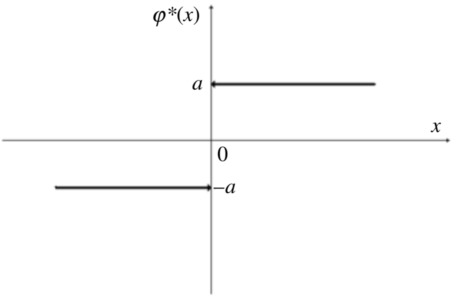

Fig. 13.2. The nonlinear transformation φ^* for the class $\mathcal{P}_1$.

leads to

$$I_F(p_\xi) \geq 4p_\xi^2(0) \geq \frac{1}{a^2} \quad \text{for any} \quad p_\xi \in \mathcal{P}_1$$

where the equality is attained when

$$p'_\xi(x)/p_\xi(x) = \lambda \mathrm{sign}(x)$$

or equivalently, for

$$p_\xi(x) = \frac{\lambda}{2} \exp\{-|x|/\lambda\}$$

For $\lambda = a$ we have

$$p_\xi(x) = \frac{a}{2} \exp\{-|x|/a\} = p_\xi^*(x)$$

and in view of (13.88)

$$I_F(p_\xi) \geq 4p_\xi^2(0) \geq \frac{1}{a^2} = I_F(p_\xi^*)$$

So, the noise distribution value within $\mathcal{P}_1$ is $p_\xi^*(x)$. Lemma is proven. □

Corollary 13.8. *The robustness of* $\mathcal{P}_1$ *version of the procedure (13.128) contains*

$$\boxed{\varphi(x) = \varphi^*(x) = -I_F^{-1}\left(p_\xi^*\right)\frac{d}{dv}\ln p_\xi^*(v) = a\ \mathrm{sign}(x)} \tag{13.146}$$

(see Fig. 13.2).

Class $\mathcal{P}_2$ (*of all symmetric distributions with a bounded variance*):

$$\boxed{\mathcal{P}_2 := \left\{p_\xi : \int_{\mathbb{R}} x^2 p_\xi(x)\,ds \leq \sigma^2 < \infty\right\}} \tag{13.147}$$

Lemma 13.10. (on class $\mathcal{P}_2$)

$$\boxed{p_\xi^*(x) = \arg\inf_{p_\xi \in \mathcal{P}_2} I_F\left(p_\xi\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}} \tag{13.148}$$

that is, the $\mathcal{P}_2$ distribution density value is the Gaussian one given by (13.148).

Proof. Taking in (13.145) $\varphi(x) = x$ for all $p_\xi \in \mathcal{P}_2$ we get

$$I_F(p_\xi) \geq 1 \bigg/ \int_{\mathbb{R}} x^2 p_\xi(x)\,dx \geq 1/\sigma^2$$

where the equality is attained when

$$p'_\xi(x)/p_\xi(x) = \lambda x, \quad \lambda \text{ is any constant}$$

or equivalently, for

$$p_\xi(x) = \frac{1}{\sqrt{2\pi/\lambda}}\exp\left\{-\frac{\lambda x^2}{2}\right\}$$

For $\lambda = \sigma^{-2}$ we have

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2}{2\sigma^2}\right\} = p_\xi^*(x)$$

and in view of (13.87)

$$I_F(p_\xi) \geq 1 \bigg/ \int_{\mathbb{R}} x^2 p_\xi(x)\,dx \geq 1/\sigma^2 = I_F(p_\xi^*)$$

So, the noise distribution value within $\mathcal{P}_1$ is $p_\xi^*(x)$. Lemma is proven. □

Corollary 13.9. *The robustness of $\mathcal{P}_2$ version of the procedure (13.128) contains*

$$\boxed{\varphi(x) = \varphi^*(x) = -I_F^{-1}\left(p_\xi^*\right)\frac{d}{dv}\ln p_\xi^*(v) = x} \tag{13.149}$$

(see Fig. 13.3) which means that the standard LSM algorithm (13.29) is the robustness within the class $\mathcal{P}_2$.

Class $\mathcal{P}_3$ (*of all symmetric 'approximately normal' distributions*):

$$\boxed{\mathcal{P}_2 := \left\{p_\xi : p_\xi(x) = (1-\alpha)\,p_{\mathcal{N}(0,\sigma^2)}(x) + \alpha q(x)\right\}} \tag{13.150}$$

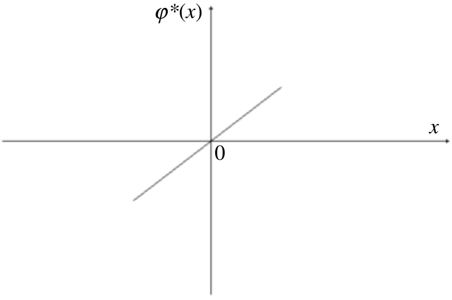

Fig. 13.3. The 'nonlinear' transformation φ^* for the class $\mathcal{P}_2$.

where $p_{\mathcal{N}(0,\sigma)}(x)$ is the centered Gaussian distribution density with variance σ^2 and $q(x)$ is another distribution density. The parameter $\alpha \in [0, 1]$ characterizes the level of the effect of a '*dirtying*' distribution $q(x)$ to the basic one $p_{\mathcal{N}(0,\sigma)}(x)$.

Lemma 13.11. (on class $\mathcal{P}_3$)

$$\boxed{\begin{aligned} p_\xi^*(x) &= \arg \inf_{p_\xi \in \mathcal{P}_3} I_F\left(p_\xi\right) \\ &= \begin{cases} \dfrac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{x^2}{2\sigma^2}\right\} \textit{ for } |x| \le \Delta \\ \dfrac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{\Delta|x|}{\sigma^2} + \dfrac{\Delta^2}{2\sigma^2}\right\} \textit{ for } |x| > \Delta \end{cases} \end{aligned}} \tag{13.151}$$

where Δ is a solution of the transcendent equation

$$\boxed{\frac{1}{1-\alpha} = \int_{-\Delta}^{\Delta} p_{N(0,\sigma)}(x)\,dx + 2p_{N(0,\sigma)}(\Delta)\frac{\sigma^2}{\Delta}} \tag{13.152}$$

that is, the value of $\mathcal{P}_3$ distribution density is the Gaussian one for $|x| \le \Delta$ and the Laplace type for $|x| > \Delta$.

Proof. **(without details)** From (13.150) it follows that

$$p_\xi(x) \ge (1-\alpha)\, p_{\mathcal{N}(0,\sigma^2)}(x)$$

So, we need to solve the following variational problem:

$$\inf_{p_\xi : p_\xi \ge (1-\alpha) p_{N(0,\sigma^2)}} I_F\left(p_\xi\right)$$

As is shown in Tsypkin and Polyak (1977), its solution is (13.151). □

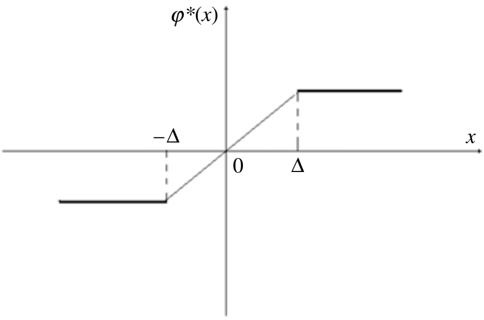

Fig. 13.4. The nonlinear transformation φ^* for class $\mathcal{P}_3$.

Corollary 13.10. *The robustness of $\mathcal{P}_3$ version of the procedure (13.128) contains*

$$\boxed{\begin{aligned}\varphi(x) &= \varphi^*(x) = -I_F^{-1}\left(p_\xi^*\right)\frac{d}{dv}\ln p_\xi^*(v) \\ &= \begin{cases} x & \text{for } |x| \le \Delta \\ \Delta \operatorname{sign}(x) & \text{for } |x| > \Delta \end{cases}\end{aligned}} \tag{13.153}$$

(see Fig. 13.4).

Class $\mathcal{P}_4$ (*of all symmetric 'approximately uniform' distributions*):

$$\boxed{\mathcal{P}_4 := \left\{p_\xi : p_\xi(x) = (1-\alpha)\, p_{U(0,a)}(x) + \alpha q(x)\right\}} \tag{13.154}$$

where

$$p_{U(0,a)}(x) := \frac{1}{2a}\chi\left(|x| \le a\right)$$

is the centered uniform distribution and $q(x)$ is another distribution density. The parameter $\alpha \in [0, 1]$ characterizes the level of the effect of a '*dirtying*' distribution $q(x)$ to the basic one $p_{U(0,a)}(x)$.

Lemma 13.12. (on class $\mathcal{P}_4$)

$$\boxed{\begin{aligned}p_\xi^*(x) &= \arg\inf_{p_\xi \in \mathcal{P}_4} I_F\left(p_\xi\right) \\ &= \begin{cases} \dfrac{1-\alpha}{2a} & \text{for } |x| \le a \\ \dfrac{1-\alpha}{2a}\exp\left\{-(1-\alpha)\dfrac{|x|-a}{\alpha a}\right\} & \text{for } |x| > a \end{cases}\end{aligned}} \tag{13.155}$$

that is, the value of $\mathcal{P}_4$ distribution density is the uniform one for $|x| \le a$ and the Laplace type for $|x| > a$.

Proof. From (13.154) it follows that

$$p_\xi(\pm a) \geq \frac{(1-\alpha)}{2a}$$

Taking in (13.145)

$$\varphi(x) = \begin{matrix} 0 & \text{for } |x| \leq a \\ \text{sign}(x) & \text{for } |x| > a \end{matrix}$$

we get

$$I_F(p_\xi) \geq \left[p_\xi(-a) + p_\xi(a)\right]^2 \Bigg/ \left[\int_{-\infty}^{-a} p_\xi(x)\,dx + \int_a^{\infty} p_\xi(x)\,dx\right]$$
$$\geq \frac{(1-\alpha)^2}{\alpha a^2}$$

so that the equality is attained when

$$p'_\xi(x)/p_\xi(x) = \lambda\varphi(x), \quad \lambda \text{ is any constant}$$

or equivalently, for

$$p_\xi(x) = \begin{cases} \text{Const for } |x| \leq a \\ \lambda\dfrac{\alpha}{2}\exp\{-\lambda(|x|-a)\} \quad \text{for } |x| > a \end{cases}$$

For $\lambda = \frac{(1-\alpha)}{\alpha a}$ we have

$$p_\xi(x) = \begin{cases} \dfrac{1-\alpha}{2a} \quad \text{for } |x| \leq a \\ \dfrac{1-\alpha}{2a}\exp\left\{-(1-\alpha)\dfrac{|x|-a}{\alpha a}\right\} \quad \text{for } |x| > a \end{cases} = p^*_\xi(x)$$

and in view of (13.87)

$$I_F(p_\xi) \geq \left[p_\xi(-a) + p_\xi(a)\right]^2 \Bigg/ \left[\int_{-\infty}^{-a} p_\xi(x)\,dx + \int_a^{\infty} p_\xi(x)\,dx\right]$$
$$\geq \frac{(1-\alpha)^2}{\alpha a^2} = I_F(p^*_\xi)$$

So, the value of the noise distribution within $\mathcal{P}_1$ is $p^*_\xi(x)$. Lemma is proven. □

Corollary 13.11. *The robustness of $\mathcal{P}_4$ version of the procedure (13.128) contains*

$$\boxed{\begin{aligned} \varphi(x) = \varphi^*(x) &= -I_F^{-1}\left(p^*_\xi\right)\frac{d}{dv}\ln p^*_\xi(v) \\ &= \begin{cases} 0 \quad \textit{for } |x| \leq a \\ \dfrac{1-\alpha}{\alpha a}\text{sign}(x) \quad \textit{for } |x| > a \end{cases} \end{aligned}} \tag{13.156}$$

(see Fig. 13.5).

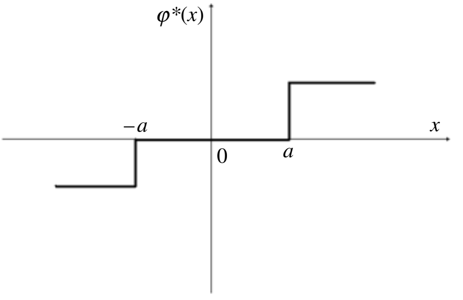

Fig. 13.5. The nonlinear transformation φ^* for class $\mathcal{P}_4$.

Some other classes of uncertainty are analyzed in Tsypkin (1984) and Devyaterikov and Poznyak (1985).

13.7.3 Robust identification of dynamic (autoregression) models

In the case of the dynamic *autoregression* model (AR-model) where the generalized inputs are dependent on the state of the system, the matrix $\mathcal{R}$ depends on p_ξ too, and therefore the problem we deal with is the complete problem (13.140), namely,

$$\sup_{p_\xi \in \mathcal{P}} \left[I_F \left(p_\xi \right) \mathcal{R} \right]^{-1} \tag{13.157}$$

For the AR-model (13.3)

$$y_{n+1} = \sum_{s=0}^{L_a} a_s y_{n-s} + \xi_n = \mathbf{a}^\top \mathbf{v}_n + \xi_n$$
$$\mathbf{a}^\top = \left(a_0, \ldots, a_{L_a} \right), \quad \mathbf{v}_n^\top = \left(y_n, y_{n-1}, \ldots, y_{n-L_a} \right)$$

we have

$$\frac{1}{n} \sum_{t=0}^{n} \mathsf{E} \left\{ \mathbf{v}_t \mathbf{v}_t^\top \right\} \to \mathcal{R}$$

where $\mathcal{R}$ satisfies (13.102), i.e.

$$\mathcal{R} = A \mathcal{R} A + \sigma^2 \Xi_0$$

with

$$A = \begin{Vmatrix} a_0 & a_1 & \cdots & \cdots & a_{L_a} \\ \mathbf{1} & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{1} & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & \mathbf{1} & 0 \end{Vmatrix}, \quad \Xi_0 := \begin{Vmatrix} \mathbf{1} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{Vmatrix}$$

Obviously, $\mathcal{R}$ can be represented as

$$\mathcal{R} = \sigma^2 \mathcal{R}_0$$

where $\mathcal{R}_0$ is the solution of

$$\mathcal{R} = A\mathcal{R}A + \Xi_0$$

so that the problem (13.157) is reduced to

$$\sup_{p_\xi \in \mathcal{P}} \left[\sigma^2\left(p_\xi\right) I_F\left(p_\xi\right)\right]^{-1}$$

or equivalently, to

$$\boxed{\inf_{p_\xi \in \mathcal{P}} \left[\sigma^2\left(p_\xi\right) I_F\left(p_\xi\right)\right]} \tag{13.158}$$

Consider now some classes $\mathcal{P}$ of *a priori* informative noise distributions and solutions of the problem (13.158) within these classes.

Class $\mathcal{P}_1^{AR}$ (*containing among others the Gaussian distribution* $p_{\mathcal{N}(0,\sigma_0^2)}(x)$).

Lemma 13.13. (on class $\mathcal{P}_1^{AR}$)

$$\boxed{p_\xi^*(x) = \arg \inf_{p_\xi \in \mathcal{P}_1^{AR}} \left[\sigma^2\left(p_\xi\right) I_F\left(p_\xi\right)\right] = p_{\mathcal{N}(0,\sigma_0^2)}(x)} \tag{13.159}$$

that is, the value of $\mathcal{P}_1^{AR}$ *distribution density is exactly the Gaussian distribution* $p_{\mathcal{N}(0,\sigma_0^2)}(x)$.

Proof. Taking in (13.144)

$$f(x) = x, \quad \varphi(x) = p'_\xi(x)/p_\xi(x)$$

we get

$$\sigma^2 I_F\left(p_\xi\right) \geq \left(\int_{\mathbb{R}} x p'_\xi(x)\,dx\right)^2 = \left(\int_{\mathbb{R}} p_\xi(x)\,dx\right)^2 = 1$$

such that the equality is attained when

$$p'_\xi(x)/p_\xi(x) = \lambda x$$

which leads to

$$p_\xi(x) = \frac{1}{\sqrt{2\pi/\lambda}} \exp\left\{-\frac{\lambda x^2}{2}\right\}$$

But since

$$I_F\left(p_{\mathcal{N}(0,\sigma_0^2)}\right) = \sigma_0^{-2}$$

from the inequality above we have

$$\sigma^2\left(p_\xi\right) I_F\left(p_\xi\right) \geq 1 = \sigma_0^2 I_F\left(p_{\mathcal{N}(0,\sigma_0^2)}\right)$$

which means that

$$p_\xi^*(x) = p_{\mathcal{N}(0,\sigma_0^2)}(x)$$

Lemma is proven. □

Corollary 13.12. *The robust on $\mathcal{P}_1^{AR}$ version of the procedure (13.128) contains*

$$\boxed{\varphi(x) = \varphi^*(x) = -I_F^{-1}\left(p_\xi^*\right)\frac{d}{dv}\ln p_\xi^*(v) = x}$$

Class $\mathcal{P}_2^{AR}$ (*containing all centered distributions with a variance not less than a given value*):

$$\boxed{\mathcal{P}_2^{AR} := \left\{p_\xi : \int_{\mathbb{R}} x^2 p_\xi(x)\,dx \geq \sigma_0^2\right\}} \tag{13.160}$$

Lemma 13.14. (on class $\mathcal{P}_2^{AR}$)

$$\boxed{p_\xi^*(x) = \arg\inf_{p_\xi\in\mathcal{P}_2^{AR}:\sigma^2(p_\xi)=\sigma_0^2} I_F\left(p_\xi\right)} \tag{13.161}$$

that is, the value of $\mathcal{P}_2^{AR}$ distribution density $p_\xi^*(x)$ coincides with the value distribution density on the classes $\mathcal{P}_i$ characterizing distribution uncertainties for static (R-models) provided that

$$\boxed{\sigma^2\left(p_\xi^*(x)\right) = \sigma_0^2} \tag{13.162}$$

Proof. It follows directly from the inequality

$$\sigma^2\left(p_\xi\right) I_F\left(p_\xi\right) \geq \sigma_0^2 I_F\left(p_\xi\right)$$

□

14 Filtering, Prediction and Smoothing

Contents

In this chapter we will concentrate on the problem of states estimation for discrete-time and continuous-time processes based on some available observations of these processes which are obviously statistically dependent. Below we will suppose that an unobservable process $x = x(t, \omega)$ and an observable process $y = y(t, \omega)$ are jointly distributed on some underlying probability space.

The *objective* is to form an estimate of the process $x(t, \omega)$ at time $[t + \tau]$, that is, $x(t + \tau, \omega)$ using the observations $\{y(s, \omega), s \in [0, t]\}$ where the time argument t may belong to discrete set $\{t_0 = 0, t_1, t_2, \dots\}$ or an interval $[0, T]$. This estimate $\hat{x}(t + \tau, \omega)$ is required to be optimal with respect to the least square criterion. For

- $\tau = 0$ this task is usually called *filtering*;
- $\tau > 0$ it is called *prediction*;
- $\tau < 0$ it is referred to as *smoothing*.

The physical device generating any one of these estimates from the observed data is called *a filter*.

14.1 Estimation of random vectors

We start with simple (but fundamental) results concerning an estimates construction when available measurements and random processes to be estimated are finite collections, or, in other words, finite-dimensional vectors.

14.1.1 Problem formulation

Consider a collection of random variables

$$\left\{x_i = x_i(\omega) \in \mathbb{R}, i = \overline{1, n};\ y_i = y_i(\omega) \in \mathbb{R}, i = \overline{1, m}\right\}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that the variables, namely,

$$\left\{y_i = y_i(\omega) \in \mathbb{R}, i = \overline{1, m}\right\}$$

generates a σ-algebra

$$\mathcal{F}_m := \sigma\left(y_i = y_i(\omega) \in \mathbb{R}, i = \overline{1, m}\right)$$

We will refer to them as '*available data*' or '*measurement*'. Define

$$x = x(\omega) = (x_1(\omega), \ldots, x_n(\omega)) \in \mathbb{R}^n$$
$$y = y(\omega) = (y_1(\omega), \ldots, y_m(\omega)) \in \mathbb{R}^m$$

and consider the following problem:

Problem 14.1. (on least square estimate) *Based on the available data $y \sim \mathcal{F}_m$ obtain the best mean-square estimate*

$$\boxed{\hat{x} = \hat{x}(y) \in \mathbb{R}^n, \quad \hat{x} \sim \mathcal{F}_m} \tag{14.1}$$

of the vector $x \in \mathbb{R}^n$ (which may be statistically dependent on y), namely, we will find

$$\boxed{\hat{x}^* := \arg\min_{\hat{x}\sim\mathcal{F}_m} E\left\{\left\|x - \hat{x}\right\|^2\right\}} \tag{14.2}$$

14.1.2 Gauss–Markov theorem

The theorem (known also as the '*Orthogonal Projection Theorem*' treating the result as a projection to the subspace of given data) below is a keystone in '*theory of random vectors estimation*' which is more frequently referred to as '*filtering theory*', or simply, '*filtering*'.

Theorem 14.1. (Gauss–Markov) *Within a class of random $\mathcal{F}_m$-measurable vectors $\hat{x}$ with bounded second moments the* ***best mean-square estimate*** *$\hat{x}^*$ of the random vector $x \in \mathbb{R}^n$ is* ***its conditional mathematical expectation*** $\mathsf{E}\{x/\mathcal{F}_m\}$*, that is,*

$$\boxed{\hat{x}^* = \arg\min_{\hat{x}\sim\mathcal{F}_m} \mathsf{E}\left\{\left\|x - \hat{x}\right\|^2\right\} \overset{a.s.}{=} \mathsf{E}\{x/\mathcal{F}_m\}} \tag{14.3}$$

Proof. The following identities hold:

$$\begin{aligned}\mathsf{E}\left\{\left\|x - \hat{x}\right\|^2\right\} &= \mathsf{E}\left\{\left\|\left(x - \hat{x}^*\right) + \left(\hat{x}^* - \hat{x}\right)\right\|^2\right\} \\ &= \mathsf{E}\left\{\left\|x - \hat{x}^*\right\|^2\right\} + \mathsf{E}\left\{\left\|\hat{x}^* - \hat{x}\right\|^2\right\} + 2\mathsf{E}\left\{\left(x - \hat{x}^*\right)^{\intercal}\left(\hat{x}^* - \hat{x}\right)\right\}\end{aligned}$$

Notice that

$$\begin{aligned}&\mathsf{E}\left\{\left(x - \hat{x}^*\right)^{\intercal}\left(\hat{x}^* - \hat{x}\right)\right\} = \mathsf{E}\left\{\mathsf{E}\left\{\left(x - \hat{x}^*\right)^{\intercal}\left(\hat{x}^* - \hat{x}\right)/\mathcal{F}_m\right\}\right\} \\ &\quad = \mathsf{E}\left\{\mathsf{E}\left\{\left(x - \hat{x}^*\right)^{\intercal}/\mathcal{F}_m\right\}\left(\hat{x}^* - \hat{x}\right)\right\} = \mathsf{E}\left\{\left(\mathsf{E}\{x/\mathcal{F}_m\} - \hat{x}^*\right)^{\intercal}\left(\hat{x}^* - \hat{x}\right)\right\} = 0\end{aligned}$$

Therefore

$$\mathsf{E}\left\{\left\|x - \hat{x}\right\|^2\right\} = \mathsf{E}\left\{\left\|x - \hat{x}^*\right\|^2\right\} + \mathsf{E}\left\{\left\|\hat{x}^* - \hat{x}\right\|^2\right\} \geq \mathsf{E}\left\{\left\|x - \hat{x}^*\right\|^2\right\}$$

where the right-hand side does not depend on $\hat{x}$. But this lower bound is achieved if $\hat{x}^* \overset{a.s.}{=} \hat{x}$. Theorem is proven. □

Example 14.1. *Suppose that two random variables* $(x, y) \in \mathbb{R}_+ \times \mathbb{R}$ *have the joint distribution density* $f(x, y)$ *on* $\mathbb{R}_+ \times \mathbb{R}$ *given by*

$$f(x, y) = \frac{1}{\sqrt{2\pi}} \frac{\beta}{\Gamma(\alpha)} x^{\alpha-1/2} \exp\left\{-\left(\beta + \frac{y^2}{2}\right) x\right\}, \quad \alpha, \beta \in \mathbb{R}_+$$

where

$$\Gamma(\alpha) := \int_{t=0}^{\infty} t^{\alpha-1} e^{-t} dt$$

Then

$$\hat{x}^* \overset{a.s.}{=} \mathsf{E}\{x/\sigma(y)\} = \int_{x=0}^{\infty} x f(x/y) dx = \frac{\alpha + 1/2}{\beta + y^2/2}$$

and one can observe that this 'best estimate' is ***not linear*** *in the observation* y.

14.1.3 Linear unbiased estimates

Definition 14.1.

(a) *The class* $\mathcal{K}_y$ *of all* ***linear estimates*** $\hat{x} \sim \mathcal{F}_m$ $\left(\hat{x} = \hat{x}(y) \in \mathbb{R}^n\right)$, *based on available data* $y \in \mathbb{R}^m$, *is defined by the following relation:*

$$\boxed{\hat{x} = K y + \tilde{x}} \tag{14.4}$$

where $K \in \mathbb{R}^{n \times m}$ *is a constant (deterministic) matrix and* $\tilde{x} \in \mathbb{R}^n$ *is a constant (deterministic) vector.*

(b) *A linear estimate* $\hat{x} \in \mathcal{K}_y$ *is called* ***unbiased*** *if*

$$\boxed{\mathsf{E}\left\{\hat{x}\right\} = \mathsf{E}\{x\} := \bar{x}} \tag{14.5}$$

Lemma 14.1. (on a structure of linear unbiased estimates) *Any linear unbiased estimate has the following representation:*

$$\boxed{\hat{x} = \bar{x} + K(y - \bar{y})} \tag{14.6}$$

where

$$\boxed{\bar{y} := \mathsf{E}\{y\}} \tag{14.7}$$

Proof. From (14.4), (14.5) and (14.7) it follows that

$$\bar{x} = \mathsf{E}\left\{\hat{x}\right\} = K\mathsf{E}\{y\} + \tilde{x} = K\bar{y} + \tilde{x}$$

and hence, $\tilde{x} = \bar{x} - K\bar{y}$ (which implies structure). Lemma is proven. □

The next results concerns the *'best'* selection of the matrix K in (14.6) which guarantees the minimum of the mean-square error $\mathsf{E}\left\{\|x-\hat{x}\|^2\right\}$ within the class of all linear unbiased estimates.

First, define the so-called *centered random vectors*:

$$\boxed{\mathring{x} := x - \bar{x}, \quad \mathring{y} := y - \bar{y}} \tag{14.8}$$

Theorem 14.2. *For any available data $y \in \mathbb{R}^m$ with a nonsingular auto-covariation*

$$\boxed{R_{\mathring{y},\mathring{y}} := \mathsf{E}\left\{\mathring{y}\mathring{y}^{\intercal}\right\} > 0} \tag{14.9}$$

the best linear unbiased estimate, minimizing $\mathsf{E}\left\{\|x-\hat{x}\|^2\right\}$, is given by (14.6) with

$$\boxed{K = K^* = R_{\mathring{x},\mathring{y}} R_{\mathring{y},\mathring{y}}^{-1}} \tag{14.10}$$

where $R_{\mathring{x},\mathring{y}} := \mathsf{E}\{\mathring{x}\mathring{y}^{\intercal}\}$, that is,

$$\boxed{\hat{x} = \bar{x} + R_{\mathring{x},\mathring{y}} R_{\mathring{y},\mathring{y}}^{-1} (y - \bar{y})} \tag{14.11}$$

Proof. One has

$$\begin{aligned}
\mathsf{E}\left\{\|x-\hat{x}\|^2\right\} &= \mathsf{E}\left\{\operatorname{tr}\|x-\hat{x}\|^2\right\} = \mathsf{E}\left\{\operatorname{tr}\left\{\left(x-\hat{x}\right)\left(x-\hat{x}\right)^{\intercal}\right\}\right\} \\
&= \mathsf{E}\left\{\operatorname{tr}\left\{(x-\bar{x}-K\mathring{y})(x-\bar{x}-K\mathring{y})^{\intercal}\right\}\right\} \\
&= \operatorname{tr}\left\{\mathsf{E}\left\{(\mathring{x}-K\mathring{y})(\mathring{x}-K\mathring{y})^{\intercal}\right\}\right\} \\
&= \operatorname{tr}\left\{\mathsf{E}\left\{\mathring{x}\mathring{x}^{\intercal}\right\} - \mathsf{E}\left\{\mathring{x}\mathring{y}^{\intercal}\right\}K^{\intercal} - K\mathsf{E}\left\{\mathring{y}\mathring{x}^{\intercal}\right\} + K\mathsf{E}\left\{\mathring{y}\mathring{y}^{\intercal}\right\}K^{\intercal}\right\} \\
&= \operatorname{tr}\left\{R_{\mathring{x},\mathring{x}} - R_{\mathring{x},\mathring{y}}K^{\intercal} - KR_{\mathring{x},\mathring{y}}^{\intercal} + KR_{\mathring{y},\mathring{y}}K^{\intercal}\right\} \\
&= \operatorname{tr}\left\{R_{\mathring{x},\mathring{x}} - R_{\mathring{x},\mathring{y}}R_{\mathring{y},\mathring{y}}^{-1}R_{\mathring{x},\mathring{y}}^{\intercal}\right\} \\
&\quad + \operatorname{tr}\left\{\left(KR_{\mathring{y},\mathring{y}}^{1/2} - R_{\mathring{x},\mathring{y}}R_{\mathring{y},\mathring{y}}^{-1/2}\right)\left(KR_{\mathring{y},\mathring{y}}^{1/2} - R_{\mathring{x},\mathring{y}}R_{\mathring{y},\mathring{y}}^{-1/2}\right)^{\intercal}\right\} \\
&\geq \operatorname{tr}\left\{R_{\mathring{x},\mathring{x}} - R_{\mathring{x},\mathring{y}}R_{\mathring{y},\mathring{y}}^{-1}R_{\mathring{x},\mathring{y}}^{\intercal}\right\}
\end{aligned}$$

So, the right-hand side of the last inequality is independent of K and the equality is attained if and only if

$$KR_{\mathring{y},\mathring{y}}^{1/2} - R_{\mathring{x},\mathring{y}}R_{\mathring{y},\mathring{y}}^{-1/2} = 0 \tag{14.12}$$

which implies (15.37). Theorem is proven. □

Corollary 14.1. *Under the conditions of Theorem 14.2*

$$\min_{K:\hat{x}=\bar{x}+K(y-\bar{y})} \mathsf{E}\left\{\left\|x-\hat{x}\right\|^2\right\} = \operatorname{tr}\left\{R_{\mathring{x},\mathring{x}} - R_{\mathring{x},\mathring{y}} R_{\mathring{y},\mathring{y}}^{-1} R_{\mathring{x},\mathring{y}}^{\mathsf{T}}\right\} \tag{14.13}$$

Corollary 14.2. *The optimal gain-matrix selection $K = K^*$ (15.37) provides the property*

$$\mathsf{E}\left\{\left(\mathring{x} - K^*\mathring{y}\right)\mathring{y}^{\mathsf{T}}\right\} = 0 \tag{14.14}$$

that is, the vectors $(\mathring{x} - K^*\mathring{y})$ and $\mathring{y}$ are not correlated.

Proof. Multiplying both sides of (14.12) by $R_{\mathring{y},\mathring{y}}^{1/2}$ leads directly to (14.14). □

14.1.4 Lemma on normal correlation

Here we will consider Theorem 14.2 when both vectors x and y are Gaussian.

Lemma 14.2. (on a normal correlation) *In the Gaussian case when both vectors x and y are Gaussian and such that $R_{\mathring{y},\mathring{y}} > 0$ the best estimate (14.3)*

$$\hat{x}^* = \arg\min_{\hat{x}\sim\mathcal{F}_m} \mathsf{E}\left\{\left\|x-\hat{x}\right\|^2\right\} \overset{a.s.}{=} \mathsf{E}\left\{x/\mathcal{F}_m\right\}$$

within the class of all possible (maybe, nonlinear) quadratically integrable $\mathcal{F}_m$-measurable functions ***coincides*** *with the best linear unbiased estimate (14.11), namely,*

$$\begin{gathered} \hat{x}^* \overset{a.s.}{=} \mathsf{E}\left\{x/\mathcal{F}_m\right\} = \bar{x} + K^*\left(y-\bar{y}\right) \\ K^* = R_{\mathring{x},\mathring{y}} R_{\mathring{y},\mathring{y}}^{-1} \end{gathered} \tag{14.15}$$

Proof. Indeed, if both vectors $\mathring{x}$ and $\mathring{y}$ are Gaussian then the vector $\left[\mathring{x} - K^*\mathring{y}\right]$ is also Gaussian. This, in view of (14.14), implies that the vectors $\left[\mathring{x} - K^*\mathring{y}\right]$ and $\mathring{y}$ are independent since, only for Gaussian vectors, an uncorrelation means their independence. Using this fact we conclude that

$$\begin{aligned} \mathsf{E}\left\{\mathring{x} - K^*\mathring{y}/\mathcal{F}_m\right\} &= \mathsf{E}\left\{\mathring{x} - K^*\mathring{y}/\sigma\left(y\right)\right\} \\ &\overset{a.s.}{=} \mathsf{E}\left\{\mathring{x} - K^*\mathring{y}\right\} = \mathsf{E}\left\{\mathring{x}\right\} - K^*\mathsf{E}\left\{\mathring{y}\right\} = 0 \end{aligned}$$

which leads to the following identity

$$\mathsf{E}\left\{\mathring{x}/\mathcal{F}_m\right\} = K^*\mathsf{E}\left\{\mathring{y}/\mathcal{F}_m\right\} = K^*\mathring{y}$$

equivalent to (14.15). Lemma is proven. □

14.2 State-estimating of linear discrete-time processes

Here we will consider only linear discrete-time processes given by linear recurrent equations with constant or time-varying parameters. The problem to be solved is to estimate linearly $x_{n+\tau}$ from measurements $\left\{y_{0,}y_1,\ldots,y_n\right\}$.

14.2.1 Description of linear stochastic models

Consider here a linear discrete-time stochastic process given by the following linear recurrent equation:

$$\boxed{\begin{array}{c} x_{n+1} = A_n x_n + b_n + \xi_n \\ y_n = H_n x_n + h_n + w_n, \quad n = 1, 2, \ldots \end{array}} \tag{14.16}$$

where

- $x_n \in \mathbb{R}^n$ is the current *state* of the process, x_0 is a random vector with a finite second moment;
- $y_n \in \mathbb{R}^m$ is the current measurable (available) *output*;
- $b_n \in \mathbb{R}^n$ and $h_n \in \mathbb{R}^m$ are measurable (available) vectors depending on the available data $\{y_1, \ldots, y_n\}$, treated below as measurable inputs;
- $A_n \in \mathbb{R}^{n\times n}$, $H_n \in \mathbb{R}^{m\times n}$ are known deterministic matrices;
- $\xi_n \in \mathbb{R}^n$ and $w_n \in \mathbb{R}^m$ are unmeasurable **independent** *random vectors*, referred to below as the state and output noise (may be, non-stationary), respectively, both defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, which are centered and having known auto-covariance Ξ_n and W_n, i.e., for all $n = 0, 1, \ldots$

$$\boxed{\begin{array}{c} \mathsf{E}\{\xi_n\} = 0,\ \mathsf{E}\{w_n\} = 0 \\ \mathsf{E}\left\{\xi_n\xi_n^{\mathsf{T}}\right\} = \Xi_n,\ \mathsf{E}\left\{w_n w_n^{\mathsf{T}}\right\} = W_n \\ \mathsf{E}\{\xi_l \xi s\} = \Xi_s\delta_{l.s},\ \mathsf{E}\left\{w_l w_s^{\mathsf{T}}\right\} = W_s\delta_{l.s} \\ \mathsf{E}\left\{\xi_l w_s^{\mathsf{T}}\right\} = 0,\ l, s = 0, 1, \ldots \end{array}} \tag{14.17}$$

Below we will denote by

$$\sigma\left(y^n\right) := \mathcal{F}\left(y_1, \ldots, y_n\right) = \mathcal{F}_n$$

a sigma-algebra generated by the available data-collection (a prehistory) of the considered process (14.16).

14.2.2 Discrete-time Kalman filtering

Here we will be interesting in the *filtering* problem, that is,

Problem 14.2.

$$\boxed{\bar{x}_{n|n} := \arg\min_{\hat{x}\sim\mathcal{F}_n} E\left\{\left\|x_n - \hat{x}\right\|^2\right\}} \tag{14.18}$$

Theorem 14.3. (**Kalman, 1960**) *The best mean-square linear unbiased estimate* $\bar{x}_{n|n}$ *(14.18) of the state* x_n *generated by (14.16)–(14.17), based on the available data* $\{y_1, \ldots, y_n\}$*, is given by the following recurrent scheme (the '**Kalman filter**'):*

$$\boxed{\begin{gathered}\bar{x}_{n|n} = \bar{x}_{n|n-1} + K_n^* \left(y_n - \bar{y}_{n|n-1}\right) \\ \bar{x}_{n|n-1} := A_{n-1}\bar{x}_{n-1|n-1} + b_{n-1} \\ \bar{y}_{n|n-1} = H_n\bar{x}_{n|n-1} + h_n\end{gathered}} \tag{14.19}$$

where

$$\boxed{\begin{gathered}K_n^* = S_n^{\mathsf{T}} D_n^{-1} \\ S_n = H_n \left(\Xi_{n-1} + A_{n-1} P_{n-1} A_{n-1}^{\mathsf{T}}\right) \\ D_n = W_n + S_n H_n^{\mathsf{T}} \\ P_n := \mathsf{E}\left\{\left(x_n - \bar{x}_{n|n}\right)\left(x_n - \bar{x}_{n|n}\right)^{\mathsf{T}}\right\} \\ P_n = A_{n-1} P_{n-1} A_{n-1}^{\mathsf{T}} + \Xi_n - S_n^{\mathsf{T}} D_n^{-1} S_n \\ \bar{x}_{0|0} = \mathsf{E}\{x_0\},\ \ P_0 = \operatorname{cov}\{x_0\} := \mathsf{E}\left\{\left(x_0 - \bar{x}_{0|0}\right)\left(x_0 - \bar{x}_{0|0}\right)^{\mathsf{T}}\right\}\end{gathered}} \tag{14.20}$$

provided that $D_n > 0$.

Proof. Any state estimate $\hat{x}_n$ is linear with respect to the last observation y_n if it may be represented as

$$\hat{x}_n = K_n y_n + c_n \tag{14.21}$$

where $K_n \in \mathbb{R}^{n\times m}$ and $c_n \in \mathbb{R}^n$ are $\mathcal{F}_{n-1} := \mathcal{F}(y_1, \ldots, y_{n-1})$-measurable matrix and vector, respectively. Define $\hat{x}_{n|n-1} := \mathsf{E}\left\{\hat{x}_n/\mathcal{F}_{n-1}\right\}$ which leads to

$$\hat{x}_{n|n-1} \overset{a.s.}{=} K_n \bar{y}_{n|n-1} + c_n$$

where $\bar{y}_{n|n-1} := \mathsf{E}\{y_n/\mathcal{F}_{n-1}\}$. Therefore $c_n \overset{a.s.}{=} \hat{x}_{n|n-1} - K_n \bar{y}_{n|n-1}$ and, hence, by (14.21)

$$\hat{x}_n \overset{a.s.}{=} \hat{x}_{n|n-1} + K_n \left(y_n - \bar{y}_{n|n-1}\right) \tag{14.22}$$

By the Gauss–Markov theorem 14.1 the best mean-square estimate of x_n is

$$\hat{x}_n^* = \bar{x}_{n|n} := \mathsf{E}\{x_n/\mathcal{F}_n\}$$

which by (14.22) gives

$$\begin{gathered}\bar{x}_{n|n} = \bar{x}_{n|n-1} + K_n \left(y_n - \bar{y}_{n|n-1}\right) \\ \bar{x}_{n|n-1} = \mathsf{E}\{\mathsf{E}\{x_n/\mathcal{F}_n\}/\mathcal{F}_{n-1}\} = \mathsf{E}\{x_n/\mathcal{F}_{n-1}\}\end{gathered} \tag{14.23}$$

Now notice that $\bar{x}_{n|n-1}$ and $\bar{y}_{n|n-1}$ can be easily obtained, applying $\mathsf{E}\{\cdot/\mathcal{F}_{n-1}\}$ to (14.16):

$$\begin{aligned} &\bar{x}_{n|n-1} \overset{a.s.}{=} A_{n-1}\bar{x}_{n-1|n-1} + b_{n-1} \\ &\bar{x}_{0|0} = \mathsf{E}\{x_0/\mathcal{F}_0\} = \mathsf{E}\{x_0/\mathcal{F}_0\} \\ &\bar{y}_{n|n-1} \overset{a.s.}{=} H_n\bar{x}_{n|n-1} + h_n \end{aligned} \tag{14.24}$$

For the state estimation error $\Delta_n := x_n - \bar{x}_{n|n}$, using (14.16), (14.23) and (14.24), one has

$$\begin{aligned} \Delta_n &= A_{n-1}x_{n-1} + b_{n-1} + \xi_{n-1} - \left[\bar{x}_{n|n-1} + K_n\left(y_n - \bar{y}_{n|n-1}\right)\right] \\ &= A_{n-1}x_{n-1} + b_{n-1} + \xi_{n-1} - A_{n-1}\bar{x}_{n-1|n-1} \\ &\quad - b_{n-1} - K_n\left(H_n x_n + h_n + w_n - H_n\left[A_{n-1}\bar{x}_{n-1|n-1} + b_{n-1}\right] - h_n\right) \\ &= (I - K_nH_n)A_{n-1}\Delta_{n-1} + (I - K_nH_n)\xi_{n-1} - K_n w_n \end{aligned} \tag{14.25}$$

The relation (14.25) implies

$$\begin{aligned} P_n &:= \mathsf{E}\left\{\Delta_n\Delta_n^{\mathsf{T}}\right\} \\ &= (I - K_nH_n)A_{n-1}P_{n-1}A_{n-1}^{\mathsf{T}}(I - K_nH_n)^{\mathsf{T}} + K_nW_nK_n^{\mathsf{T}} \\ &\quad + (I - K_nH_n)\Xi_{n-1}(I - K_nH_n)^{\mathsf{T}} \\ &= A_{n-1}P_{n-1}A_{n-1}^{\mathsf{T}} - K_nS_n - S_n^{\mathsf{T}}K_n^{\mathsf{T}} + K_nD_nK_n + \Xi_{n-1} \\ &= A_{n-1}P_{n-1}A_{n-1}^{\mathsf{T}} + \Xi_{n-1} - S_n^{\mathsf{T}}D_n^{-1}S_n \end{aligned}$$

$$\left(K_nD_n^{1/2} - S_n^{\mathsf{T}}D_n^{-1/2}\right)\left(K_nD_n^{1/2} - S_n^{\mathsf{T}}D_n^{-1/2}\right)^{\mathsf{T}} \geq A_{n-1}P_{n-1}A_{n-1}^{\mathsf{T}} + \Xi_{n-1} - S_n^{\mathsf{T}}D_n^{-1}S_n$$

where the equality is attained when $K_nD_n^{1/2} - S_n^{\mathsf{T}}D_n^{-1/2} = 0$ which completes the proof. □

Corollary 14.3. *Assuming that ξ_n, w_n are **Gaussian** and b_n, h_n are measurable vectors **linearly** depending on the available data $\{y_1, \ldots, y_n\}$, it follows that the Kalman filter (14.19)–(14.20) is **optimal** not only among all linear estimates but among all possible (may be, nonlinear) estimates which may be constructed based on available data.*

Proof. This results from the fact that in this case the sequence $\{x_n\}_{n\geq 0}$ is Gaussian, and hence, by Lemma 14.2, $\bar{x}_{n|n} = \mathsf{E}\{x_n/\mathcal{F}_n\}$ coincides with $\hat{x}_n = \hat{x}_{n|n-1} + K_n\left(y_n - \bar{y}_{n|n-1}\right)$ where K_n minimizes

$$\mathsf{E}\left\{\|\Delta_n\|^2\right\} = \operatorname{tr}\mathsf{E}\left\{\Delta_n\Delta_n^{\mathsf{T}}\right\} = \operatorname{tr}P_n$$ □

Corollary 14.4. *The relations (14.20) can be represented in more compact format as*

$$\boxed{\begin{aligned} K_n^* &= P_{n|n-1}H_n^{\mathsf{T}}\left[W_n + H_nP_{n|n-1}H_n^{\mathsf{T}}\right]^{-1} \\ P_{n|n-1} &= A_{n-1}P_{n-1}A_{n-1}^{\mathsf{T}} + \Xi_{n-1} \\ P_n &= \left(I - K_n^*H_n\right)P_{n|n-1} \end{aligned}} \tag{14.26}$$

14.2.3 Discrete-time prediction and smoothing

Consider below the sub-class dynamic models (14.16)–(14.17) where the 'input' vectors $b_n \in \mathbb{R}^n$ and $h_n \in \mathbb{R}^m$ are *deterministic* and *measurable* (available). By the Gauss–Markov theorem 14.1 the *best mean-square estimate* $\hat{x}^*_{n+\tau}$ of the model state vector $x_{n+\tau}$ for any $\tau = -n, \ldots, -1, 0, 1, \ldots$ is

$$\boxed{\hat{x}^*_{n+\tau} = \bar{x}_{n+\tau|n} := \mathsf{E}\{x_{n+\tau}/\mathcal{F}_n\}} \tag{14.27}$$

14.2.3.1 Prediction

This case corresponds to $\tau = 1, 2, \ldots$ in (14.27). Taking into account that the noises ξ_n, w_n are independent and centered, the application of the operator $\mathsf{E}\{\cdot/\mathcal{F}_n\}$ to both sides of (14.16) with the following *back-iteration* implies

$$\begin{aligned}
\bar{x}_{n+\tau|n} &= \mathsf{E}\{x_{n+\tau}/\mathcal{F}_n\} = A_{n+\tau-1}\mathsf{E}\{x_{n+\tau-1}/\mathcal{F}_n\} + b_{n+\tau-1} \\
&= A_{n+\tau-1}\left[A_{n+\tau-2}\mathsf{E}\{x_{n+\tau-2}/\mathcal{F}_n\} + b_{n+\tau-2}\right] + b_{n+\tau-1} \\
&= \left[A_{n+\tau-1}\cdots A_n\right]\mathsf{E}\{x_n/\mathcal{F}_n\} \\
&\quad + \left[A_{n+\tau-1}\cdots A_{n+1}\right] b_n + \cdots + A_{n+\tau-1}b_{n+\tau-2} + b_{n+\tau-1} \\
&= \left(\prod_{k=0}^{\tau-1} A_{n+k}\right)\bar{x}_{n|n} + \sum_{k=0}^{\tau-1}\left(\prod_{s=k}^{\tau-2} A_{n+1+s}\right) b_{n+k}
\end{aligned}$$

(here we accept that $\prod_{s=k}^{m} A_s = I$ if $m < k$). So, we are ready to formulate the main result dealing with a prediction design.

Theorem 14.4. (on prediction) *If within the sub-class dynamic models (14.16)–(14.17) the 'input' vectors $b_n \in \mathbb{R}^n$ and $h_n \in \mathbb{R}^m$ are deterministic and measurable (available) then the **best mean square prediction** for τ-steps ahead* $(\tau = 1, 2, \ldots)$

$$\boxed{\bar{x}_{n+\tau|n} := \mathsf{E}\{x_{n+\tau}/\mathcal{F}_n\} = \arg\min_{\hat{x}\sim\mathcal{F}_n} E\left\{\left\|x_{n+\tau} - \hat{x}\right\|^2\right\}} \tag{14.28}$$

as a function of the available data $\{y_1, \ldots, y_n\}$ *is given by*

$$\boxed{\bar{x}_{n+\tau|n} = \left(\prod_{k=0}^{\tau-1} A_{n+k}\right)\bar{x}_{n|n} + \sum_{k=0}^{\tau-1}\left(\prod_{s=k}^{\tau-2} A_{n+1+s}\right) b_{n+k}} \tag{14.29}$$

where $\bar{x}_{n|n}$ is the state estimation generated by the recurrent Kalman filter (14.19)–(14.26).

The formula (14.29) can be rewritten in the recurrent form as follows:

$$\boxed{\begin{array}{c} \bar{x}_{n+\tau|n} = P^{A}_{n,n+\tau-1}\bar{x}_{n|n} + \sum_{k=0}^{\tau-1}\left(\prod_{s=k}^{\tau-2} A_{n+1+s}\right) b_{n+k} \\ P^{A}_{n,n+\tau-1} = A_{n+\tau-1}P^{A}_{n,n+\tau-2}, \ P^{A}_{n,n} = A_n \\ S_{n,n+\tau-1} = A_{n+\tau-1}S_{n,n+\tau-2} + b_{n+\tau-1}, \ S_{n,n} = b_n \end{array}} \tag{14.30}$$

where

$$\boxed{\begin{aligned} P^{A}_{n,n+\tau-1} &:= \left(\prod_{k=0}^{\tau-1} A_{n+k}\right) \\ S_{n,n+\tau-1} &:= \sum_{k=0}^{\tau-1}\left(\prod_{s=k}^{\tau-2} A_{n+1+s}\right) b_{n+k} \\ &= A_{n+\tau-1}\sum_{k=0}^{\tau-2}\left(\prod_{s=k}^{\tau-3} A_{n+1+s}\right) b_{n+k} + b_{n+\tau-1} \end{aligned}} \tag{14.31}$$

14.2.3.2 Smoothing

This case corresponds to $\tau = -1, -2, \ldots, n$ in (14.27). Again, taking into account that the noises ξ_n, w_n are independent and centered, the application of the operator $\mathsf{E}\{\cdot/\mathcal{F}_n\}$ to both sides of (14.16) with the following *forward-iteration* implies

$$\begin{aligned} \bar{x}_{n|n} &= \mathsf{E}\{x_n/\mathcal{F}_n\} = A_{n-1}\mathsf{E}\{x_{n-1}/\mathcal{F}_n\} + b_{n-1} \\ &= A_{n-1}\left[A_{n-2}\mathsf{E}\{x_{n-2}/\mathcal{F}_n\} + b_{n-2}\right] + b_{n-1} \\ &= \left[A_{n-1}\cdots A_{n-\tau}\right]\mathsf{E}\{x_{n-\tau}/\mathcal{F}_n\} \\ &\quad + \left[A_{n-1}\cdots A_{n-\tau+1}\right]b_{n-\tau} + \cdots + A_{n-1}b_{n-2} + b_{n-1} \\ &= \left(\prod_{k=-\tau}^{-1} A_{n+k}\right)\bar{x}_{n-\tau|n} + \sum_{k=1}^{\tau}\left(\prod_{s=-k}^{-2} A_{n+s+1}\right) b_{n-k} \end{aligned} \tag{14.32}$$

which leads to the following formulation of the main result dealing with a smoothing design.

Theorem 14.5. (on smoothing) *If within the sub-class dynamic models (14.16)–(14.17) the 'input' vectors $b_n \in \mathbb{R}^n$, $h_n \in \mathbb{R}^m$ are deterministic and measurable (available) and*

$$\boxed{\det A_{n+k} \neq 0 \quad \textit{for all} \quad k = -\tau, \ldots, -1} \tag{14.33}$$

then the ***best mean square smoothing*** *for τ-steps back* ($\tau = 1, 2, \ldots, n$)

$$\boxed{\bar{x}_{n-\tau|n} := \mathsf{E}\{x_{n-\tau}/\mathcal{F}_n\} = \arg\min_{\hat{x}\sim\mathcal{F}_n} E\left\{\left\|x_{n-\tau} - \hat{x}\right\|^2\right\}} \tag{14.34}$$

as a function of the available data $\{y_1, \dots, y_n\}$ *is given by*

$$\boxed{\bar{x}_{n-\tau|n} = \left(\prod_{k=-\tau}^{-1} A_{n+k}\right)^{-1} \left[\bar{x}_{n|n} - \sum_{k=1}^{\tau} \left(\prod_{s=-k}^{-2} A_{n+s+1}\right) b_{n-k}\right]} \tag{14.35}$$

where $\bar{x}_{n|n}$ *is the state estimation generated by the recurrent Kalman's filter (14.19)–(14.26).*

Proof. It follows from (14.32) if take into account that the matrix $\left(\prod_{k=-\tau}^{-1} A_{n+k}\right)$ is non-singular if (14.33) holds. □

14.3 State-estimating of linear continuous-time processes

14.3.1 Structure of an observer for linear stochastic processes

Consider here a linear stochastic process[1] given by

$$\boxed{\begin{gathered} dx\,(t,\omega) = [A(t)x\,(t,\omega) + b\,(t)]\,dt + C\,(t)\,dV_t\,(\omega) \\ x\,(0,\omega) = x_0\,(\omega) \\ dy\,(t,\omega) = H(t)x\,(t,\omega)\,dt + D\,(t)\,dW_t(\omega), \quad t \ge 0 \end{gathered}} \tag{14.36}$$

where

- $x\,(t,\omega) \in \mathbb{R}^n$ is the current *state* of the process, $x_0\,(\omega)$ is a random vector with a finite second moment;
- $b(t) \in \mathbb{R}^n$ is known measurable input;
- $y\,(t,\omega) \in \mathbb{R}^m$ is the current measurable (available) *output*;
- $A(t) \in \mathbb{R}^{n\times n}$, $H(t) \in \mathbb{R}^{m\times n}$ are known deterministic matrices;
- $V_t\,(\omega) \in \mathbb{R}^{k_v}$ and $W_t\,(\omega) \in \mathbb{R}^{k_w}$ are unmeasurable **independent** *standard Wiener processes*, referred below to as the state and output noise, respectively; namely, they

[1] If the second equation in (14.36) has another form, namely,

$$dy\,(t,\omega) = H(t)\,dx\,(t,\omega) + D\,(t)\,dW_t(\omega), \quad t \ge 0$$

then, by the first equation in (14.36), it follows that

$$\begin{aligned} &dy\,(t,\omega) = \tilde{H}(t)x\,(t,\omega)\,dt + \tilde{D}\,(t)\,d\tilde{W}_t\,(\omega) \\ &\tilde{H}(t) := H(t)A(t),\ \ \tilde{D}\,(t) := \begin{bmatrix} C\,(t) & D\,(t) \end{bmatrix} \\ &d\tilde{W}_t\,(\omega) = \begin{pmatrix} dV_t\,(\omega) \\ dW_t\,(\omega) \end{pmatrix} \end{aligned}$$

So, the considered version is equivalent to the original one given in (14.36).

are centered Gaussian vector processes with orthogonal independent increments such that

$$\boxed{E\left\{\begin{pmatrix} V_t(\omega) \\ W_t(\omega) \end{pmatrix}\begin{pmatrix} V_t(\omega) \\ W_t(\omega) \end{pmatrix}^{\mathsf{T}}\right\} = \begin{pmatrix} I_{k_v \times k_v} & 0 \\ 0 & I_{k_w \times k_w} \end{pmatrix} t} \tag{14.37}$$

Matrices $C(t) \in \mathbb{R}^{n \times k_v}$ and $D(t) \in \mathbb{R}^{m \times k_w}$ are supposed to be deterministic and known. Moreover, to simplify the presentation the following assumption will be in force hereafter:

A1. For all $t \in [0, T]$

$$\boxed{D(t) D^{\mathsf{T}}(t) > 0} \tag{14.38}$$

The *problem* under consideration is as follows.

Problem 14.3. *Design a process*

$$\hat{x}(t, \omega) = \hat{x}(y(\tau, \omega),\ \tau \in [0, t]) \in \mathbb{R}^n \tag{14.39}$$

referred to as a ***state estimate*** *(or, a state observer) such that an estimation error*

$$\Delta x(t, \omega) := \hat{x}(t, \omega) - x(t, \omega) \tag{14.40}$$

would be as small as possible (in some probabilistic sense).

Here we will use the following definition.

Definition 14.2. *We say that a 'state estimate' $\hat{x}(t, \omega)$ is generated by a* ***global*** *(full order)* ***linear differential observer*** *(or, a filter) if it satisfies the following three conditions:*

1. ***(OSDE property)****: the function $\hat{x}_t$ is the solution of the following ordinary linear stochastic differential equation*

$$\boxed{d\hat{x}(t, \omega) = \left[G(t)\hat{x}(t, \omega) + b(t)\right]dt + L(t)\,dy_t(t, \omega), \quad \hat{x}_0 \text{ is fixed}} \tag{14.41}$$

where $G(t) \in \mathbb{R}^{n \times n}$, $L(t) \in \mathbb{R}^{m \times n}$ are some deterministic matrices;

2. *(**The exact mapping property**): the trajectories $x(t, \omega)$ of the given system (14.36) and $\hat{x}(t, \omega)$ (14.41) coincide for all $t \geq 0$ with probability one, that is,*

$$\boxed{x(t, \omega) = \hat{x}(t, \omega), \quad dx(t, \omega) = d\hat{x}(t, \omega)} \tag{14.42}$$

if the initial states (14.41) coincide, i.e.,

$$x(0, \omega) \overset{a.s.}{=} \hat{x}(0, \omega)$$

and when there are no disturbances at all, that is, for all $t \geq 0$

$$C(t) = 0 \quad \textit{and} \quad D(t) = 0$$

3. *(**The asymptotic consistency property**): if the initial states of the original model and the estimating model do not coincide with some positive probability, that is,*

$$\mathsf{E}\left\{\left\|x(0,\omega)-\hat{x}(0,\omega)\right\|^2\right\}>0$$

but still there are no disturbances, namely,

$$C(t)=0 \quad \text{and} \quad D(t)=0$$

then the estimates $\hat{x}(t,\omega)$ should satisfy (if we are interested in the state estimation on a infinite time-interval)

$$\boxed{\mathsf{E}\left\{\left\|x(t,\omega)-\hat{x}(t,\omega)\right\|^2\right\}\underset{t\to\infty}{\to}0} \tag{14.43}$$

Lemma 14.3. *Both models (14.36) and (14.41) satisfy the condition 2 in 14.2 if and only if $G(t)$ and $L(t)$ in (14.41) are as follows:*

$$\boxed{G(t)=A(t)-L(t)H(t)} \tag{14.44}$$

for almost all $t\geq 0$.

Proof. Since by the condition 2 $C(t)=0$ and $D(t)=0$, it follows that

$$\begin{aligned} d(\Delta x(t,\omega)) &= G(t)\hat{x}(t,\omega)\,dt+[L(t)H(t)-A(t)]\,x(t,\omega)\,dt \\ &= [\theta(t)-b(t)]\,dt+G(t)\Delta x(t,\omega)\,dt \\ &\quad +[L(t)H(t)-A(t)+G(t)]\,x(t,\omega)\,dt \end{aligned} \tag{14.45}$$

(a) *Necessity*. Putting $\Delta x(t,\omega)\overset{a.s.}{=}0$ and $d(\Delta x(t,\omega))\overset{a.s.}{=}0$, we get

$$[L(t)H(t)-A(t)+G(t)]\,x(t,\omega)\overset{a.s.}{=}0$$

which should be valid for any initial conditions $x(0,\omega)\overset{a.s.}{=}\hat{x}(0,\omega)$, and hence, for any $x(t,\omega)$. This implies the identity

$$L(t)H(t)-A(t)+G(t)=0$$

for almost all $t\geq 0$.
(b) *Sufficiency*. Suppose that (14.44) holds. Then by (14.45) we have

$$d(\Delta x(t,\omega))=G(t)\Delta x(t,\omega)\,dt \tag{14.46}$$

which, in view of the condition $\Delta x(0,\omega)\overset{a.s.}{=}0$ implies

$$\Delta x(t,\omega)=\Phi_G(t,0)\,\Delta x(0,\omega)\overset{a.s.}{=}0$$

where $\Phi_G(t,0)$ is the fundamental matrix of the linear vector equation (14.46). Lemma is proven. □

Using (14.44), one can represent the linear observer (filter) as

$$\boxed{\begin{aligned} d\hat{x}(t,\omega) = {} & \left[A(t)\hat{x}(t,\omega) + b(t)\right]dt \\ & + L(t)\left[dy_t(t,\omega) - H(t)\hat{x}(t,\omega)\,dt\right], \quad \hat{x}_0 \text{ is fixed} \end{aligned}} \tag{14.47}$$

Claim 14.1. *Any global linear differential observer (or, a filter) GL (14.41) should have the structure (14.47)*[2] *which repeats the linear regular part*

$$A(t)\hat{x}(t,\omega) + b(t) \tag{14.48}$$

and additionally, contains the correction term

$$L(t)\left[dy_t(t,\omega) - H(t)\hat{x}(t,\omega)\,dt\right] \tag{14.49}$$

where the matrix $L(t)$ is referred to as the ***observer gain-matrix****. The quality of the state estimating process (or filtering) obviously depends on the selection of this observer gain-matrix $L(t)$.*

Lemma 14.4. *Both models (14.36) and (14.41) satisfy condition 3 in 14.2 if and only if the linear time-varying system*

$$d(\Delta x(t,\omega)) = G(t)\,\Delta x(t,\omega)\,dt \tag{14.50}$$

is ***asymptotically stable****, or in other words, when*

$$\boxed{\Phi_G(t,0) \underset{t\to\infty}{\to} 0} \tag{14.51}$$

with $\Phi_G(t,0)$ satisfying

$$\begin{aligned} & \frac{d}{dt}\Phi_G(t,0) = [A(t) - L(t)\,H(t)]\,\Phi_G(t,0) \\ & \Phi_G(0,0) = I \end{aligned} \tag{14.52}$$

Proof. By (14.45) we have

$$d(\Delta x(t,\omega)) = G(t)\,\Delta x(t,\omega)\,dt$$

which leads to

$$\Delta x(t,\omega) = \Phi_G(t,0)\,\Delta x(0,\omega) \tag{14.53}$$

---

2 This system is referred to as '*Luenberger-structure observer*'.

and, as the result, to

$$\begin{aligned}\mathsf{E}\left\{\|\Delta x(t,\omega)\|^2\right\} &= \mathsf{E}\left\{\|\Phi_G(t,0)\,\Delta x(0,\omega)\|^2\right\} \\ &\le \|\Phi_G(t,0)\|^2\,\mathsf{E}\left\{\|\Delta x(0,\omega)\|^2\right\}\end{aligned} \tag{14.54}$$

(a) *Necessity*. Evidently, it results from (14.53).
(b) Sufficiency follows from (14.54).
Lemma is proven. □

Remark 14.1. *In a stationary case, when all matrices $A(t)$, $L(t)$ and $H(t)$ are constant, i.e.,*

$$A(t) = A, \quad L(t) = L, \quad H(t) = H$$

the existence of the gain-matrix L, providing the Hurwitz property (stability) for the matrix $G = A - LH$, is guaranteed by the condition that the pair (A, H) is ***observable*** *(see, for example, Poznyak (2008)). The observer (or filter) (14.47) with constant gain-matrix L is called the* ***Luenberger-type observer***.

Remark 14.2. *Obviously, if we consider the state estimation on a finite time-interval $[0, T]$ $(T < \infty)$ then property 3 in Definition 14.2 is not essential.*

So, below we will design the gain matrix $L(t)$ such that the mean-square state estimation error $\mathsf{E}\left\{\|x(t,\omega) - \hat{x}(t,\omega)\|^2\right\}$ should be as low as possible within the class of global linear differential observers (14.41) subjected to the constraint (14.51).

14.3.2 Continuous-time linear filtering

14.3.2.1 First two moments of the estimation error

Under noise presence the dynamics of the state estimation error $\Delta x(t,\omega)$ is governed by the linear stochastic differential equation:

$$\boxed{d(\Delta x(t,\omega)) = G(t)\,\Delta x(t,\omega)\,dt + G_v(t)\,dv(t,\omega)} \tag{14.55}$$

where

$$\boxed{\begin{gathered} G(t) = A(t) - L(t)\,H(t) \\ G_v(t) := \left[L(t)\,D(t) \quad [-C(t)]\right] \\ v(t,\omega) := \begin{pmatrix} W_t(\omega) \\ V_t(\omega)\end{pmatrix}\end{gathered}} \tag{14.56}$$

Lemma 14.5. *The first two moments*

$$\boxed{\begin{gathered} m(t) := \mathsf{E}\{\Delta x(t,\omega)\} \\ Q(t) := \mathsf{E}\{[\Delta x(t,\omega)][\Delta x(t,\omega)]^{\mathsf{T}}\}\end{gathered}} \tag{14.57}$$

of $\Delta x\,(t,\omega)$ *satisfy the following ODEs:*

$$\boxed{\dot{m}\,(t) = G\,(t)\,m\,(t), \quad m\,(0) = \mathsf{E}\,\{\Delta x\,(0,\omega)\}} \tag{14.58}$$

and

$$\boxed{\begin{aligned} \dot{Q}\,(t) &= G\,(t)\,Q\,(t) + G\,(t)\,Q^{\mathsf{T}}(t) \\ &\quad + L\,(t)\,D\,(t)\,D^{\mathsf{T}}\,(t)\,L^{\mathsf{T}}\,(t) + C\,(t)\,C^{\mathsf{T}}\,(t) \\ Q\,(0) &= \mathsf{E}\,\{x\,(0,\omega)\,x^{\mathsf{T}}\,(0,\omega)\} \end{aligned}} \tag{14.59}$$

Proof. It directly follows from Corollary 12.4 (see relations (12.85)–(12.88)). □

14.3.2.2 Mean-square gain-matrix optimization

The rigorous derivation of the optimal state-observer (known as the continuous-time Kalman filter) for linear stochastic models (14.36) can be found in Davis (1977). Here we will present another method of derivation which (on our opinion) is more relevant to an engineering audience.

Theorem 14.6. (on the continuous-time Kalman filter) *If*

1. *all matrices* $A\,(t)$, $C\,(t)$, $H\,(t)$ *and* $D\,(t)$ *in (14.36) are continuous almost everywhere (a.e.) on some interval* $[0,T]$*;*
2. *and the assumption A.1 (14.38) holds*

then the best (in mean-square sense) gain-matrix $L\,(t) = L^*\,(t)$ *in the global linear observer (14.47) given by*

$$\boxed{\begin{aligned} d\hat{x}\,(t,\omega) &= \left[A(t)\hat{x}\,(t,\omega) + b(t)\right]dt \\ &\quad + L\,(t)\left[dy_t\,(t,\omega) - H(t)\hat{x}\,(t,\omega)\,dt\right] \\ &\quad \hat{x}_0 \text{ is fixed} \end{aligned}} \tag{14.60}$$

is as follows:

$$\boxed{L\,(t) = L^*\,(t) := P\,(t)\,H^{\mathsf{T}}(t)\left[D\,(t)\,D^{\mathsf{T}}\,(t)\right]^{-1}} \tag{14.61}$$

where $P\,(t)$ *satisfies the following differential Riccati equation*

$$\boxed{\begin{aligned} \dot{P}\,(t) &= A(t)P\,(t) + P\,(t)\,A(t)^{\mathsf{T}} + C\,(t)\,C^{\mathsf{T}}\,(t) \\ &\quad - P\,(t)\,H^{\mathsf{T}}(t)\,[D\,(t)\,D^{\mathsf{T}}\,(t)]^{-1}\,H(t)P\,(t) \\ P\,(0) &= \mathsf{E}\,\{\Delta x\,(0,\omega)\,\Delta^{\mathsf{T}}x\,(0,\omega)\} \end{aligned}} \tag{14.62}$$

so that for all $t \geq 0$

$$\boxed{P(t) = \mathsf{E}\left\{\Delta x(t,\omega)\,\Delta^{\mathsf{T}} x(t,\omega)\right\}} \tag{14.63}$$

Proof.

1. The relation (14.59) can be represented as

$$\begin{aligned}\dot{Q}(t) &= A(t)Q(t) + Q(t)A(t)^{\mathsf{T}} + C(t)C^{\mathsf{T}}(t) \\ &\quad - L(t)H(t)Q(t) - Q(t)[L(t)H(t)]^{\mathsf{T}} + L(t)D(t)D^{\mathsf{T}}(t)L^{\mathsf{T}}(t) \\ &= A(t)Q(t) + Q(t)A(t)^{\mathsf{T}} + C(t)C^{\mathsf{T}}(t) \\ &\quad - Q(t)H^{\mathsf{T}}(t)\left[D(t)D^{\mathsf{T}}(t)\right]^{-1}H(t)Q(t) + \mathcal{R}(t \mid L(t))\end{aligned} \tag{14.64}$$

where the matrix $0 \leq \mathcal{R}(t \mid L(t)) \in \mathbb{R}^{n\times n}$ is

$$\begin{aligned}\mathcal{R}(t \mid L(t)) = &\left(L(t)[D(t)D^{\mathsf{T}}(t)]^{1/2} - Q(t)H^{\mathsf{T}}(t)[D(t)D^{\mathsf{T}}(t)]^{-1/2}\right) \\ &\cdot \left(L(t)[D(t)D^{\mathsf{T}}(t)]^{1/2} - Q(t)H^{\mathsf{T}}(t)[D(t)D^{\mathsf{T}}(t)]^{-1/2}\right)^{\mathsf{T}}\end{aligned}$$

Taking $L(t) = L^{*}(t)$ as in (14.61) provides

$$\mathcal{R}(t \mid L(t)) = 0$$

for all $t \geq 0$. Then (14.64) becomes

$$\begin{aligned}\dot{Q}(t) &= A(t)Q(t) + Q(t)A(t)^{\mathsf{T}} + C(t)C^{\mathsf{T}}(t) \\ &\quad - Q(t)H^{\mathsf{T}}(t)\left[D(t)D^{\mathsf{T}}(t)\right]^{-1}H(t)Q(t) := S(t \mid Q(t))\end{aligned} \tag{14.65}$$

Denote $Q(t) = P(t)$ if it satisfies (14.65).

2. Let us prove an auxiliary statement.

Auxiliary statement. If

1.

$$\boxed{Q(0) \geq P(0)} \tag{14.66}$$

2. all matrices $A(t)$, $C(t)$, $H(t)$ and $D(t)$ in (14.36) are continuous almost everywhere (a.e.) on some interval $[0, T]$,
3. and the assumption A.1 (14.38) holds

then

$$\boxed{Q(t) \geq P(t)} \tag{14.67}$$

for any $L(t)$, that is, for any $z \in \mathbb{R}^{n}$

$$\boxed{\begin{array}{c} z^{\mathsf{T}}\Delta(t)z \geq 0 \\ \Delta^{\mathsf{T}}(t) = \Delta(t) := Q(t) - P(t)\end{array}} \tag{14.68}$$

Proof of auxiliary statement. Since

$$\dot{Q}(t) = S(t \mid Q(t)) + \mathcal{R}(t \mid L(t))$$
$$\mathcal{R}(t \mid L(t)) \geq 0$$
$$\dot{P}(t) = S(t \mid P(t))$$

for a time-partition $0 = t_0 < t_1 < \cdots < t_k < \cdots$ and any $t \in [t_k, t_{k+1})$ we have

$$\begin{aligned}\Delta(t) &:= Q(t) - P(t) = Q(t_k) - P(t_k) \\ &\quad + \int_{s=t_k}^{t} [S(s \mid Q(s)) - S(s \mid P(s))]\, ds + \int_{s=t_k}^{t} \mathcal{R}(s \mid L(s)) ds \\ &\geq Q(t_k) - P(t_k) + \int_{s=t_k}^{t} [S(s \mid Q(s)) - S(s \mid P(s))]\, ds \end{aligned} \tag{14.69}$$

Now we can apply the induction method.
(a) Consider $k = 0$. If $L(t) = L^*(t)$ almost everywhere (a.e.) on $[t_0, t_1)$, then $S(t \mid Q(t)) \overset{a.a.}{=} S(t \mid P(t))$ and (14.69) leads to

$$\Delta(t) \geq \Delta(t_0) \geq 0$$

Consider now the situation when $L(t) \neq L^*(t)$ at some set $\mathcal{T} \subset [t_0, t_1)$ of a non-zero Lebesgue measure. Since $Q(t)$ and $P(t)$ are absolutely continuous and $S(t \mid Q(t))$ is (a.e.)-continuous function of t by the assumptions A.2–A.3, then for any $z \in z \in \mathbb{R}^n$ and small enough (but non-zero) t_1 we have

$$\begin{aligned} &z^{\mathsf{T}} [S(t \mid Q(t)) - S(t \mid P(t))]\, z \\ &\quad = z^{\mathsf{T}} S(t \mid Q(t))\, z - \min_{L(t)} z^{\mathsf{T}} S(t \mid Q(t))\, z \geq 0 \end{aligned}$$

and therefore,

$$\int_{s=t_0}^{t_1} [S(s \mid Q(s)) - S(s \mid P(s))]\, ds \geq 0$$

and hence,

$$\Delta(t) \geq \Delta(t_0) + \int_{s=t_0}^{t_1} [S(s \mid Q(s)) - S(s \mid P(s))]\, ds \geq \Delta(t_0) \geq 0$$

(b) Suppose $\Delta(t_k) \geq 0$ for some $k > 0$. Then repeating the same consideration as for the case $k = 0$ we obtain the desired result. □

Now the proof of the main theorem follows directly if we take into account that under the accepted conditions the *auxiliary statement* holds for any $L(t)$. Theorem is proven. □

Remark 14.3. *In general, the global linear observer (14.47) generates a class of linear estimates $\hat{x}(t,\omega)$ of the state $x(t,\omega)$, based on the measurements $y(\tau,\omega)\,|_{\tau\in[0,t]}$, which minimizes the mean-square estimation error $\mathsf{E}\{\Delta x(t,\omega)\,\Delta^{\mathsf{T}} x(t,\omega)\}$, and, therefore, taking into account that both Wiener processes in (14.36) are Gaussian provided that $x(t,\omega)$ is Gaussian too, we may conclude (see Lemma 14.2) that*

$$\hat{x}(t,\omega) = \mathsf{E}\{x(t,\omega) \mid y(\tau,\omega), \tau \in [0,t]\} \tag{14.70}$$

Example 14.2. (the Ornstein–Uhlenback process) *Consider the Ornstein–Uhlenback process $\{x(t,\omega)\}$ satisfying the following linear stochastic differential equation:*

$$\boxed{\begin{array}{c} dx(t,\omega) = -\alpha x(t,\omega)\,dt + \sigma dV_t(\omega) \\ x(0,\omega) \overset{a.s.}{=} 0 \in \mathbb{R}, \quad \alpha \geq 0,\ \sigma > 0 \\ \mathsf{E}\left\{x_0^2(\omega)\right\} = q_0, \quad t \geq 0 \end{array}} \tag{14.71}$$

provided that the measurable information $\{y(t,\omega)\}$ is generated by

$$\boxed{dy(t,\omega) = x(t,\omega)\,dt + g dW_t(\omega), \quad g > 0} \tag{14.72}$$

Both scalar Wiener processes $V_t(\omega)$ and $W_t(\omega)$ are standard Gaussian independent processes, namely,

$$\begin{aligned} &\mathsf{E}\{V_t(\omega)\} = \mathsf{E}\{W_t(\omega)\} = 0 \\ &\mathsf{E}\left\{V_t^2(\omega)\right\} = \mathsf{E}\left\{W_t^2(\omega)\right\} = t \end{aligned}$$

According to Theorem 14.6, the best state estimate of this process $\hat{x}(t,\omega)$, constructed based on $y(\tau,\omega)\,|_{\tau\in[0,t]}$, is given by

$$\boxed{\begin{aligned} d\hat{x}(t,\omega) &= -\alpha\hat{x}(t,\omega)\,dt \\ &\quad + p(t)\,g^{-2}(t)\left[dy(t,\omega) - \hat{x}(t,\omega)\,dt\right] \\ \dot{p}(t) &= -2\alpha p(t) - p^2(t)\,g^{-2} + \sigma^2 \\ \hat{x}(0,\omega) &\overset{a.s.}{=} 0,\ \ p(0) = q_0 \end{aligned}} \tag{14.73}$$

The function $p(t)$ can be expressed analytically (using the solution for the Bernoulli differential equation) as

$$\boxed{\begin{array}{c} p(t) = p_1 + \dfrac{(p_1 - p_2)}{\dfrac{(q_0 - p_2)}{(q_0 - p_1)}e^{2\beta t} - 1} \\ \beta = \sqrt{\alpha^2 + \sigma^2 g^{-2}} \\ p_1 = g^2(\beta - \alpha), \quad p_2 = -g^2(\beta + \alpha) \end{array}} \tag{14.74}$$

14.3.3 Continuous-time prediction and smoothing

14.3.3.1 Prediction

By (14.36) it follows that for any $\tau > 0$

$$x(t+\tau,\omega) = \Phi(t+\tau,t)\,x(t,\omega) + \Phi(t+\tau,t)\int_{s=t}^{t+\tau}\Phi^{-1}(s,t)\,b(s)\,ds + \Phi(t+\tau,t)\int_{s=t}^{t+\tau}\Phi^{-1}(s,t)\,C(s)\,dV_s(\omega)$$

and, taking into account that the last two terms are independent of $\{y(s,\omega), s \in [0,t]\}$ and in view of (14.70), one has

$$\begin{aligned}\mathsf{E}\{x(t+\tau,\omega) \mid y(s,\omega), s \in [0,t]\} &= \Phi(t+\tau,t)\,\mathsf{E}\{x(t,\omega) \mid y(s,\omega), s \in [0,t]\} \\ &\quad + \Phi(t+\tau,t)\int_{s=t}^{t+\tau}\Phi^{-1}(s,t)\,b(s)\,ds \\ &= \Phi(t+\tau,t)\,\hat{x}(t,\omega) + \Phi(t+\tau,t)\int_{s=t}^{t+\tau}\Phi^{-1}(s,t)\,b(s)\,ds\end{aligned}$$

Here $\Phi(t,s)$ is the transition matrix of (14.36) satisfying for any $t, s \geq 0$

$$\frac{\partial}{\partial t}\Phi(t,s) = A(t)\Phi(t,s), \quad \Phi(t,t) = I \tag{14.75}$$

As a result, we may conclude that the *'best'* (in mean-square sense) *prediction*

$$\hat{x}(t+\tau \mid t) = \mathsf{E}\{x(t+\tau,\omega) \mid y(s,\omega), s \in [0,t]\}$$

is given by

$$\boxed{\hat{x}(t+\tau \mid t) = \Phi(t+\tau,t)\,\hat{x}(t,\omega) + \Phi(t+\tau,t)\int_{s=t}^{t+\tau}\Phi^{-1}(s,t)\,b(s)\,ds} \tag{14.76}$$

where $\Phi(t+\tau,t)$ satisfies (14.75) and $\hat{x}(t,\omega)$ is the Kalman filtering state estimate given by (14.60).

14.3.3.2 Smoothing

Analogously, by (14.36) it follows that for any $\tau > 0$

$$x(t,\omega) = \Phi(t,t-\tau)\, x(t-\tau,\omega) + \Phi(t,t-\tau) \int_{s=t-\tau}^{t} \Phi^{-1}(s,t-\tau)\, b(s)\, ds + \Phi(t,t-\tau) \int_{s=t-\tau}^{t} \Phi^{-1}(s,t-\tau)\, C(s)\, dV_s(\omega)$$

which leads to

$$\begin{aligned} \hat{x}(t,\omega) &= \mathsf{E}\{x(t,\omega) \mid y(s,\omega), s \in [0,t]\} \\ &= \Phi(t,t-\tau)\, \mathsf{E}\{x(t-\tau,\omega) \mid y(s,\omega), s \in [0,t]\} \\ &\quad + \Phi(t,t-\tau) \int_{s=t-\tau}^{t} \Phi^{-1}(s,t-\tau)\, b(s)\, ds \end{aligned}$$

So, the 'best' *smoothing*-estimate

$$\hat{x}(t-\tau,\omega) = \mathsf{E}\{x(t-\tau,\omega) \mid y(s,\omega), s \in [0,t]\}$$

is given by

$$\boxed{\hat{x}(t-\tau,\omega) = \Phi^{-1}(t,t-\tau)\, \hat{x}(t,\omega) - \int_{s=t-\tau}^{t} \Phi^{-1}(s,t-\tau)\, b(s)\, ds} \qquad (14.77)$$

where again $\hat{x}(t,\omega)$ is the Kalman filtering state estimate given by (14.60).

This page intentionally left blank

15 Stochastic Approximation

Contents

15.1 Outline of chapter

In certain statistical applications (such as bioassay, sensitivity testing, or fatigue trials) some problems arising can be conveniently attacked using the so-called *stochastic approximation method* (SAM) which requires minimum distributional assumptions. Traditionally, referring to publications on SAM, the book by Wasan (1969) is mentioned. SAM is closely related to recursive *least squares* (see Chapter 13 of this book) and to the estimation of parameters of a nonlinear regression (Albert and Gardner, 1967). A comprehensive discussion of both stochastic approximation and recursive estimation and their relationship is provided by Nevel'son and Khas'minski (1972). More recent material can be found in Kushner and Yin (1997). The control engineering literature also contains many applications of SAM, basically related to identification problems (see, for example, Tsypkin (1971) and Saridis (1977)).

Quite a large number of stochastic approximation schemes have been discussed in the literature, but they essentially amount to modifications of two basic schemes:

- the *Robbins–Monro procedure* (Robbins and Monro, 1951), dealing with a nonlinear regression problem when only measurements of a regression function corrupted by noise are available,
- the *Kiefer–Wolfowitz procedure* (Kiefer and Wolfowitz, 1952), dealing with an optimization problem when only measurements of a function to be optimized corrupted by noise are available in any predetermined point.

When noise in measurements is an independent process or a martingale-difference, it is standard practice to iterate the approximation procedure and obtain as the end product a martingale plus other terms which are asymptotically negligible.

In this chapter we shall give strong laws of convergence as well as asymptotic normality theorem and iterated logarithm results for both the Robbins–Monro and the Kiefer–Wolfowitz procedures. We also will present some extensions related to the stochastic gradient algorithm, its robustification and the conditions when these procedures work under correlated noises.

15.2 Stochastic nonlinear regression

15.2.1 Nonlinear regression problem

Consider the problem of finding a root of nonlinear vector-function $f : \mathbb{R}^N \to \mathbb{R}^N$ (which is referred to as *a nonlinear regression function*)

$$\boxed{f(x) = 0} \tag{15.1}$$

based on sequence measurements

$$\boxed{y_n = f(x_{n-1}) + \xi_n} \tag{15.2}$$

where

- the points $\{x_n\}_{n\geq 0}$ (where these measurements are realized) can be predetermined by a special rule (an active experiment), suggested by a designer, or can be defined *a priori* (a passive experiment); and
- $\{\xi_n\}_{n\geq 1}$ is a stochastic noise sequence disturbing the observations $\{y_n\}_{n\geq 1}$.

The following assumptions (and their combinations), concerning the regression function, will be in force throughout this section:

A1 The root $x^* \in \mathbb{R}^N$ of the equation (15.1), satisfying

$$\boxed{f(x^*) = 0} \tag{15.3}$$

exists and is unique.

A2 For some positive constants k, K and for all $x \in \mathbb{R}^N$

$$\boxed{k\|x - x^*\|^2 \leq (f(x), x - x^*) \leq K\|x - x^*\|^2} \tag{15.4}$$

A3 For all $x \in \mathbb{R}^N$ as $x \to x^*$

(a)

$$\boxed{\begin{array}{c} f(x) = A_1(x - x^*) + o(\|x - x^*\|) \\ A_1 = \nabla f(x^*), \ \det \nabla f(x^*) \neq 0 \end{array}} \tag{15.5}$$

(b)

$$\boxed{\begin{array}{l} f(x) = A_1(x - x^*) \\ \qquad\qquad + (A_2(x - x^*), (x - x^*)) + o\left(\|x - x^*\|^2\right) \\ \quad A_1 = \nabla f(x^*), \ \det \nabla f(x^*) \neq 0, \quad \|A_2\| < \infty \end{array}} \tag{15.6}$$

The illustration of a nonlinear regression $f(x)$ satisfying the assumption A1–A2 in the scalar case ($N = 1$) is given in Fig. 15.1.

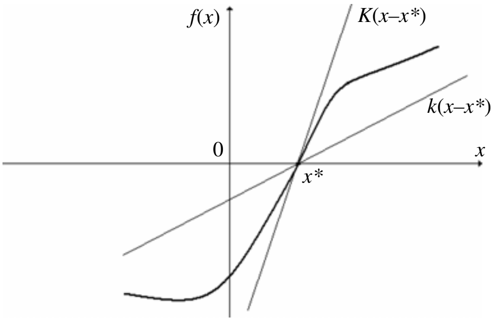

Fig. 15.1. A nonlinear regression function.

15.2.2 Robbins–Monro procedure and convergence with probability one

To find a root of the nonlinear regression (15.1) consider the following recurrent procedure:

$$\boxed{\begin{array}{c} x_n = x_{n-1} - \Gamma_n y_n \\ n = 1, 2, \ldots \end{array}} \tag{15.7}$$

where $\Gamma_n \in \mathbb{R}^{N \times N}$ is the gain matrix referred to as a current 'step-size' of the procedure.

Theorem 15.1. **(Robbins and Monro, 1951)**[1] *Suppose that the assumptions **A1–A2** are satisfied and the noise sequence is a martingale-difference with a bounded conditional covariation, i.e.,*

$$\boxed{\mathsf{E}\{\xi_n \mid \mathcal{F}_{n-1}\} \overset{a.s.}{=} 0, \quad \mathsf{E}\left\{\xi_n \xi_n^\intercal \mid \mathcal{F}_{n-1}\right\} \overset{a.s.}{\leq} \Xi < \infty} \tag{15.8}$$

where $\mathcal{F}_n := \sigma(x_0, \xi_1, \ldots, \xi_n)$. *If*

$$\boxed{\Gamma_n = \gamma_n I} \tag{15.9}$$

with the scalar sequence $\{\gamma_n\}_{n \geq 1}$ *satisfying*

$$\boxed{\gamma_n \leq \frac{2k}{K}(1 - \rho), \quad \sum_{n=1}^{\infty} \gamma_n = 0, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty} \tag{15.10}$$

for some $\rho \in (0, 1)$, *then*

$$\boxed{x_n \underset{n \to \infty}{\overset{a.s.}{\to}} x^*} \tag{15.11}$$

under any fixed initial conditions x_0.

---

[1] In the original paper (Robbins and Monro, 1951) only the scalar case $N = 1$ was considered.

Proof. For $\bar{v}_n := \|x_n - x^*\|^2$ in view of (15.7) it follows that

$$\begin{aligned}\bar{v}_n &= \left\|\left(x_{n-1} - x^*\right) - \gamma_n y_n\right\|^2 \\ &= \bar{v}_{n-1} - 2\gamma_n \left(y_n, x_{n-1} - x^*\right) + \gamma_n^2 \|y_n\|^2\end{aligned} \tag{15.12}$$

Taking from both sides the conditional mathematical expectation and observing that x_{n-1} is $\mathcal{F}_{n-1}$-measurable we get

$$\mathsf{E}\{\bar{v}_n \mid \mathcal{F}_{n-1}\} \le \bar{v}_{n-1} - 2\gamma_n \left(f(x_{n-1}), x_{n-1} - x^*\right) + \gamma_n^2 \left[\|f(x_{n-1})\|^2 + \mathrm{tr}\,\Xi\right]$$

which, in view of A2 and (15.10), implies

$$\begin{aligned}\mathsf{E}\{\bar{v}_n \mid \mathcal{F}_{n-1}\} &\le \bar{v}_{n-1} - 2k\gamma_n \bar{v}_{n-1} + \gamma_n^2 \left[K v_{n-1} + \mathrm{tr}\,\Xi\right] \\ &= \bar{v}_{n-1}\left[1 - 2k\gamma_n\left(1 - \gamma_n \frac{K}{2k}\right)\right] + \gamma_n^2 \mathrm{tr}\,\Xi \\ &\le \bar{v}_{n-1}\left(1 - 2k\rho\gamma_n\right) + \gamma_n^2 \mathrm{tr}\,\Xi\end{aligned} \tag{15.13}$$

Applying then Lemma 7.9 with

$$\eta_n := \bar{v}_n, \quad \lambda_{n+1} := 2k\rho\gamma_n, \quad \nu_n = 0, \quad \theta_n := \gamma_n^2 \mathrm{tr}\,\Xi$$

we conclude (15.11). Theorem is proven. □

Lemma 15.1. *The conditions (15.10) are satisfied, for example, if*

$$\boxed{\gamma_n := \frac{a}{(n+b)^\gamma}, \quad a > 0,\ b \ge 0,\ \frac{1}{2} < \gamma \le 1} \tag{15.14}$$

and the best rate of the convergence (15.11) $n^{\varkappa^*}$ with $\varkappa^* < 1$ is achieved for

$$\boxed{\gamma = \gamma^* = 1, \quad \frac{1}{2k\rho} < \gamma_0} \tag{15.15}$$

namely,

$$\boxed{\|x_n - x^*\|^2 \overset{a.s.}{=} o_\omega\left(\frac{1}{n^{\varkappa^*}}\right)} \tag{15.16}$$

Proof. It follows directly from Lemma 7.12 with

$$v_n := n^\varkappa, \quad \alpha_n := 2k\rho\gamma_n, \quad \beta_n := \gamma_n^2 \mathrm{tr}\,\Xi$$

being applied to (15.13) if we notice that

$$\begin{aligned}\frac{v_{n+1} - v_n}{\alpha_n v_n} &= \frac{(n+1)^\varkappa - n^\varkappa}{(2k\rho\gamma_n)\, n^\varkappa} = \frac{(n+b)^\gamma}{2k\rho\gamma_0}\left(\frac{\varkappa + o(n^{-1})}{n}\right) \\ &= \frac{\varkappa}{2k\rho\gamma_0}\left(\frac{1 + o(n^{-1})}{n^{1-\gamma}}\right) \to \mu = \begin{cases} 0 & \text{if} \quad \gamma < 1 \\ \dfrac{\varkappa}{2k\rho\gamma_0} & \text{if} \quad \gamma = \gamma^* = 1 \end{cases}\end{aligned}$$

and

$$\sum_{n=1}^{\infty} \beta_n \upsilon_n = a^2 \text{tr}\, \Xi \sum_{n=1}^{\infty} \frac{n^{\varkappa}}{(n+b)^{2\gamma}} \leq a^2 \text{tr}\, \Xi \sum_{n=1}^{\infty} \frac{1}{n^{2\gamma-\varkappa}} < \infty$$

if $2\gamma - \varkappa > 1$, or equivalently, when $\varkappa < 2\gamma - 1$. Therefore, maximum admissible $\varkappa$ is

$$\varkappa = \varkappa^* < 2\gamma^* - 1 = 1$$

Lemma is proven. □

Theorem 15.2. *The Theorem 15.1 remains valid if instead of (15.9) we take*

$$\boxed{\begin{array}{c} \Gamma_n = \gamma_n \Gamma, \quad 0 < \Gamma = \Gamma^{\intercal} \in \mathbb{R}^{N \times N} \\ \gamma_n \leq \dfrac{k}{K} (1-\rho) \end{array}} \tag{15.17}$$

Proof. It practically repeats the proof of Theorem 15.1 with only one difference: instead of $\bar{\upsilon}_n := \|x_n - x^*\|^2$ one can consider

$$\bar{\upsilon}_n (\Gamma) := \left\| x_n - x^* \right\|_{\Gamma^{-1}}^2$$

which leads to the following relations:

$$\begin{aligned} \bar{\upsilon}_n (\Gamma) &= \left\| \left(x_{n-1} - x^*\right) - \gamma_n \Gamma y_n \right\|_{\Gamma^{-1}}^2 = \bar{\upsilon}_{n-1} (\Gamma) \\ &\quad - 2\gamma_n \left(\Gamma^{-1} \Gamma y_n, x_{n-1} - x^*\right) + \gamma_n^2 \left\| \Gamma y_n \right\|_{\Gamma^{-1}}^2 \\ &= \bar{\upsilon}_{n-1} (\Gamma) - 2\gamma_n \left(y_n, x_{n-1} - x^*\right) + \gamma_n^2 y_n^{\intercal} \Gamma y_n \end{aligned}$$

and, therefore,

$$\begin{aligned} \mathsf{E} \{\bar{\upsilon}_n (\Gamma) \mid \mathcal{F}_{n-1}\} &\leq \bar{\upsilon}_{n-1} (\Gamma) - 2\gamma_n \left(f(x_{n-1}), x_{n-1} - x^*\right) \\ &\quad + 2\gamma_n^2 \left[\| f(x_{n-1}) \|^2 + \text{tr}\{\Xi\Gamma\} \right] \\ &\leq \bar{\upsilon}_{n-1} (\Gamma) - 2k\gamma_n \bar{\upsilon}_{n-1} (\Gamma) + 2\gamma_n^2 \left[K \upsilon_{n-1} + \text{tr}\{\Xi\Gamma\} \right] \\ &= \bar{\upsilon}_{n-1} (1 - 2k\rho\gamma_n) + \gamma_n^2 \text{tr}\{\Xi\Gamma\} \end{aligned} \tag{15.18}$$

Then the result follows directly from Lemma 7.9 with

$$\eta_n := \bar{\upsilon}_n, \quad \lambda_{n+1} := 2k\rho\gamma_n, \quad \nu_n = 0, \quad \theta_n := \gamma_n^2 \text{tr}\{\Xi\Gamma\}$$

Theorem is proven. □

Corollary 15.1. *Under the conditions of Theorem 15.2*

$$\boxed{\mathsf{E}\{\bar{\upsilon}_n (\Gamma)\} \leq O(\nu_n^{-1})} \tag{15.19}$$

where the numerical sequence $\{\nu_n\}_{n\geq 1}$ satisfies the conditions

$$\boxed{\begin{gathered} 0 < \nu_n < \nu_{n+1} \\ \lim_{n\to\infty} \frac{\nu_n - \nu_{n-1}}{\tilde{\alpha}\nu_{n-1}} = \mu < 1, \quad \limsup_{n\to\infty} \nu_n \gamma_n = p \end{gathered}} \tag{15.20}$$

Proof. Taking the mathematical expectation from both sides of (15.18) leads to

$$\begin{aligned} &\mathsf{E}\{\bar{v}_n(\Gamma)\} \leq \mathsf{E}\{\bar{v}_{n-1}(\Gamma)\}(1-\tilde{\alpha}_n) + \tilde{\beta}_n \\ &\tilde{\alpha}_n := 2k\rho\gamma_n, \ \tilde{\beta}_n := \gamma_n^2 \mathrm{tr}\{\Xi\Gamma\} \end{aligned}$$

For $w_n := \nu_n \mathsf{E}\{\bar{v}_n(\Gamma)\}$ it follows (for large enough n)

$$\begin{aligned} w_n &\leq w_{n-1}(1-\tilde{\alpha}_n)\frac{\nu_n}{\nu_{n-1}} + \nu_n\tilde{\beta}_n \\ &= w_{n-1}(1-\tilde{\alpha}_n)\left(1 + \frac{\nu_n - \nu_{n-1}}{\nu_{n-1}}\right) + \nu_n\tilde{\beta}_n \\ &= w_{n-1}\left(1 - \tilde{\alpha}_n + \tilde{\alpha}_n\left[\frac{\nu_n - \nu_{n-1}}{\tilde{\alpha}\nu_{n-1}}\right] - \tilde{\alpha}_n\frac{\nu_n - \nu_{n-1}}{\nu_{n-1}}\right) + \nu_n\tilde{\beta}_n \\ &= w_{n-1}\left(1 - \tilde{\alpha}_n\left[1 - [\mu + o(1)](1-\tilde{\alpha}_n)\right]\right) + \nu_n\tilde{\beta}_n \\ &= w_{n-1}(1 - \tilde{\alpha}_n[1 - \mu + o(1)]) + \nu_n\tilde{\beta}_n \end{aligned}$$

By Lemma 16.14 from Poznyak (2008) it follows that

$$\limsup_{n\to\infty} w_n \leq \limsup_{n\to\infty} \frac{\nu_n\tilde{\beta}_n}{\tilde{\alpha}_n(1-\mu)} = \mathsf{Const}\cdot p$$

which implies (15.19). Corollary is proven. □

15.2.3 Asymptotic normality

From this point we will consider the class of the gain matrices

$$\boxed{\Gamma_n = \frac{\Gamma}{n+b}, \quad 0 < \Gamma = \Gamma^{\mathsf{T}} \in \mathbb{R}^{N\times N}, \ b \geq 0} \tag{15.21}$$

Lemma 15.2. (on $\sqrt{n}$-equivalence) *If*

1.

$$\|x_n - x^*\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

*2. the assumption **A3**(a) holds,*

3. Γ_n *is as in (15.21) with* $b = 0$*, and, additionally,*

$$\boxed{\lambda_- := \lambda_{\min}\left(A_1^{\intercal}\Gamma + \Gamma A_1\right) > 0} \tag{15.22}$$

then the process $\{x_n\}_{n\geq 1}$*, generated by (15.7), and the linear recurrence*

$$\boxed{\begin{array}{c} \tilde{x}_n = \tilde{x}_{n-1} - \Gamma_n\left[A_1\left(\tilde{x}_n - x^*\right) + \xi_n\right] \\ \tilde{x}_0 = x_0, \quad A_1 = \nabla f(x^*) \end{array}} \tag{15.23}$$

are $\sqrt{n}$***-equivalent****, namely,*

$$\boxed{\sqrt{n}\left(x_n - \tilde{x}_n\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0} \tag{15.24}$$

Proof. One has

$$\begin{aligned} w_n &:= \sqrt{n}\left(x_n - \tilde{x}_n\right) \\ &= \sqrt{n}\left(x_{n-1} - \tilde{x}_{n-1} - \Gamma_n\left[f\left(x_{n-1}\right) - A_1\left(\tilde{x}_n - x^*\right)\right]\right) \\ &= \sqrt{n}\left(x_{n-1} - \tilde{x}_{n-1} - \Gamma_n\left(A_1\left[x_{n-1} - \tilde{x}_n\right] + o\left(\left\|x_n - x^*\right\|\right)\right)\right) \\ &= \sqrt{n}\left[I - \Gamma_n A_1\right]\left[x_{n-1} - \tilde{x}_n\right] + \sqrt{n}\Gamma_n o\left(\left\|x_n - x^*\right\|\right) \\ &= \frac{\sqrt{n}}{\sqrt{n-1}}\left[I - \Gamma_n A_1\right] w_{n-1} + \sqrt{n}\Gamma_n o\left(\left\|x_n - x^*\right\|\right) \end{aligned}$$

The approximation

$$\frac{\sqrt{n}}{\sqrt{n-1}} = \sqrt{1 + \frac{1}{n-1}} = 1 + \frac{1 + o(n^{-1})}{2n}$$

and the representation

$$o\left(\left\|x - x^*\right\|\right) = \left\|x - x^*\right\| o(1)$$

imply

$$\begin{aligned} w_n &= \left[1 + \frac{1 + o(n^{-1})}{2n}\right]\left[I - \frac{\Gamma}{n+b}A_1\right] w_{n-1} + \frac{\sqrt{n}\Gamma}{n+b} o\left(\left\|x_n - x^*\right\|\right) \\ &= \left[I - \frac{\Gamma\left(1 + o(n^{-1})\right)}{n}A_1 - \frac{1 + o(n^{-1})}{2n^2}\Gamma A_1\right] w_{n-1} + \frac{\Gamma}{n} w_{n-1} o(1) \\ &= \left[I - \frac{\Gamma A_1}{n}\left(1 + o(1)\right)\right] w_{n-1} \end{aligned}$$

and, hence,

$$\left\|w_n\right\|^2 \leq \left\|I - \frac{\Gamma A_1}{n}\left(1 + o(1)\right)\right\|^2 \left\|w_{n-1}\right\|^2$$

where

$$\begin{aligned}
&\left\| I - \frac{\Gamma}{n}(A_1 + o(1)) \right\|^2 \\
&\quad = \lambda_{\max}\left(\left[I - \frac{\Gamma A_1}{n}(1 + o(1)) \right]^{\intercal} \left[I - \frac{\Gamma A_1}{n}(1 + o(1)) \right] \right) \\
&\quad = \lambda_{\max}\left(\left[I - \frac{1 + o(1)}{n}\left[A_1^{\intercal}\Gamma + \Gamma A_1 \right] + \frac{A_1^{\intercal}\Gamma^2 A_1}{n^2}(1 + o(1))^2 \right] \right) \\
&\quad \le \lambda_{\max}\left(\left[I - \frac{1 + o(1)}{n}\left[A_1^{\intercal}\Gamma + \Gamma A_1 \right] \right] \right) + O\left(\frac{1}{n^2} \right) \\
&\quad = 1 - \frac{1 + o(1)}{n}\lambda_{\min}\left(A_1^{\intercal}\Gamma + \Gamma A_1 \right) + O\left(\frac{1}{n^2} \right) = 1 - \frac{\lambda_- + o(1)}{n}
\end{aligned}$$

Using the inequality $1 + x \le e^x$ valid for any $x \in \mathbb{R}$ we get

$$1 - \frac{\lambda_- + o(1)}{n} \le \exp\left\{ -\frac{\lambda_- + o(1)}{n} \right\}$$

which implies

$$\begin{aligned}
\|w_n\|^2 &\le \left(1 - \frac{\lambda_- + o(1)}{n} \right) \|w_{n-1}\|^2 \\
&\le \prod_{k=n_0}^{n} \left(1 - \frac{\lambda_- + o(1)}{k} \right) \|w_{k-1}\|^2
\end{aligned}$$

Select n_0 large enough such that for all $n \ge n_0$

$$\lambda_- + o(1) \ge c > 0$$

Then the last inequalities imply $(a.s.)$

$$\|w_n\|^2 \le \exp\left\{ -\sum_{k=n_0}^{n} \frac{\lambda_- + o(1)}{k} \right\} \le \exp\left\{ -c\sum_{k=n_0}^{n} \frac{1}{k} \right\} \underset{n\to\infty}{\to} 0$$

which completes the proof. Lemma is proven. □

The lemma above permits us to analyze the *rate of convergence* of the original procedure (15.7) using the $\sqrt{n}$*-equivalent procedure* (15.23):

$$\boxed{\tilde{x}_n = \tilde{x}_{n-1} - \frac{1}{n}\Gamma\left[\nabla f(x^*)(\tilde{x}_n - x^*) + \xi_n \right]} \tag{15.25}$$

For $\tilde{\Delta}_n := \tilde{x}_n - x^*$ the recurrence (15.25) gives

$$\tilde{\Delta}_n = \left[I - \frac{1}{n}\Gamma\nabla f(x^*) \right] \tilde{\Delta}_{n-1} - \frac{1}{n}\Gamma\xi_n \tag{15.26}$$

so that, for $\theta_n = \sqrt{n}\tilde{\Delta}_n$ it follows

$$\begin{aligned}
\theta_n &= \sqrt{1+\frac{1}{n}}\left[I - \frac{1}{n}\Gamma\nabla f(x^*)\right]\theta_{n-1} - \frac{1+O\left(n^{-1}\right)}{\sqrt{n}}\Gamma\xi_n \\
&= \left(1+\frac{1+O\left(n^{-1}\right)}{2n}\right)\left[I - \frac{1}{n}\Gamma\nabla f(x^*)\right]\theta_{n-1} - \frac{1+O\left(n^{-1}\right)}{\sqrt{n}}\Gamma\xi_n \\
&= \left(I+\frac{1+O\left(n^{-1}\right)}{n}\left[\frac{1}{2}I - \Gamma\nabla f(x^*)\right]\right)\theta_{n-1} - \frac{1+O\left(n^{-1}\right)}{\sqrt{n}}\Gamma\xi_n
\end{aligned}$$

or, in another form,

$$\begin{aligned}
\theta_n &= B_n\theta_{n-1} + \zeta_n \\
B_n &:= I + \frac{1+O\left(n^{-1}\right)}{n}\left[\frac{1}{2}I - \Gamma\nabla f(x^*)\right] \\
\tilde{\xi}_n &:= -\frac{1+O\left(n^{-1}\right)}{\sqrt{n}}\Gamma\xi_n
\end{aligned}$$

Iterating back gives

$$\begin{aligned}
\theta_n &= \left(\prod_{t=n_0}^{n} B_t\right)\theta_{n_0-1} + \sum_{t=n_0}^{n}\left(\prod_{s=t+1}^{n} B_t\right)\tilde{\xi}_t \\
&= \prod_{t=n_0}^{n} B_t\theta_{n_0-1} + \frac{1}{\sqrt{n}}\left(\sqrt{n}\prod_{t=n_0}^{n} B_t\right)\sum_{t=n_0}^{n}\sqrt{t}\left(\sqrt{t}\prod_{s=n_0}^{t} B_t\right)^{-1}\tilde{\xi}_t \\
&= \prod_{t=n_0}^{n} B_t\theta_{n_0-1} + \frac{1}{\sqrt{n}}A_{n,n_0}^{-1}\sum_{t=n_0}^{n} A_{t,n_0}\zeta_n
\end{aligned}$$

where

$$A_{n,n_0}^{-1} := \sqrt{n}\prod_{t=n_0}^{n} B_t, \quad \zeta_n := \sqrt{n}\tilde{\xi}_n = -\left(1+O\left(n^{-1}\right)\right)\Gamma\xi_n$$

Suppose that

A4 the matrix

$$\boxed{S := \left[\frac{1}{2}I - \Gamma\nabla f(x^*)\right]} \tag{15.27}$$

is Hurwitz (stable).

Then $\prod_{t=n_0}^{n} B_t \to 0$ when $n \to \infty$, and therefore for some $\alpha > 0$

$$\boxed{\theta_n = \frac{1}{\sqrt{n}} A_{n,n_0}^{-1} \sum_{t=n_0}^{n} A_{t,n_0} \zeta_n + O_\omega\left(e^{-\alpha n}\right)} \tag{15.28}$$

Now we are ready to formulate the main result on the asymptotic normality of the procedure (15.7).

Theorem 15.3. (on asymptotic normality) *If under the assumptions* ***A1***, ***A2*** *and* ***A3****(a) in the stochastic nonlinear regression procedure (15.7) with the gain matrix* $\Gamma_n = \dfrac{\Gamma}{n+b}$ *as in (15.21) where the matrix* Γ *satisfies*

$$\boxed{\begin{array}{c} \lambda_- := \lambda_{\min}\left(\nabla^\intercal f(x^*)\Gamma + \Gamma^\intercal \nabla f(x^*)\right) > 0 \\ S := \left[\dfrac{1}{2} I - \Gamma \nabla f(x^*)\right] \text{ is Hurwitz} \end{array}} \tag{15.29}$$

there exists a limit

$$\boxed{R := \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathsf{E}\left\{\xi_t \xi_t^\intercal\right\} > 0} \tag{15.30}$$

then

$$\boxed{\sqrt{n}\left(x_n - x^*\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, K)} \tag{15.31}$$

where the matrix $K = K^\intercal > 0$ *is the solution of the following algebraic matrix Lyapunov equation:*

$$\boxed{SK + KS^\intercal = -\Gamma R \Gamma^\intercal} \tag{15.32}$$

Proof. Let us verify condition 1 of Lemma 8.11:

$$\begin{aligned} A &:= \lim_{n\to\infty} n\left(A_n^{-1} A_{n-1} - I\right) \\ &= \lim_{n\to\infty} n\left(\sqrt{n} \prod_{t=n_0}^{n} B_t \frac{1}{\sqrt{n-1}} \left(\prod_{t=n_0}^{n-1} B_t\right)^{-1} - I\right) = \lim_{n\to\infty} n\left(\sqrt{\frac{n}{n-1}} B_n - I\right) \\ &= \lim_{n\to\infty} n\left(\left(1 + \frac{1 + O\left(n^{-1}\right)}{2n}\right)\left(I + \frac{1 + O\left(n^{-1}\right)}{n}\left[\frac{1}{2} I - \Gamma \nabla f(x^*)\right]\right) - I\right) \end{aligned}$$

$$= \lim_{n\to\infty}\left(\left(\frac{1+O\left(n^{-1}\right)}{2}\right)I + \left(1+O\left(n^{-1}\right)\right)\left[\frac{1}{2}I - \Gamma\nabla f(x^*)\right]\right)$$
$$= I - \Gamma\nabla f(x^*)$$

and therefore, by assumption **A4** the matrix

$$A - \frac{1}{2}I = \frac{1}{2}I - \Gamma\nabla f(x^*) = S$$

is Hurwitz. So, condition 1 of Lemma 8.11 is fulfilled. Condition 2 is fulfilled too by the assumption of this theorem concerning the properties of $\{\xi_n\}_{n\geq 1}$. Then the desired result follows directly from Theorem 8.15 and Lemma 8.11. Theorem is proven. □

15.2.4 Logarithmic iterative law

Theorem 15.4. (on LIL) *If in Theorem 15.3 the condition* ***A3****(a) is strengthened to* ***A3****(b) then for large enough n*

$$\boxed{x_n = x^* + K^{1/2}\bar{\zeta}_n(\omega)\sqrt{2n^{-1}\ln\ln n}} \tag{15.33}$$

where $\bar{\zeta}_n(\omega)$ is a random vector that has its set of limit points confined in n-dimensional sphere $\{x : \|x\| \leq 1\}$ such that

$$\boxed{\limsup_{n\to\infty}\left\|\bar{\zeta}_n(\omega)\right\| \overset{a.s.}{=} 1} \tag{15.34}$$

Proof. Here we do not give a complete detailed proof, but consider this result only as an illustration of Theorem 8.16 in Chapter 8 (see details in Hall and Heyde (1980)). □

15.2.5 Step-size optimization

As follows from (15.32) the matrix K, defining the rate of convergence of the procedure (15.7)–(15.21) depends on the selection of matrix Γ, namely,

$$\boxed{K = K(\Gamma)} \tag{15.35}$$

and to optimize the convergence rate one has to select the step-size parameter Γ to try to solve the matrix minimization problem:

$$\boxed{\begin{array}{c} K(\Gamma) \to \min\limits_{\Gamma} \\ \text{subject to (15.29)} \end{array}} \tag{15.36}$$

The following lemma gives the solution of this problem in the case of stationary noises, and therefore, defines the optimal parametric matrix $\Gamma = \Gamma^*$ defining the best form of the stochastic nonlinear regression procedure (15.7)–(15.21).

Lemma 15.3. *Under conditions of Theorem 15.3 when*

$$R = \Xi_\xi := \mathsf{E}\left\{\xi_n \xi_n^\intercal\right\} > 0 \quad \textit{for any} \quad n = 1, 2, \ldots$$

for any admissible Γ *satisfying (15.29) the following matrix inequality holds:*

$$\boxed{K(\Gamma) \geq K^* = \left[\nabla f(x^*)\right]^{-1} \Xi_\xi \left[\nabla^\intercal f(x^*)\right]^{-1}} \tag{15.37}$$

where the equality is achieved for

$$\boxed{\Gamma^* = \left[\nabla f(x^*)\right]^{-1}} \tag{15.38}$$

so that $K^* = K(\Gamma^*)$.

Proof. For the nonnegative definite matrix H, equal to

$$\begin{aligned} 0 \leq H &:= \int\limits_{t \in T} \begin{pmatrix} F(t) \\ G(t) \end{pmatrix} \begin{pmatrix} F(t) \\ G(t) \end{pmatrix}^\intercal dt \\ &= \begin{pmatrix} \int\limits_{t \in T} F(t) F(t)^\intercal dt & \int\limits_{t \in T} F(t) G(t)^\intercal dt \\ \int\limits_{t \in T} G(t) F(t)^\intercal dt & \int\limits_{t \in T} G(t) G(t)^\intercal dt \end{pmatrix} \end{aligned} \tag{15.39}$$

where $F(t) \in \mathbb{R}^{N_F \times M}$, $G(t) \in \mathbb{R}^{N_G \times M}$ are functional quadratically integrable matrices, by the Schur complement (see Lemma 13.5) application, it follows that

$$\begin{aligned} &\int\limits_{t \in T} F(t) F(t)^\intercal dt \\ &\quad \geq \int\limits_{t \in T} F(t) G(t)^\intercal dt \left(\int\limits_{t \in T} G(t) G(t)^\intercal dt \right)^{-1} \int\limits_{t \in T} G(t) F(t)^\intercal dt \end{aligned} \tag{15.40}$$

provided that $\int\limits_{t \in T} G(t) G(t)^\intercal dt > 0$. Now notice that the solution K of the matrix Lyapunov equation can be expressed (see, for example, Lemma 9.1 in Poznyak (2008)) as

$$K = \int\limits_{t=0}^{\infty} e^{St} \Gamma \Xi_\xi \Gamma^\intercal e^{S^\intercal t} dt$$

(here we have used the fact that the matrix S is Hurwitz). Putting

$$F(t) := e^{St} \Gamma \Xi_\xi^{1/2} \quad \text{and} \quad G(t) := e^{-\frac{1}{2} I t} \Xi_\xi^{-1/2}$$

in (15.40) we obtain

$$K = \int_{t=0}^{\infty} F(t) F^{\intercal}(t) dt$$
$$\geq \int_{t=0}^{\infty} e^{St} \Gamma e^{-\frac{1}{2}It} dt \left(\int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Xi_{\xi}^{-1} e^{-\frac{1}{2}It} dt \right)^{-1} \int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Gamma^{\intercal} e^{S^{\intercal}t} dt$$

Since Γ and $e^{-\frac{1}{2}It}$ commute, i.e.,

$$\Gamma e^{-\frac{1}{2}It} = e^{-\frac{1}{2}It} \Gamma$$

and applying the formula

$$\int_{t=0}^{\infty} e^{At} dt = -A^{-1} \quad \text{(if } A \text{ is Hurwitz)}$$

we get

$$K = \int_{t=0}^{\infty} F(t) F^{\intercal}(t) dt$$
$$\geq \int_{t=0}^{\infty} e^{\left(S - \frac{1}{2}I\right)t} dt \Gamma \left(\int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Xi_{\xi}^{-1} e^{-\frac{1}{2}It} dt \right)^{-1} \Gamma^{\intercal} \int_{t=0}^{\infty} e^{\left(S - \frac{1}{2}I\right)^{\intercal} t} dt$$
$$= \left(S - \frac{1}{2}I\right)^{-1} \Gamma \left(\int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Xi_{\xi}^{-1} e^{-\frac{1}{2}It} dt \right)^{-1} \Gamma^{\intercal} \left(S^{\intercal} - \frac{1}{2}I\right)^{-1}$$
$$= \left(\Gamma \nabla f(x^*)\right)^{-1} \Gamma \left(\int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Xi_{\xi}^{-1} e^{-\frac{1}{2}It} dt \right)^{-1} \Gamma^{\intercal} \left(\nabla^{\intercal} f(x^*) \Gamma^{\intercal}\right)^{-1}$$
$$= \left(\nabla f(x^*)\right)^{-1} \left(R^*\right)^{-1} \left[\nabla^{\intercal} f(x^*)\right]^{-1}$$

where

$$R^* = \int_{t=0}^{\infty} e^{-\frac{1}{2}It} \Xi_{\xi}^{-1} e^{-\frac{1}{2}It} dt$$

is the solution of the matrix Lyapunov equation

$$\left(-\frac{1}{2}I\right) R^* + R^* \left(-\frac{1}{2}I\right)^{\intercal} = -\Xi_{\xi}^{-1}$$

which can be found analytically:

$$R^* = \Xi_\xi^{-1}$$

Therefore,

$$K \geq \left(\nabla f(x^*)\right)^{-1} \left(\Xi_\xi^{-1}\right)^{-1} \left[\nabla^\intercal f(x^*)\right]^{-1} = \left(\nabla f(x^*)\right)^{-1} \Xi_\xi \left[\nabla^\intercal f(x^*)\right]^{-1}$$

which proves (15.37). Direct substitution (15.38) into (15.32) leads to $K^* = K(\Gamma^*)$ which concludes the proof. □

Now one can make the following conclusion.

Conclusion 15.1. *The **best convergence rate** of the procedure (15.7)–(15.21) is achieved for $\Gamma = \Gamma^*$ (15.38) and this procedure is*

$$\boxed{x_n = x_{n-1} - \frac{\left[\nabla f(x^*)\right]^{-1}}{n} y_n} \tag{15.41}$$

*Obviously, it is **unrealizable** since x^* and, hence, $\nabla f(x^*)$ is a priori unknown. The question is: 'Is it really possible to find any realizable approximation of this optimal procedure?' The **positive answer** was given in Ruppert (1988) and rigorously proven in Polyak (1990). The next subsection introduces the reader to the details of this problem.*

15.2.6 Ruppert–Polyak version with averaging

Consider the following version of the recurrent procedure (15.7)–(15.21) with the additional averaging

$$\boxed{\begin{array}{c} x_n = x_{n-1} - \dfrac{\Gamma}{n} y_n, \quad n = 1, 2, \ldots \\ \bar{x}_n := \dfrac{1}{n} \displaystyle\sum_{t=1}^{n} x_t = \bar{x}_{n-1} - \dfrac{1}{n} \left(\bar{x}_{n-1} - x_n\right) \end{array}} \tag{15.42}$$

Suppose that the conditions of Lemma 15.2 are fulfilled. Therefore, we have the convergence with probability one, i.e.,

$$\left\| x_n - x^* \right\| \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

and, moreover, we know that

$$\sqrt{n}\left(x_n - \tilde{x}_n\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

where the process $\{\tilde{x}_n\}_{n\geq 1}$ is generated by (15.25):

$$\tilde{x}_n = \tilde{x}_{n-1} - \frac{1}{n}\Gamma\left[\nabla f(x^*)\left(\tilde{x}_n - x^*\right) + \xi_n\right]$$

So, for $\Delta_n := x_n - x^*$, $\tilde{\Delta}_n := \tilde{x}_n - x^*$ and $\bar{\Delta}_n := \bar{x}_n - x^*$ we have (see (15.26))

$$\tilde{\Delta}_n = \left[I - \frac{1}{n}\Gamma\nabla f(x^*) \right] \tilde{\Delta}_{n-1} - \frac{1}{n}\Gamma\xi_n$$
$$\bar{\Delta}_n := \frac{1}{n}\sum_{t=1}^{n}\Delta_t = \bar{\Delta}_{n-1}\left(1 - \frac{1}{n}\right) + \frac{1}{n}\Delta_n$$

Taking into account that

$$\sqrt{n}\left(\Delta_n - \tilde{\Delta}_n\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

we derive that

$$\frac{1}{n}\Delta_n = \frac{1}{n}\left(\Delta_n - \tilde{\Delta}_n\right) + \frac{1}{n}\tilde{\Delta}_n \overset{a.s.}{=} \frac{1}{n}\tilde{\Delta}_n + \frac{o_\omega(1)}{(n)^{3/2}}$$

and, hence,

$$\boxed{\begin{aligned} &\tilde{\Delta}_n = \left[I - \frac{1}{n}\Gamma\nabla f(x^*) \right] \tilde{\Delta}_{n-1} - \frac{1}{n}\Gamma\xi_n \\ &\bar{\Delta}_n := \frac{1}{n}\sum_{t=1}^{n}\Delta_t = \frac{1}{n}\sum_{t=1}^{n} t\left(\frac{1}{t}\Delta_t\right) \\ &\quad = \frac{1}{n}\sum_{t=1}^{n} t\left(\frac{1}{t}\tilde{\Delta}_t + \frac{o_\omega(1)}{(t)^{3/2}}\right) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\Delta}_t + \frac{1}{n}\sum_{t=1}^{n}\frac{o_\omega(1)}{(t)^{3/2}} \end{aligned}} \tag{15.43}$$

One can see that by the Kronecker Lemma 8.3

$$\sqrt{n}\bar{\Delta}_n - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sqrt{t}\left(\frac{o_\omega(1)}{t^2}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

Hence,

Proposition 15.1. *To analyze the properties of the sequence*

$$\left\{\sqrt{n}\bar{\Delta}_n\right\}_{n\geq 1}$$

it is sufficient to consider instead the sequence

$$\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t\right\}_{n\geq 1}$$

The next theorem analyzes the property of this last sequence.

Theorem 15.5. *Under the conditions of Lemma 15.2 for any admissible matrices* Γ *the following* $\sqrt{n}$*-equivalence holds:*

$$\boxed{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t-\breve{\Delta}_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0} \tag{15.44}$$

where

$$\boxed{\breve{\Delta}_n := -\left[\nabla f(x^*)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_t} \tag{15.45}$$

Proof. The relations (15.43) imply

$$\tilde{\Delta}_{n-1} = n\left[\Gamma\nabla f(x^*)\right]^{-1}\left(\tilde{\Delta}_{n-1}-\tilde{\Delta}_n\right)-\left[\nabla f(x^*)\right]^{-1}\xi_n$$

and therefore,

$$\begin{aligned}\tilde{\theta}_n &:= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t-\breve{\Delta}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\Delta}_t+\left[\nabla f(x^*)\right]^{-1}\xi_t\right)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\Delta}_{t-1}+\left[\nabla f(x^*)\right]^{-1}\xi_t\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\Delta}_t-\tilde{\Delta}_{t-1}\right)\\ &= \left[\Gamma\nabla f(x^*)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}t\left(\tilde{\Delta}_{t-1}-\tilde{\Delta}_t\right)+\frac{1}{\sqrt{n}}\left(\tilde{\Delta}_n-\tilde{\Delta}_0\right)\end{aligned}$$

Show that

$$\tilde{\theta}_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0 \tag{15.46}$$

By Theorem 15.2 $\tilde{\Delta}_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ which implies

$$\frac{1}{\sqrt{n}}\left(\tilde{\Delta}_n-\tilde{\Delta}_0\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0$$

So, to prove (15.46) it is sufficient to show that

$$\tilde{\theta}'_n := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}t\left(\tilde{\Delta}_{t-1}-\tilde{\Delta}_t\right) \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0 \tag{15.47}$$

Integration (summation) by part, applying the Abel identity (see Lemma 12.2 in Poznyak (2008)), gives

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}t\left(\tilde{\Delta}_{t-1}-\tilde{\Delta}_t\right)\\ &\quad=\sqrt{n}\sum_{t=1}^{n}\left(\tilde{\Delta}_{t-1}-\tilde{\Delta}_t\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[t-(t-1)]\sum_{s=1}^{t-1}\left(\tilde{\Delta}_{s-1}-\tilde{\Delta}_t\right)\\ &\sqrt{n}\left(\tilde{\Delta}_0-\tilde{\Delta}_n\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\Delta}_0-\tilde{\Delta}_t\right)=-\sqrt{n}\tilde{\Delta}_n+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t\end{aligned}$$

Then, by the Chebyshev inequality (4.9) and in view of (15.19) for $\nu_n = n$, we derive

$$\mathsf{P}\left\{\sqrt{n}\left\|\tilde{\Delta}_n\right\| > \varepsilon\right\} = \mathsf{P}\left\{\left\|\tilde{\Delta}_n\right\| > \frac{\varepsilon}{\sqrt{n}}\right\} \le \frac{n}{\varepsilon^2}\mathsf{E}\left\{\left\|\tilde{\Delta}_n\right\|^2\right\}$$

Since by (15.19) $\mathsf{E}\left\{\left\|\tilde{\Delta}_n\right\|^2\right\} = O\left(n^{-1}\right)$ for any small enough $\epsilon > 0$ there exists a number $n_0(\epsilon)$ such that for all $n \ge n_0(\epsilon)$ we have

$$\mathsf{E}\left\{\left\|\tilde{\Delta}_n\right\|^2\right\} \le \epsilon n^{-1} \tag{15.48}$$

we may conclude that

$$\mathsf{P}\left\{\sqrt{n}\left\|\tilde{\Delta}_n\right\| > \varepsilon\right\} \le \frac{n}{\varepsilon^2}\mathsf{E}\left\{\left\|\tilde{\Delta}_n\right\|^2\right\} \le \frac{\epsilon}{\varepsilon^2} = \epsilon_0$$

if we take $\epsilon = \varepsilon^2\epsilon_0$. Since ϵ_0 is arbitrary we conclude that for any given $\varepsilon > 0$

$$\mathsf{P}\left\{\sqrt{n}\left\|\tilde{\Delta}_n\right\| > \varepsilon\right\} \underset{n\to\infty}{\to} 0$$

Analogously,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\Delta}_t = \frac{1}{\sqrt{n}}\sum_{t=n_0(\epsilon)}^{n}\tilde{\Delta}_t + O_\omega\left(n^{-1/2}\right)$$

and in view of (15.48),

$$\begin{aligned}
\mathsf{P}\left\{\frac{1}{\sqrt{n}}\left\|\sum_{t=n_0(\epsilon)}^{n}\tilde{\Delta}_t\right\| > \varepsilon\right\} &\le \mathsf{P}\left\{\sum_{t=n_0(\epsilon)}^{n}\left\|\tilde{\Delta}_t\right\| > \sqrt{n}\varepsilon\right\} \\
&\le \frac{1}{\sqrt{n}\varepsilon}\mathsf{E}\left\{\sum_{t=n_0(\epsilon)}^{n}\left\|\tilde{\Delta}_t\right\|\right\} = \frac{1}{\sqrt{n}\varepsilon}\sum_{t=n_0(\epsilon)}^{n}\mathsf{E}\left\{\left\|\tilde{\Delta}_t\right\|\right\} \\
&\le \frac{\epsilon}{\sqrt{n}\varepsilon}\sum_{t=n_0(\epsilon)}^{n} O\left(\frac{1}{\sqrt{t}}\right) = \mathrm{C}\frac{\epsilon}{\sqrt{n}\varepsilon}\left(\sqrt{n} - \sqrt{n_0(\epsilon)}\right) \\
&\le \mathrm{C}\frac{\epsilon}{\varepsilon} \le \epsilon_0, \quad \mathrm{C} \in (0, \infty)
\end{aligned}$$

and take $\epsilon = \mathrm{C}^{-1}\varepsilon\epsilon_0$. Since ϵ_0 is arbitrary we conclude that for any given $\varepsilon > 0$

$$\mathsf{P}\left\{\frac{1}{\sqrt{n}}\left\|\sum_{t=n_0(\epsilon)}^{n}\tilde{\Delta}_t\right\| > \varepsilon\right\} \underset{n\to\infty}{\to} 0$$

Therefore $\tilde{\theta}'_n \overset{\mathsf{P}}{\underset{n\to\infty}{\to}} 0$ which completes the proof. □

So, now we may formulate the following conclusion.

Conclusion 15.2. *Under the conditions of Lemma 15.2 for any admissible matrices Γ in procedure (15.42)*

$$\boxed{\begin{gathered}\sqrt{n}\left(\bar{x}_n - x^*\right) - \breve{\Delta}_n \underset{n\to\infty}{\overset{\mathsf{P}}{\to}} 0 \\ \breve{\Delta}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(-\left[\nabla f(x^*)\right]^{-1}\xi_t\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, K^*\right) \\ K^* = \left[\nabla f(x^*)\right]^{-1}\Xi_\xi\left[\nabla^\intercal f(x^*)\right]^{-1}\end{gathered}}$$

which means that

$$\boxed{\sqrt{n}\left(\bar{x}_n - x^*\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, K^*\right)}$$

or, in other words, the Ruppert–Polyak procedure (15.42), containing averaging, converges with the ***maximum possible convergence rate*** *K^* even if it does not involve any information on $\left[\nabla f(x^*)\right]^{-1}$ which is required (as follows from (15.41)) for the realization of the best gain matrix $\Gamma^* = \left[\nabla f(x^*)\right]^{-1}$ (15.38).*

Some generalizations related with nonlinear transformation of measurements can be found in Nazin (2001).

15.2.7 Convergence under dependent noises

Consider again procedure (15.7)

$$\boxed{\begin{gathered}x_n = x_{n-1} - \gamma_n\Gamma y_n \\ y_n = f\left(x_{n-1}\right) + \zeta_n, \quad 0 < \Gamma = \Gamma^\intercal \in \mathbb{R}^{N\times N}\end{gathered}} \tag{15.49}$$

where $\{\gamma_n\}_{n\geq 1}$ satisfies (15.10) and $\{\zeta_n\}_{n\geq 1}$ is a centered noise sequence which now may be dependent such that there exist its characteristics

$$\boxed{\begin{gathered}\|\zeta_n\|_{L_p} := \left(\left\|\zeta_{1,n}\right\|_{L_p}, \ldots, \left\|\zeta_{1,N}\right\|_{L_p}\right) \\ \left\|\zeta_{i,n}\right\|_{L_p} := \left(\mathsf{E}\left\{\left|\zeta_{i,n}\right|^p\right\}\right)^{1/p}, \quad p \geq 2 \\ q_{n,m}^{(q)} := \left\|\mathsf{E}\left\{\zeta_n \mid \mathcal{F}_m\right\}\right\|_{L_q}, \quad q = \frac{p}{p-1}, \; n > m \\ \Xi_{t,k} := \mathsf{E}\left\{\zeta_n\zeta_k^\intercal\right\}, \quad t, k = 1, 2, \ldots; \; \sigma_t^2 := \operatorname{tr}\left\{\Xi_{t,t}\right\}\end{gathered}} \tag{15.50}$$

Remark 15.1. *For independent (or martingale-difference) centered sequences $\{\zeta_n\}_{n\geq 1}$ we have*

$$\boxed{q_{n,m}^{(q)} = 0 \;\left(n > m\right), \quad \Xi_{t,k} = 0 \;\left(t \neq k\right)} \tag{15.51}$$

First, let us show the following fact (hereafter we follow Poznyak and Tchikin (1985)).

Lemma 15.4. *Let $\{x_n\}_{n\geq 1}$ be generated by (15.49) and $\{u_n\}_{n\geq 1}$ governed by*

$$\boxed{u_n = u_{n-1} - \gamma_n \left(u_{n-1} + \Gamma\zeta_n\right), \quad u_0 \overset{a.s.}{=} 0} \tag{15.52}$$

*Then under the conditions **A1** and **A2***

$$\boxed{\limsup_{n\to\infty} \left\|(x_n - x^*) - u_n\right\|^2_{\Gamma^{-1}} \leq \text{Const} \cdot \limsup_{n\to\infty} \|u_n\|^2} \tag{15.53}$$

for all $\omega \in \Omega$.

Proof. Define $w_n := (x_n - x^*) - u_n$. Then, by (15.49) and (15.52), it follows that

$$\begin{aligned}
\|w_n\|^2_{\Gamma^{-1}} &:= \|w_{n-1} - \gamma_n\Gamma y_n + \gamma_n\left(u_{n-1} + \Gamma\zeta_n\right)\|^2_{\Gamma^{-1}} \\
&= \|w_{n-1} - \gamma_n\left(\Gamma f\left(x_{n-1}\right) - u_{n-1}\right)\|^2_{\Gamma^{-1}} \\
&= \|w_{n-1}\|^2_{\Gamma^{-1}} - 2\gamma_n\left(w_{n-1}, \Gamma^{-1}\left(\Gamma f\left(x_{n-1}\right) - u_{n-1}\right)\right) \\
&\quad + \gamma_n^2\left\|\Gamma f\left(x_{n-1}\right) - u_{n-1}\right\|^2_{\Gamma^{-1}} \\
&= \|w_{n-1}\|^2_{\Gamma^{-1}} - 2\gamma_n\left(x_{n-1} - x^*, f\left(x_{n-1}\right)\right) \\
&\quad + 2\gamma_n\left(u_{n-1}, f\left(x_{n-1}\right)\right) + 2\gamma_n\left(w_{n-1}, u_{n-1}\right) \\
&\quad + \gamma_n^2\left\|\Gamma f\left(x_{n-1}\right) - u_{n-1}\right\|^2_{\Gamma^{-1}}
\end{aligned}$$

Then, by the assumption **A2** and in view of the λ-inequality (see Lemma 12.1 in Poznyak (2008))

$$2\left(a, b\right) \leq \varepsilon\|a\|^2 + \varepsilon^{-1}\|b\|^2, \quad \varepsilon > 0$$

with $\varepsilon = \varepsilon_n$, we get

$$\begin{aligned}
\|w_n\|^2_{\Gamma^{-1}} &\leq \|w_{n-1}\|^2_{\Gamma^{-1}} - 2\gamma_n k\left\|x_{n-1} - x^*\right\|^2 \\
&\quad + \gamma_n\left(\varepsilon\|u_{n-1}\|^2 + 2\varepsilon^{-1}\left[K\left\|x_{n-1} - x^*\right\|^2 + \|\Gamma\|\,\|w_n\|^2_{\Gamma^{-1}}\right]\right) \\
&\quad + 2\gamma_n^2\left\|\Gamma^{-1}\right\|\left(K\left\|x_{n-1} - x^*\right\|^2 + \|u_{n-1}\|^2\right) \\
&= \|w_{n-1}\|^2_{\Gamma^{-1}}\left(1 + 2\gamma_n\varepsilon^{-1}\|\Gamma\|\right) \\
&\quad - 2\gamma_n k\left\|x_{n-1} - x^*\right\|^2\left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n\left\|\Gamma^{-1}\right\|K\right) \\
&\quad + \gamma_n\left(\varepsilon + 2\gamma_n\left\|\Gamma^{-1}\right\|\right)\|u_{n-1}\|^2 \leq \|w_n\|^2_{\Gamma^{-1}}\left(1 + 2\gamma_n\varepsilon^{-1}\|\Gamma\|\right) \\
&\quad - 2\gamma_n k\left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n\left\|\Gamma^{-1}\right\|K\right)\left(\|w_{n-1}\|^2 + \|u_{n-1}\|^2\right) \\
&\quad + 2\gamma_n k\left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n\left\|\Gamma^{-1}\right\|K\right)\left(\varepsilon_1\|w_{n-1}\|^2 + \varepsilon_1^{-1}\|u_{n-1}\|^2\right) \\
&\quad + \gamma_n\left(\varepsilon + 2\gamma_n\left\|\Gamma^{-1}\right\|\right)\|u_{n-1}\|^2
\end{aligned}$$

$$
\begin{aligned}
&= \|w_n\|^2_{\Gamma^{-1}} \left(1 + 2\gamma_n \left[\varepsilon^{-1}\|\Gamma\| - k\left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n \left\|\Gamma^{-1}\right\| K\right)(1-\varepsilon_1)\right]\right) \\
&\quad + \gamma_n \|u_{n-1}\|^2 \left[2k\left(\varepsilon_1^{-1} - \left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n \left\|\Gamma^{-1}\right\| K\right)\right) + \varepsilon + 2\gamma_n \left\|\Gamma^{-1}\right\|\right]
\end{aligned}
$$

Taking ε and ε_1 such that

$$
\begin{aligned}
&c_0 := 1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n \left\|\Gamma^{-1}\right\| K \geq c_0 > 0 \\
&k\left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n \left\|\Gamma^{-1}\right\| K\right)(1-\varepsilon_1) - \varepsilon^{-1}\|\Gamma\| \geq c > 0
\end{aligned}
$$

we finally obtain

$$
\begin{aligned}
&\|w_n\|^2_{\Gamma^{-1}} \leq \|w_{n-1}\|^2_{\Gamma^{-1}}(1 - 2\gamma_n c) + \beta_n\gamma_n\|u_{n-1}\|^2 \\
&\beta_n := 2k\left(\varepsilon_1^{-1} - \left(1 - 2\varepsilon^{-1}\frac{K}{k} - \gamma_n\left\|\Gamma^{-1}\right\| K\right)\right) + \varepsilon + 2\gamma_n\left\|\Gamma^{-1}\right\|
\end{aligned}
$$

and by Lemma 16.14 in Poznyak (2008) on linear recurrent inequalities it follows that

$$
\begin{aligned}
&\limsup_{n\to\infty}\|w_n\|^2_{\Gamma^{-1}} \leq \text{Const}\limsup_{n\to\infty}\|u_{n-1}\|^2 \\
&\text{Const} = c^{-1}\left[k\left(\varepsilon_1^{-1} - \left(1 - 2\varepsilon^{-1}\frac{K}{k}\right)\right) + \varepsilon/2\right]
\end{aligned}
$$

which completes the proof. □

Conclusion 15.3. *This theorem shows that to prove any type of convergence to zero for* $\{x_n - x^*\}_{n\geq 1}$ *it is sufficient to show the corresponding convergence to zero for* $\{u_n\}_{n\geq 1}$.

The last result permits us to formulate the following important statement concerning the convergence of the Robbins–Monro procedure under dependent noise sequences.

Theorem 15.6. *Under conditions **A1–A2** procedure (15.49), where* $\{\gamma_n\}_{n\geq 1}$ *satisfies (15.10) and is monotonically decreasing such that*

$$\boxed{n\gamma_n \underset{n\to\infty}{\to} \text{const}} \tag{15.54}$$

provides the convergence

$$\boxed{x_n - x^* \underset{n\to\infty}{\to} 0} \tag{15.55}$$

*(a) **with probability one** (or a.s.) if*

$$\boxed{\left\|\sum_{n=1}^{\infty}\gamma_n\zeta_n\right\| \overset{a.s.}{<} \infty} \tag{15.56}$$

(b) ***in mean-square sense*** *if*

$$\boxed{\left\| \sum_{n=1}^{\infty} \gamma_n \zeta_n \right\|_{L_2} < \infty} \tag{15.57}$$

or, in other words, the convergence (in some probabilistic sense) of the series $\sum_{n=1}^{\infty} \gamma_n \zeta_n$ *implies the corresponding convergence of procedure (15.49).*

Proof. Using the conclusion above it is sufficient to state the convergence of sequence $\{u_n\}_{n \geq 1}$ (15.52) for which we have

$$\begin{aligned} u_n &= u_{n-1} - \gamma_n \left(u_{n-1} + \Gamma \zeta_n\right) \\ &= u_{n-1}\left(1 - \gamma_n\right) - \gamma_n \Gamma \zeta_n = -\Gamma \pi_n \sum_{t=1}^{n} \pi_t^{-1} \gamma_t \zeta_t \end{aligned}$$

where

$$\pi_n := \prod_{s=1}^{n} \left(1 - \gamma_s\right)$$

Since (by the inequality $1 + x \leq e^x$ valid for any $x \in \mathbb{R}$)

$$\pi_n \leq \exp\left\{ - \sum_{s=1}^{n} \gamma_s \right\} \underset{n \to \infty}{\to} 0$$

then by the Kronecker Lemma 8.3 the desired result, concerning a.s.-convergence, follows. As for mean-square convergence, notice that by the Abel identity (see Lemma 12.2 in Poznyak (2008))

$$\pi_n \sum_{t=n_0}^{n} \pi_t^{-1} \gamma_t \zeta_t = s_n - \pi_n \sum_{t=n_0}^{n} \left(\pi_t^{-1} - \pi_{t-1}^{-1} \right) s_{t-1}$$

where

$$s_n := \sum_{t=n_0}^{n} \gamma_t \zeta_t \underset{n \to \infty}{\overset{L_2}{\to}} s^*$$

Therefore,

$$\begin{aligned} \pi_n \sum_{t=n_0}^{n} \pi_t^{-1} \gamma_t \zeta_t = \left[s_n - s^* \right] - \frac{\pi_n}{\pi_{n_0}} s^* \\ - \pi_n \sum_{t=n_0}^{n} \left(\pi_t^{-1} - \pi_{t-1}^{-1} \right) \left[s_{t-1} - s^* \right] \end{aligned}$$

and hence, in view of the Jensen inequality (see Corollary 16.18 in Poznyak (2008))

$$\left(\sum_{t=1}^{n}\varphi_t\right)^2 \le n\left(\sum_{t=1}^{n}\varphi_t^2\right)$$

and taking into account that $\pi_t^{-1} - \pi_{t-1}^{-1} > 0$, we get

$$\begin{aligned}
\left\|\pi_n \sum_{t=1}^{n} \pi_t^{-1}\gamma_t\zeta_t\right\|_{L_2}^2 &\le 3\|s_n - s^*\|_{L_2}^2 + 3\frac{\pi_n}{\pi_{n_0}}\|s^*\|^2 \\
&\quad + 3\left\|\pi_n \sum_{t=n_0}^{n}\left(\pi_t^{-1} - \pi_{t-1}^{-1}\right)[s_{t-1} - s^*]\right\|_{L_2}^2 \\
&\le 3\|s_n - s^*\|_{L_2}^2 + 3\frac{\pi_n}{\pi_{n_0}}\|s^*\|^2 \\
&\quad + 3\pi_n^2(n - n_0)\sum_{t=n_0}^{n}\pi_t^{-2}\left(\pi_t^{-1} - \pi_{t-1}^{-1}\right)\|s_{t-1} - s^*\|_{L_2}^2
\end{aligned}$$

$$\begin{aligned}
&3\left\|s_n - s^*\right\|_{L_2}^2 + 3\frac{\pi_n}{\pi_{n_0}}\left\|s^*\right\|^2 + 3\pi_n^2(n - n_0)\sum_{t=n_0}^{n}\pi_t^{-2}\gamma_t^2\left\|s_{t-1} - s^*\right\|_{L_2}^2 \\
&= 3\|s_n - s^*\|_{L_2}^2 + 3\frac{\pi_n}{\pi_{n_0}}\|s^*\|^2 \\
&\quad + 3[\text{const} + 0(1)]\pi_n^2(n - n_0)\sum_{t=n_0}^{n}[\pi_t(t - n_0)]^{-2}\left\|s_{t-1} - s^*\right\|_{L_2}^2 \underset{n\to\infty}{\to} 0
\end{aligned}$$

by the Teöplitz Lemma 8.2 being applied to the third term. Theorem is proven. □

Corollary 15.2. (on a.s.-convergence) *Procedure (15.49) provides the convergence* $x_n - x^* \overset{a.s.}{\underset{n\to\infty}{\to}} 0$ *if*

$$\boxed{\sum_{n=1}^{\infty}\left[\gamma_n^2\sigma_n^2 + \gamma_n q_{n,n-1}^{(2)}\sum_{t=1}^{n-1}\gamma_t\sigma_t\right] < \infty} \tag{15.58}$$

Proof. Consider the series $\left\{s_n := \sum\limits_{t=1}^{n}\gamma_t\zeta_t\right\}_{n\ge 1}$. Then for $\mathcal{F}_n := \sigma(\zeta_1, \ldots, \zeta_n)$ we have

$$\mathsf{E}\left\{\|s_n\|^2 \mid \mathcal{F}_{n-1}\right\} \overset{a.s}{=} \|s_{n-1}\|^2 + 2\gamma_n\sum_{t=1}^{n-1}\gamma_t\zeta_t^{\mathsf{T}}\mathsf{E}\{\zeta_n \mid \mathcal{F}_{n-1}\} + \gamma_n^2\mathsf{E}\left\{\|\zeta_n\|^2 \mid \mathcal{F}_{n-1}\right\}$$

which permits us to conclude that (in view of Lemma 7.8)

$$\limsup_{n\to\infty} s_n^2 \overset{a.s.}{<} \infty$$

if

$$\sum_{t=1}^{\infty}\left[\gamma_n^2 \mathsf{E}\left\{\|\zeta_n\|^2 \mid \mathcal{F}_{n-1}\right\} + \gamma_n \sum_{t=1}^{n-1} \gamma_t \|\zeta_t\| \, \|\mathsf{E}\{\zeta_n \mid \mathcal{F}_{n-1}\}\|\right] \overset{a.s.}{<} \infty$$

But any series of nonnegative random variables converges with probability one if the corresponding series of mathematical expectations converges, namely, when

$$\sum_{t=1}^{\infty}\left[\gamma_n^2 \sigma_n^2 + \gamma_n \sum_{t=1}^{n-1} \gamma_t \mathsf{E}\{\|\zeta_t\| \, \|\mathsf{E}\{\zeta_n \mid \mathcal{F}_{n-1}\}\|\}\right] < \infty \tag{15.59}$$

Taking into account the Cauchy–Schwartz inequality

$$\mathsf{E}\{\|\zeta_t\| \, \|\mathsf{E}\{\zeta_n \mid \mathcal{F}_{n-1}\}\|\} \le \sqrt{\mathsf{E}\left\{\|\mathsf{E}\{\zeta_n \mid \mathcal{F}_{n-1}\}\|^2\right\}}\sqrt{\mathsf{E}\left\{\|\zeta_t\|^2\right\}} \le q_{n,n-1}^{(2)} \sigma_t$$

we conclude that series (15.59) converges if

$$\sum_{t=1}^{\infty}\left[\gamma_n^2 \sigma_n^2 + \gamma_n q_{n,n-1}^{(2)} \sum_{t=1}^{n-1} \gamma_t \sigma_t\right] < \infty$$

which takes place by (15.58). Corollary is proven. □

Corollary 15.3. (on a mean-square convergence) *Procedure (15.49) provides the convergence* $\mathsf{E}\left\{\|x_n - x^*\|^2\right\} \underset{n\to\infty}{\to} 0$ *if*

$$\boxed{\sum_{t=1}^{\infty}\sum_{s=1}^{\infty} \gamma_t \gamma_s \operatorname{tr}\left\{\Xi_{t,s}\right\} < \infty} \tag{15.60}$$

Proof. Again, for the series $\left\{s_n := \sum_{t=1}^{n} \gamma_t \zeta_t\right\}_{n\ge1}$ we have

$$\mathsf{E}\left\{s_n^2\right\} \overset{a.s}{=} \sum_{t=1}^{n}\sum_{s=1}^{n} \gamma_t \gamma_s \mathsf{E}\left\{\zeta_t^{\mathsf{T}} \zeta_s\right\} = \sum_{t=1}^{n}\sum_{s=1}^{n} \gamma_t \gamma_s \operatorname{tr}\left\{\Xi_{t,s}\right\}$$

which prove the corollary. □

Remark 15.2. *When* $\{\zeta_n\}_{n\ge1}$ *is the sequence of* ***independent*** *random variables one has*

$$q_{n,n-1}^{(2)} = 0$$

and the conditions (15.58) become

$$\boxed{\sum_{n=1}^{\infty} \gamma_n^2 \sigma_n^2 < \infty} \tag{15.61}$$

15.3 Stochastic optimization

15.3.1 Stochastic gradient method

Here we will consider the following optimization problem: for a differentiable function $f : \mathbb{R}^N \to \mathbb{R}$ find its minimum, that is,

$$\boxed{f(x) \to \min_{x \in \mathbb{R}^N}} \tag{15.62}$$

provided that at any point $x_n \in \mathbb{R}^N$ its measurable gradient $\nabla f(x_n)$, disturbed by a stochastic noise ζ_n, namely, the vector

$$\boxed{y_n = \nabla f(x_n) + \zeta_n} \tag{15.63}$$

is available.

First, notice that if a minimum point x^* is unique, i.e.,

$$f(x) > f\left(x^*\right), \quad x \neq x^*$$

then the optimality condition for a point x to be an extremum point is

$$\boxed{\nabla f(x) = 0} \tag{15.64}$$

which exactly coincides with the stochastic nonlinear regression problem (15.1)–(15.2) considered above. This means that *we do not need to prove any new results* concerning the stochastic minimization procedure

$$\boxed{\begin{array}{c} x_n = x_{n-1} - \gamma_n \Gamma y_n \\ y_n = \nabla f(x_{n-1}) + \zeta_n, \quad 0 < \Gamma = \Gamma^\intercal \in \mathbb{R}^{N \times N} \end{array}} \tag{15.65}$$

Assuming that the function f is strongly convex, and, hence (see Chapter 21 in Poznyak (2008)), satisfies the inequalities

$$\boxed{k\|x - x^*\|^2 \le (\nabla f(x), x - x^*) \le K\|x - x^*\|^2} \tag{15.66}$$

for some positive constants k, K and for all $x \in \mathbb{R}^N$, it is sufficient to change in all statements above the vector function f to the gradient $\nabla f(x)$ and we do not need to re-prove all of the theorems on convergence and the rate of convergence. So, below we will formulate the main theorems concerning the properties of the stochastic gradient procedure (15.65).

Theorem 15.7. (on convergence under dependent noises) *Let $\{\gamma_n\}_{n\ge1}$ in (15.65) satisfy the conditions (15.10) and (15.54) and $\{\zeta_n\}_{n\ge1}$ be a centered noise sequence which may be dependent on the characteristics (15.50). Suppose also that the function f is differentiable and its gradient satisfies the Lipschitz condition. Then* *(a)* $x_n - x^* \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ *if*

$$\boxed{\sum_{n=1}^{\infty}\left[\gamma_n^2\sigma_n^2 + \gamma_n q_{n,n-1}^{(2)}\sum_{t=1}^{n-1}\gamma_t\sigma_t\right] < \infty} \tag{15.67}$$

(b) $\mathsf{E}\left\{\|x_n - x^*\|^2\right\} \underset{n\to\infty}{\to} 0$ *if*

$$\boxed{\sum_{t=1}^{\infty}\sum_{s=1}^{\infty}\gamma_t\gamma_s \operatorname{tr}\left\{\Xi_{t,s}\right\} < \infty} \tag{15.68}$$

Theorem 15.8. (on asymptotic normality) *If under the assumptions of Theorem 15.7 the function f is twice differentiable in the optimal point and in the stochastic gradient procedure (15.65) with $\Gamma_n = \dfrac{\Gamma}{n+b}$ the matrix Γ satisfies*

$$\boxed{\begin{array}{c}\lambda_- := \lambda_{\min}\left(\nabla^2 f(x^*)\Gamma + \Gamma^\intercal\nabla^2 f(x^*)\right) > 0\\ S := \left[\dfrac{1}{2}I - \Gamma\nabla^2 f(x^*)\right] \textit{is Hurwitz}\end{array}} \tag{15.69}$$

there exists a limit

$$\boxed{R := \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathsf{E}\left\{\xi_t\xi_t^\intercal\right\} > 0} \tag{15.70}$$

and $\{\xi_t\}_{t\geq 1}$ is a martingale-difference sequence ($q_{n,n-1}^{(2)} = 0$), then

$$\boxed{\sqrt{n}\,(x_n - x^*) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, K)} \tag{15.71}$$

where the matrix $K = K^\intercal > 0$ is the solution of the following algebraic matrix Lyapunov equation:

$$\boxed{SK + KS^\intercal = -\Gamma R\Gamma^\intercal} \tag{15.72}$$

and the best rate of the convergence is achieved for

$$\boxed{\Gamma^* = \left[\nabla^2 f(x^*)\right]^{-1}} \tag{15.73}$$

so that

$$\boxed{K(\Gamma) \geq K^* = K(\Gamma^*) = \left(\nabla^2 f(x^*)\right)^{-1} \Xi_\xi \left[\nabla^2 f(x^*)\right]^{-1}} \tag{15.74}$$

if $\{\xi_t\}_{t\geq 1}$ *is supposed to be independent and stationary providing* $\Xi_\xi = R$.

Theorem 15.9. (on LIL for stochastic gradient) *If in Theorem 15.8 the condition of twice differentiability in the optimal point is strengthened to its continuity ($\nabla^2 f(x) \in C^2$), then for large enough n*

$$\boxed{x_n = x^* + K^{1/2}\bar{\zeta}_n(\omega)\sqrt{2n^{-1}\ln\ln n}} \tag{15.75}$$

where $\bar{\zeta}_n(\omega)$ *is a random vector having its set of limit points confined in n-dimensional sphere* $\{x : \|x\| \leq 1\}$ *such that*

$$\boxed{\limsup_{n\to\infty} \left\|\bar{\zeta}_n(\omega)\right\| \overset{a.s.}{=} 1} \tag{15.76}$$

Theorem 15.10. (The Ruppert–Polyak extension) *For any admissible matrices* Γ *in procedure (15.65) for the averaged vectors*

$$\boxed{\bar{x}_n := \frac{1}{n}\sum_{t=1}^{n} x_t = \bar{x}_{n-1} - \frac{1}{n}(\bar{x}_{n-1} - x_n)} \tag{15.77}$$

we have

$$\boxed{\begin{gathered}\sqrt{n}(\bar{x}_n - x^*) - \breve{\Delta}_n \underset{n\to\infty}{\overset{P}{\to}} 0 \\ \breve{\Delta}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(-\left[\nabla^2 f(x^*)\right]^{-1}\xi_t\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, K^*\right) \\ K^* = \left[\nabla^2 f(x^*)\right]^{-1}\Xi_\xi\left[\nabla^2 f(x^*)\right]^{-1}\end{gathered}}$$

which means that

$$\boxed{\sqrt{n}(\bar{x}_n - x^*) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, K^*)}$$

or, in other words, the Ruppert–Polyak extension of the procedure (15.65), containing averaging (15.77), converges with the ***maximum possible convergence rate*** K^* *even if it does not involve any information on* $\left[\nabla^2 f(x^*)\right]^{-1}$ *which is required for the realization of the best gain matrix* $\Gamma^* = \left[\nabla f(x^*)\right]^{-1}$ *(15.73).*

15.3.2 Kiefer–Wolfowitz procedure

As has been done above, we will consider the optimization problem (15.62) provided that at any point $x \in \mathbb{R}^N$ there is measurable only the corresponding value of the minimized function $f(x)$, disturbed by a stochastic scalar noise ζ_n; namely, the scalar

$$\boxed{y_n(x) = f(x) + \zeta_n} \tag{15.78}$$

is available.

Consider the recurrent procedure which generates the stochastic sequence $\{x_n\}_{n\geq 0}$ according to the following rule:

$$\begin{gathered}
x_n = x_{n-1} - \frac{\gamma_n}{2\alpha_n}\Gamma Y_n \\
0 \leq \gamma_n, \quad \sum_{n=1}^{\infty} \gamma_n = 0 \\
\boxed{\begin{gathered}
0 < \alpha_n, \quad 0 < \Gamma = \Gamma^{\intercal} \in \mathbb{R}^{N\times N} \\
Y_n := \sum_{i=1}^{N} \left[y_n(x_{n-1} + \alpha_n e_i) - y_n(x_{n-1} - \alpha_n e_i) \right] e_i \in \mathbb{R}^N \\
e_i^{\intercal} := \Bigg(\underbrace{0, 0, \ldots, 0, 1}_{i}, 0, \ldots, 0 \Bigg) \in \mathbb{R}^N
\end{gathered}}
\end{gathered} \tag{15.79}$$

We shall call this scheme the generalized version of the Kiefer–Wolfowitz procedure which originally was suggested in Kiefer and Wolfowitz (1952) for the scalar case $N = 1$.

Before the presentation of the convergence analysis of the procedure (15.79) we will need the following auxiliary result.

Lemma 15.5. (on a gradient approximation) *If*

A1, the minimized function $f : \mathbb{R}^N \to \mathbb{R}$, is differentiable and its gradient satisfies the Lipschitz condition

then

$$\boxed{\begin{gathered}
\frac{1}{2\alpha_n} Y_n = \nabla f(x_{n-1}) + \bar{\zeta}_n \\
\bar{\zeta}_n = w_n + \frac{1}{2\alpha_n} \sum_{i=1}^{N} \left(\zeta_{i,n}^{+} - \zeta_{i,n}^{-} \right) e_i
\end{gathered}} \tag{15.80}$$

where w_n is the vector satisfying

$$\boxed{\|w_n\| \leq \frac{N L_{\nabla}}{2} \alpha_n} \tag{15.81}$$

and $\zeta_{i,n}^{+}$, $\zeta_{i,n}^{-}$ are noises in the measurements of the function f in the points $f(x_{n-1}+\alpha_n e_i)$ and $f(x_{n-1}-\alpha_n e_i)$ respectively.

Proof. By assumption A1 (see Chapter 21 in Poznyak (2008)), it follows that

$$\begin{aligned} y_n(x_{n-1}+\alpha_n e_i) &= f(x_{n-1}+\alpha_n e_i)+\zeta_{i,n}^{+} \\ &= f(x_{n-1})+(\nabla f(x_{n-1}),\alpha_n e_i)+w_{i,n}^{+}+\zeta_{i,n}^{+} \\ &= f(x_{n-1})+\alpha_n\frac{\partial}{\partial x_i}f(x_{n-1})+w_{i,n}^{+}+\zeta_{i,n}^{+} \end{aligned}$$

where $\zeta_{i,n}^{+}$ is the noise of measurement, $w_{i,n}^{+}$ is a value satisfying

$$\left|w_{i,n}^{+}\right| \le \frac{L_\nabla}{2}\|\alpha_n e_i\|^2 = \frac{L_\nabla}{2}\alpha_n^2$$

and L_∇ is the Lipschitz constant for the gradient $\nabla f(x)$. Analogously

$$\begin{aligned} y_n(x_{n-1}-\alpha_n e_i) &= f(x_{n-1}-\alpha_n e_i)+\zeta_{i,n}^{-} \\ &= f(x_{n-1})-(\nabla f(x_{n-1}),\alpha_n e_i)+w_{i,n}^{-}+\zeta_{i,n}^{-} \\ &= f(x_{n-1})-\alpha_n\frac{\partial}{\partial x_i}f(x_{n-1})+w_{i,n}^{-}+\zeta_{i,n}^{-} \end{aligned}$$

$$\left|w_{i,n}^{-}\right| \le \frac{L_\nabla}{2}\|\alpha_n e_i\|^2 = \frac{L_\nabla}{2}\alpha_n^2$$

Then

$$\begin{aligned} \frac{1}{2\alpha_n}Y_n &:= \frac{1}{2\alpha_n}\sum_{i=1}^{N}\left[2\alpha_n\frac{\partial}{\partial x_i}f(x_{n-1})+w_{i,n}^{+}-w_{i,n}^{-}+\zeta_{i,n}^{+}-\zeta_{i,n}^{-}\right] \\ &= \sum_{i=1}^{N}\left[\frac{\partial}{\partial x_i}f(x_{n-1})+\frac{1}{2\alpha_n}\left(w_{i,n}^{+}-w_{i,n}^{-}+\zeta_{i,n}^{+}-\zeta_{i,n}^{-}\right)\right]e_i \\ &= \nabla f(x_{n-1})+\bar{\zeta}_n \end{aligned}$$

which implies (15.80) with

$$w_n := \frac{1}{2\alpha_n}\sum_{i=1}^{N}\left(w_{i,n}^{+}-w_{i,n}^{-}\right)e_i$$

Lemma is proven. □

By this lemma one can apply now Theorem 15.7 on the convergence of the stochastic gradient procedure where instead of $\{\zeta_n\}_{n\ge1}$ the sequence $\{\bar{\zeta}_n\}_{n\ge1}$ is used. So, we get

Theorem 15.11. (KW-procedure under dependent noises) *Let $f:\mathbb{R}^N\to\mathbb{R}$ be a strongly convex function satisfying assumption A1 of Lemma 15.5 and x^* be its minimum point. Suppose that there exist*

$$q_{n,n-1}^{(2)} := \max_i \max\left\{\left\|\mathsf{E}\left\{\zeta_{i,n}^{+}\mid\mathcal{F}_{n-1}\right\}\right\|_{L_2}, \left\|\mathsf{E}\left\{\zeta_{i,n}^{-}\mid\mathcal{F}_{n-1}\right\}\right\|_{L_2}\right\}$$

$$
\begin{aligned}
&{}^{is}\Xi_{t,k}^{++} := \mathsf{E}\left\{\zeta_{i,t}^{+}\zeta_{s,k}^{+\intercal}\right\}, \quad {}^{is}\Xi_{t,k}^{--} := \mathsf{E}\left\{\zeta_{i,t}^{-}\zeta_{s,k}^{-\intercal}\right\}, \quad t,k = 1,2,\ldots;\\
&{}^{is}\Xi_{t,k}^{+-} := \mathsf{E}\left\{\zeta_{i,t}^{+}\zeta_{s,k}^{-\intercal}\right\}, \quad {}^{is}\Xi_{t,k}^{-+} := \mathsf{E}\left\{\zeta_{i,t}^{-}\zeta_{s,k}^{+\intercal}\right\}\\
&\sigma_t^2 := \max_i \max \left\{\mathsf{E}\left\{\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)^{\intercal}\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)\right\}\right\}
\end{aligned}
$$

Then for the KW-procedure (15.79)

(a) $x_n - x^* \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ *if*

$$
\boxed{\sum_{n=1}^{\infty}\left[\gamma_n\alpha_n + \frac{\gamma_n^2}{\alpha_n^2}\sigma_n^2 + \frac{\gamma_n}{\alpha_n}q_{n,n-1}^{(2)}\sum_{t=1}^{n-1}\frac{\gamma_t}{\alpha_t}\sigma_t\right] < \infty} \tag{15.82}
$$

(b) $\mathsf{E}\left\{\|x_n - x^*\|^2\right\} \underset{n\to\infty}{\to} 0$ *if*

$$
\boxed{\begin{array}{c}
\sum_{n=1}^{\infty}\gamma_n\alpha_n < \infty\\
\sum_{t=1}^{\infty}\sum_{s=1}^{\infty}\frac{\gamma_t\gamma_s}{\alpha_t\alpha_s}\operatorname{tr}\left\{{}^{is}\Xi_{t,k}^{++} + {}^{is}\Xi_{t,k}^{--} - {}^{is}\Xi_{t,k}^{+-} - {}^{is}\Xi_{t,k}^{-+}\right\} < \infty
\end{array}} \tag{15.83}
$$

Proof. We have convergence if the series

$$
\sum_{s=1}^{\infty}\gamma_t\bar{\zeta}_t = \sum_{s=1}^{\infty}\gamma_t w_t + \sum_{s=1}^{\infty}\gamma_t\frac{1}{2\alpha_t}\sum_{i=1}^{N}\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)e_i
$$

converges in the corresponding probabilistic sense for which, obviously, the convergence of two series is sufficient

$$
\left\|\sum_{s=1}^{\infty}\gamma_t w_t\right\| \le \sum_{s=1}^{\infty}\gamma_t\|w_t\| \le \frac{NL_{\nabla}}{2}\sum_{s=1}^{\infty}\gamma_t\alpha_n < \infty
$$

and

$$
\begin{aligned}
\left\|\sum_{s=1}^{\infty}\gamma_t\frac{1}{2\alpha_t}\sum_{i=1}^{N}\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)e_i\right\| &\le \frac{1}{2}\sum_{i=1}^{N}\left\|\sum_{s=1}^{\infty}\frac{\gamma_t}{\alpha_t}\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)e_i\right\|\\
&\le \frac{1}{2}\sum_{i=1}^{N}\left(\left\|\sum_{s=1}^{\infty}\frac{\gamma_t}{\alpha_t}\zeta_{i,t}^{+}e_i\right\| + \left\|\sum_{s=1}^{\infty}\frac{\gamma_t}{\alpha_t}\zeta_{i,t}^{-}e_i\right\|\right)
\end{aligned}
$$

The proof results directly from Corollaries 15.2 and 15.3 where $\{\gamma_n\}_{n\ge1}$ is changed to $\left\{\gamma_t\alpha_t^{-1}\right\}_{n\ge1}$. Here we need to take into account that

$$
\mathsf{E}\left\{\left(\zeta_{i,t}^{+} - \zeta_{i,t}^{-}\right)\left(\zeta_{s,k}^{+} - \zeta_{s,k}^{-}\right)^{\intercal}\right\} = \operatorname{tr}\left\{{}^{is}\Xi_{t,k}^{++} + {}^{is}\Xi_{t,k}^{--} - {}^{is}\Xi_{t,k}^{+-} - {}^{is}\Xi_{t,k}^{-+}\right\}
$$

Theorem is proven. □

Corollary 15.4. *If in Theorem 15.11 the measurement noises $\zeta_{i,t}^{+}$, $\zeta_{i,t}^{-}$ are martingale-differences (or independent) then $q_{n,n-1}^{(2)} = 0$ and the conditions of the convergence with probability one (15.82) become*

$$\boxed{\sum_{n=1}^{\infty} \left[\gamma_n \alpha_n + \gamma_n^2 \alpha_n^{-2} \sigma_n^2 \right] < \infty} \tag{15.84}$$

Some other properties (such as asymptotic normality, LIL, the optimal selection of the gain matrix Γ, the version with averaging) can be considered analogously to those in the previous sections.

15.3.3 Random search method

Again as before, we will consider the optimization problem (15.62) under available measurements (15.78).

Consider the following recurrence, generating the stochastic sequence $\{x_n\}_{n\geq 0}$ according to the rule:

$$\boxed{\begin{array}{c} x_n = x_{n-1} - \dfrac{\gamma_n}{\alpha_n} N \Gamma Y_n \\ 0 \leq \gamma_n, \quad \displaystyle\sum_{n=1}^{\infty} \gamma_n = 0 \\ 0 < \alpha_n, \quad 0 < \Gamma = \Gamma^{\intercal} \in \mathbb{R}^{N \times N} \\ Y_n := v_n y_n \left(x_{n-1} + \alpha_n v_n \right), \quad v_n \in \mathbb{R}^N \\ y_n \left(x_{n-1} + \alpha_n v_n \right) = f \left(x_{n-1} + \alpha_n v_n \right) + \zeta_n \end{array}} \tag{15.85}$$

where $\{v_n\}_{n\geq 1}$ is a sequence of independent random vectors uniformly distributed on the unite N-dimensional sphere so that

$$\boxed{\|v_n\|^2 = 1, \quad \mathsf{E}\{v_n\} = 0, \quad \mathsf{E}\left\{v_n v_n^{\intercal}\right\} = N^{-1} I} \tag{15.86}$$

This random sequence is especially introduced providing the so-called '*minimum point search property*' justifying the name of the procedure (15.85) as the *random search procedure*.

To analyze the property of the random search procedure (15.85) we need to introduce two σ-algebras:

$$\mathcal{G}_{n-1} := (x_0; \zeta_1, v_1; \ldots; \zeta_{n-1}, v_{n-1}; \zeta_n)$$
$$\mathcal{F}_{n-1} := (x_0; \zeta_1, v_1; \ldots; \zeta_{n-1}, v_{n-1};)$$

Obviously,

$$\mathcal{F}_{n-1} \subset \mathcal{G}_{n-1}$$

so that

$$\mathsf{E}\{\cdot \mid \mathcal{F}_{n-1}\} \overset{a.s.}{=} \mathsf{E}\{\mathsf{E}\{\cdot \mid \mathcal{G}_{n-1}\} \mid \mathcal{F}_{n-1}\}$$

If $f : \mathbb{R}^N \to \mathbb{R}$ is a strongly convex function, satisfying assumption A1 of Lemma 15.5, then

$$\begin{aligned}
&\alpha_n^{-1} N \mathsf{E}\{Y_n \mid \mathcal{G}_{n-1}\} \\
&\quad = \alpha_n^{-1} N \mathsf{E}\left\{v_n \left[f(x_{n-1} + \alpha_n v_n) + \zeta_n\right] \mid \mathcal{G}_{n-1}\right\} \\
&\quad = \alpha_n^{-1} N \mathsf{E}\{v_n f(x_{n-1} + \alpha_n v_n) \mid \mathcal{G}_{n-1}\} \\
&\quad = \alpha_n^{-1} N \mathsf{E}\left\{v_n f(x_{n-1}) + \alpha_n v_n v_n^{\intercal} \nabla f(x_{n-1}) + v_n w_n \mid \mathcal{G}_{n-1}\right\} \\
&\quad = \nabla f(x_{n-1}) + \alpha_n^{-1} N \mathsf{E}\{v_n w_n \mid \mathcal{G}_{n-1}\}
\end{aligned}$$

where

$$|w_n| \le \frac{L_\nabla}{2} \|\alpha_n v_n\|^2 = \frac{L_\nabla}{2} \alpha_n^2$$

which gives under the definition

$$\bar{\zeta}_n := \alpha_n^{-1} N \mathsf{E}\{v_n w_n \mid \mathcal{G}_{n-1}\}$$

the following estimate

$$\left\|\bar{\zeta}_n\right\| \le \alpha_n^{-1} N \mathsf{E}\{\|v_n\| |w_n| \mid \mathcal{G}_{n-1}\} \le \frac{L_\nabla}{2} N \alpha_n$$

This means that the procedure (15.85) can be represented as

$$\boxed{\begin{gathered}
x_n = x_{n-1} - \gamma_n \Gamma\left[\nabla f(x_{n-1}) + \bar{\zeta}_n\right] \\
\bar{\zeta}_n := \alpha_n^{-1} N \mathsf{E}\{v_n w_n \mid \mathcal{G}_{n-1}\}, \quad \left\|\bar{\zeta}_n\right\| \le \frac{L_\nabla}{2} N \alpha_n
\end{gathered}} \tag{15.87}$$

Now applying the results concerning the stochastic gradient method we may formulate the following statement.

Theorem 15.12. (on random search method convergence) *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a strongly convex function satisfying assumption A1 of Lemma 15.5 and x^* be its minimum point. Then,* ***irrespective of whether noise sequence measurement*** $\{\zeta_n\}_{n\ge 1}$ ***is dependent or not****, for the Random Search procedure (15.85) $x_n - x^* \underset{n\to\infty}{\overset{a.s.}{\to}} 0$ and, in the same time, $\mathsf{E}\left\{\|x_n - x^*\|^2\right\} \underset{n\to\infty}{\to} 0$ if*

$$\boxed{\left\|\sum_{n=1}^{\infty} \gamma_n \bar{\zeta}_n\right\| < \infty} \tag{15.88}$$

in the corresponding probabilistic sense, which takes place if, additional to the properties $0 \le \gamma_n$, $\sum_{n=1}^{\infty} \gamma_n = 0$, *the sequences* $\{\gamma_n\}_{n\ge1}$ *and* $\{\alpha_n\}_{n\ge1}$ *satisfy*

$$\boxed{\sum_{n=1}^{\infty} \gamma_n \alpha_n < \infty} \tag{15.89}$$

Other properties (such as asymptotic normality, LIL, the optimal selection of the gain matrix Γ, the version with averaging and so on) can be considered in a way analogous to that done in the previous sections.

16 Robust Stochastic Control

Contents

This chapter deals with a version of the *robust stochastic maximum principle* (*RSMP*) applied to the min-max Mayer problem formulated for stochastic differential equations with the control-dependent diffusion term. The parametric families of first- and second-order adjoint stochastic processes are introduced to construct the corresponding Hamiltonian formalism. The Hamiltonian function used for the construction of the robust optimal control is shown to be equal to the Lebesgue integral over a parametric set (which may be a compact or a finite set) of the standard stochastic Hamiltonians corresponding to a fixed value of the uncertain parameter. The chapter deals with a cost function given at finite horizon and containing the mathematical expectation of a terminal term. A terminal condition, covered by a vector function, is also considered. The optimal control strategies, adapted for available information, for the wide class of uncertain systems given by a stochastic differential equation with unknown parameters from a given uncertainty set, are constructed. This problem belongs to the class of min-max stochastic optimization problems. Robust stochastic LQ control designing is discussed in detail. Two numerical examples, dealing with *production planning* and *reinsurance-dividend management*, illustrate the theoretical results.

16.1 Introduction

During the last decade, the min-max control problem, dealing with different classes of nonlinear systems, has received much attention from many researchers because of its theoretical and practical importance. Basically, the results in this area are based on two classical approaches:

- *Maximum principle* (MP) (Pontryagin et al., 1969, translated from Russian); and
- *dynamic programming method* (DP) (Bellman, 1960).

In the case of a complete model description, both of them can be directly applied to construct optimal control.

Various forms of the *stochastic maximum principle* have been published in the literature (Kushner, 1972; Fleming and Rishel, 1975; Bismut, 1977, 1978; Haussman, 1981). All of these publications have usually dealt with systems whose diffusion coefficients did not contain control variables and the control region of which was assumed to be convex. In Bensoussan (1983) the case of diffusion coefficients that depend smoothly on a control variable, was considered. Later this approach was extended to the class of partially observable systems (Haussman, 1982; Bensoussan, 1992), where optimal control consists of two basic components: state estimation and control via the estimates obtained. The most advanced results concerning the maximum principle for nonlinear stochastic differential equations with controlled diffusion terms were obtained by the Fudan University group, led by X. Li (see Zhou (1991) and Yong and Zhou (1999); and see the bibliography within).

Faced with some *uncertainties* (parametric type, unmodeled dynamics, external perturbations etc.) the results above cannot be applied. There are two ways to overcome uncertainty problems:

- The first is to apply the adaptive approach (Duncan et al., 1999) to identify the uncertainty on-line and then use the resulting estimates to construct a control strategy (Duncan and Varaiya, 1971);
- The second one, which will be considered in this chapter, is to obtain a solution suitable for a class of given models by formulating a corresponding *min-max control problem*, where the maximization is taken over a set of possible uncertainties and the minimization is taken over all of the control strategies within a given set.

For *stochastic uncertain systems*, min-max control of a class of dynamic systems with mixed uncertainties was investigated in different publications. Robust (non-optimal) control for linear time-varying systems given by stochastic differential equations was studied in Poznyak and Taksar (1996) and Taksar et al. (1998) where the solution is based on the stochastic *Lyapunov analysis* with martingale technique implementation. Other problems dealing with discrete time models of deterministic and/or simplest stochastic nature and their corresponding solutions are discussed in Yaz (1991), Blom and Everdij (1993), Bernhard (1994) and Boukas et al. (1999). In Ugrinovskii and Petersen (1997) the finite horizon min-max optimal control problems of nonlinear continuous time systems with stochastic uncertainty are considered. The original problem was converted into an unconstrained stochastic game problem and a stochastic version of the S-procedure has been designed to obtain a solution.

In this chapter we explore the possibilities of the MP approach for a class of min-max control problems for uncertain systems given by a system of stochastic differential equations. Here we will follow Poznyak (2002a,b).

16.2 Problem setting

16.2.1 Stochastic uncertain systems

Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathsf{P}\right)$ be a given *filtered probability space*, that is,

- the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is complete;
- the sigma-algebra $\mathcal{F}_0$ contains all the P-null sets in $\mathcal{F}$;
- the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is right continuous: $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$.

On this probability space an m-dimensional standard Brownian motion is defined, i.e., $(W(t), t \geq 0)$ (with $W(0) = 0$) is an $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^m$-valued process such that

$$\begin{array}{c} \mathsf{E}\{W(t) - W(s) \mid \mathcal{F}_s\} = 0 \quad \mathsf{P}\text{-}a.s. \\ \mathsf{E}\left\{[W(t) - W(s)][W(t) - W(s)]^{\mathsf{T}} \mid \mathcal{F}_s\right\} = (t-s) I \quad \mathsf{P}\text{-}a.s. \\ \mathsf{P}\{\omega \in \Omega : W(0) = 0\} = 1 \end{array}$$

Consider the stochastic nonlinear controlled continuous-time process with the dynamics $x(t)$ given by

$$\begin{aligned} x(t) = x(0) + \int_{s=0}^{t} b^{\alpha}(s, x(s), u(s))\, dt \\ + \int_{s=0}^{t} \sigma^{\alpha}(s, x(s), u(s))\, dW(s) \end{aligned} \tag{16.1}$$

or, in the abstract (symbolic) form,

$$\begin{aligned} &dx(t) = b^{\alpha}(t, x(t), u(t))\, dt + \sigma^{\alpha}(t, x(t), u(t))\, dW(t) \\ &x(0) = x_0, \quad t \in [0, T]\ (T > 0) \end{aligned} \tag{16.2}$$

The first integral in (16.1) is a stochastic ordinary integral and the second one is an Itô integral. In the above $u(t) \in U$ is a control at time t and

$$\begin{aligned} &b^{\alpha} : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^n \\ &\sigma^{\alpha} : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n\times m} \end{aligned}$$

The parameter α is supposed to be *a priori unknown* and running a given parametric set $\mathcal{A}$ from a space with a countable additive measure m.

For any $\alpha \in \mathcal{A}$ denote

$$\begin{aligned} &b^{\alpha}(t, x, u) := \left(b_1^{\alpha}(t, x, u), \ldots, b_n^{\alpha}(t, x, u)\right)^{\mathsf{T}} \\ &\sigma^{\alpha}(t, x, u) := \left(\sigma^{1\alpha}(t, x, u), \ldots, \sigma^{n\alpha}(t, x, u)\right) \\ &\sigma^{j\alpha}(t, x, u) := \left(\sigma_1^{j\alpha}(t, x, u), \ldots, \sigma_m^{j\alpha}(t, x, u)\right)^{\mathsf{T}} \end{aligned}$$

It is assumed that

A1: $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration generated by $(W(t), t \geq 0)$ and augmented by the **P**-null sets from $\mathcal{F}$.
A2: (U, d) is a separable metric space with a metric d.

The following definition is used subsequently.

Definition 16.1. *The function $f : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n\times m}$ is said to be an $L_\phi\left(C^2\right)$-**mapping** if*

1. *it is Borel measurable;*
2. *it is C^2 in x for any $t \in [0,T]$ and any $u \in U$;*
3. *there exist a constant L and a modulus of continuity $\phi : [0,\infty) \to [0,\infty)$ such that for any $t \in [0,T]$ and for any $x, u, \hat{x}, u \in \mathbb{R}^n \times U \times \mathbb{R}^n \times U$*

$$\boxed{\begin{array}{c}\left\| f(t,x,u) - f\left(t,\hat{x},\hat{u}\right)\right\| \leq L\left\|x-\hat{x}\right\| + \phi\left(d\left(u,\hat{u}\right)\right)\\ \|f(t,0,u)\| \leq L\\ \left\| f_x(t,x,u) - f_x\left(t,\hat{x},\hat{u}\right)\right\| \leq L\left\|x-\hat{x}\right\| + \phi\left(d\left(u,\hat{u}\right)\right)\\ \left\| f_{xx}(t,x,u) - f_{xx}\left(t,\hat{x},\hat{u}\right)\right\| \leq \phi\left(\left\|x-\hat{x}\right\| + d\left(u,\hat{u}\right)\right)\end{array}} \tag{16.3}$$

(here $f_x(\cdot,x,\cdot)$ and $f_{xx}(\cdot,x,\cdot)$ are the partial derivatives of the first and the second order).

In view of this definition, it is also assumed that
A3: for any $\alpha \in \mathcal{A}$ both $b^\alpha(t,x,u)$ and $\sigma^\alpha(t,x,u)$ are $L_\phi\left(C^2\right)$-mappings.

Let $\mathcal{A}_0 \subset \mathcal{A}$ be measurable subsets with a finite measure, that is,

$$m(\mathcal{A}_0) < \infty$$

The following assumption concerning the right-hand side of (16.2) will be in force throughout:
A4: All components $b^\alpha(t,x,u)$, $\sigma^\alpha(t,x,u)$ are measurable with respect to α, that is, for any $i = 1,\ldots,n$, $j = 1,\ldots,m$, $c \in R^1$, $x \in R^n$, $u \in U$ and $t \in [0,T]$

$$\left\{\alpha : b_i^\alpha(t,x,u) \leq c\right\} \in \mathcal{A}$$
$$\left\{\alpha : \sigma_j^{i\alpha}(t,x,u) \leq c\right\} \in \mathcal{A}$$

Moreover, every considered function of α is assumed to be measurable with respect to α.

The only sources of uncertainty in this system description are

- the system random noise $W(t)$;
- the *a priori* unknown parameter $\alpha \in \mathcal{A}$.

It is assumed that *the past information is available* for the controller.

To emphasize the dependence of the random trajectories on the parameter $\alpha \in \mathcal{A}$ equation (16.2) is rewritten as

$$\boxed{\begin{cases} dx^{\alpha}(t) = b^{\alpha}\left(t, x^{\alpha}(t), u(t)\right)dt + \sigma^{\alpha}\left(t, x^{\alpha}(t), u(t)\right)dW(t) \\ x^{\alpha}(0) = x_0, \quad t \in [0,T]\,(T>0) \end{cases}} \tag{16.4}$$

16.2.2 Terminal condition, feasible and admissible controls

The following definitions will be used throughout this chapter.

Definition 16.2. *A stochastic control $u(\cdot)$ is called* ***feasible*** *in the stochastic sense (or, s-feasible) for the system (16.4) if*

1. $u(\cdot) \in \mathcal{U}[0,T] := \left\{u : [0,T]\times\Omega \to U \mid u(\cdot) \text{ is } \{\mathcal{F}_t\}_{t\geq 0}\text{-adapted}\right\}$
2. *$x^{\alpha}(t)$ is the unique solution of (16.4) in the sense that for any $x^{\alpha}(t)$ and $\hat{x}^{\alpha}(t)$, satisfying (16.4),*

$$\mathsf{P}\left\{\omega \in \Omega : x^{\alpha}(t) = \hat{x}^{\alpha}(t)\right\} = 1$$

The set of all s-feasible controls is denoted by $\mathcal{U}^{s}_{feas}[0,T]$. The pair $(x^{\alpha}(t); u(\cdot))$, where $x^{\alpha}(t)$ is the solution of (16.4) corresponding to this $u(\cdot)$, is called an ***s-feasible pair****.*

The assumptions **A1**–**A4** guarantee that any $u(\cdot)$ from $\mathcal{U}[0,T]$ is s-feasible.

In addition, it is required that the following terminal state constraints are satisfied:

$$\boxed{\mathsf{E}\left\{h^{j}\left(x^{\alpha}(T)\right)\right\} \geq 0 \quad (j = 1,\ldots,l)} \tag{16.5}$$

where $h^{j} : \mathbb{R}^{n} \to \mathbb{R}$ are given functions.
A5: For $j = 1,\ldots,l$ the functions h^{j} are $L_{\phi}\left(C^{2}\right)$-mappings.

Definition 16.3. *The control $u(\cdot)$ and the pair $(x^{\alpha}(t); u(\cdot))$ are called an* ***s-admissible control*** *or* ***realizing the terminal condition*** *(16.5) and an s-admissible pair, respectively, if*

1. $u(\cdot) \in \mathcal{U}^{s}_{feas}[0,T]$
2. *$x^{\alpha}(t)$ is the solution of (16.4), corresponding to this $u(\cdot)$, such that the inequalities (16.5) are satisfied.*

The set of all s-admissible controls is denoted by $\mathcal{U}^{s}_{adm}[0,T]$.

16.2.3 Highest cost function and robust optimal control

Definition 16.4. *For any scalar-valued function $\varphi(\alpha)$ bounded on $\mathcal{A}$ define* ***the m-truth (or m-essential) maximum*** *of $\varphi(\alpha)$ on $\mathcal{A}$ as follows:*

$$m\text{ - }ess\ sup_{\alpha\in\mathcal{A}}\ \varphi(\alpha) := \max \varphi^{+}$$

such that

$$m\left\{\alpha \in \mathcal{A} : \varphi(\alpha) > \varphi^{+}\right\} = 0$$

It can be easily shown (see, for example, Yoshida (1979)) that the following *integral presentation* for the truth maximum holds:

$$\boxed{m\text{ - ess sup}_{\alpha\in\mathcal{A}}\varphi(\alpha) = \sup_{\mathcal{A}_0\subset\mathcal{A}:m(\mathcal{A}_0)>0} \frac{1}{m(\mathcal{A}_0)} \int_{\mathcal{A}_0} \varphi(\alpha)dm} \tag{16.6}$$

where the Lebesgue–Stiltijes integral is taken over all subsets $\mathcal{A}_0 \subset \mathcal{A}$ with positive measure $m(\mathcal{A}_0)$.

Consider the cost function h^{α} containing a terminal term, that is,

$$\boxed{h^{\alpha} := \mathsf{E}\left\{h^0(x^{\alpha}(T))\right\}} \tag{16.7}$$

Here $h_0(x)$ is a positive, bounded and smooth *cost function* defined on $\mathbb{R}^n$. The end time-point T is assumed to be finite and $x^{\alpha}(t) \in \mathbb{R}^n$.

If an admissible control is applied, for every $\alpha \in \mathcal{A}$ we deal with the cost value $h^{\alpha} = E\left\{h_0(x^{\alpha}(T))\right\}$ calculated at the terminal point $x^{\alpha}(T) \in \mathbb{R}^n$. Since the realized value of α is *a priori* unknown, define the *worst (highest) cost*

$$\boxed{\begin{aligned} F &= \sup_{\mathcal{A}_0\subset\mathcal{A}:m(\mathcal{A}_0)>0} \frac{1}{m(\mathcal{A}_0)} \int_{\mathcal{A}_0} E\left\{h^0(x^{\alpha}(T))\right\} dm \\ &= m\text{ - ess sup}_{\alpha\in\mathcal{A}} h^{\alpha} \end{aligned}} \tag{16.8}$$

The function F depends only on the considered admissible control $u(t)$, $t_0 \le t \le t_1$.

Definition 16.5. *The control $\bar{u}(t)$, $0 \le t \le T$ is said to be* ***robust optimal*** *if*

(i) it satisfies the terminal condition, that is, it is ***admissible****;*

(ii) it achieves the ***minimal*** *worst (highest) cost F^0 (among all admissible controls satisfying the terminal condition).*

*If the dynamics $\bar{x}^{\alpha}(t)$ correspond to this robust optimal control $\bar{u}(t)$ then $\left(\bar{x}^{\alpha}(\cdot), \bar{u}(\cdot)\right)$ is called an α-**robust optimal pair**.*

Thus the *robust optimization problem* consists of finding an admissible control action $u(t),\ 0 \le t \le T$, which provides

$$F^0 := F = \min_{u(t)} m\text{ - ess}\sup_{\alpha\in\mathcal{A}} h^{\alpha} = \min_{u(t)} \max_{\lambda\in\Lambda} \int_{\lambda\in\Lambda} \lambda(\alpha)\, E\left\{h^0(x^{\alpha}(T))\right\} dm(\alpha) \tag{16.9}$$

where the minimum is taken over all admissible control strategies and the maximum over all functions $\lambda(\alpha)$ within the so-called set of *'distribution densities'* Λ defined by

$$\Lambda := \left\{ \begin{array}{c} \lambda = \lambda(\alpha) = \mu(\alpha) \left(\int_{\alpha\in\mathcal{A}} \mu(\alpha)\, dm(\alpha) \right)^{-1} \ge 0 \\ \int_{\alpha\in\mathcal{A}} \lambda(\alpha)\, dm(\alpha) = 1 \end{array} \right\} \tag{16.10}$$

This is the *stochastic min-max Bolza problem.*

16.3 Robust stochastic maximum principle

16.3.1 First- and second-order adjoint processes

The adjoint equations and the associated Hamiltonian function are introduced in this section to present the *necessary conditions* of the robust optimality for the considered class of partially unknown stochastic systems – called the *robust stochastic maximum principle* (RSMP). If in the deterministic case (Boltyanskii and Poznyak, 1999) the adjoint equations are backward ordinary differential equations and represent, in some sense, the same forward equation but in reverse time, in the stochastic case such interpretation is not applicable because any time reversal may destroy the non-anticipativeness of the stochastic solutions, that is, any obtained robust control should not depend on the future. To avoid these problems the approach given in Zhou (1991) is used that takes into account the adjoint equations introduced for any fixed value of the parameter α and, hence, some of the results from Zhou (1991) may be applied directly without any changes.

So, for any $\alpha \in \mathcal{A}$ and any admissible control $u(\cdot) \in \mathcal{U}^s_{adm}[0, T]$ consider

- *the 1-st order vector adjoint equations:*

$$\boxed{\begin{aligned} d\psi^{\alpha}(t) &= -\Bigg[b^{\alpha}_x\left(t, x^{\alpha}(t), u(t)\right)^{\intercal} \psi^{\alpha}(t) \\ &\quad + \sum_{j=1}^{m} \sigma^{\alpha j}_x \left(t, x^{\alpha}(t), u(t)\right)^{\intercal} q^{\alpha}_j(t) \Bigg] dt + q^{\alpha}(t) dW(t) \\ \psi^{\alpha}(T) &= c^{\alpha}, \quad t \in [0, T] \end{aligned}} \tag{16.11}$$

- the *2-nd order matrix adjoint equations:*

$$
\boxed{\begin{aligned}
d\Psi^{\alpha}(t) = &-\Big[b_x^{\alpha}(t, x^{\alpha}(t), u(t))^{\mathsf{T}} \Psi^{\alpha}(t) \\
&+ \Psi^{\alpha}(t) b_x^{\alpha}\left(t, x^{\alpha}(t), u(t)\right)\Big] \\
&+ \sum_{j=1}^{m} \sigma_x^{\alpha j}\left(t, x^{\alpha}(t), u(t)\right)^{\mathsf{T}} \Psi^{\alpha}(t) \sigma_x^{\alpha j}\left(t, x^{\alpha}(t), u(t)\right) \\
&+ \sum_{j=1}^{m} \Big(\sigma_x^{\alpha j}\left(t, x^{\alpha}(t), u(t)\right)^{\mathsf{T}} Q_j^{\alpha}(t) \\
&+ Q_j^{\alpha}(t) \sigma_x^{\alpha j}\left(t, x^{\alpha}(t), u(t)\right)\Big) \\
&+ H_{xx}^{\alpha}\left(t, x^{\alpha}(t), u(t), \psi^{\alpha}(t), q^{\alpha}(t)\right)\Big] dt \\
&+ \sum_{j=1}^{m} Q_j^{\alpha}(t)\, dW^j(t) \\
\Psi^{\alpha}(T) = C^{\alpha}, &\quad t \in [0, T]
\end{aligned}} \tag{16.12}
$$

Here $c^{\alpha} \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R}^n)$ is a square integrable $\mathcal{F}_T$-measurable $\mathbb{R}^n$-valued random vector, $\psi^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R}^n)$ is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^n$-valued vector random process and

$$
q^{\alpha}(t) \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{n\times m}\right)
$$

is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^{n\times m}$-valued matrix random process. Similarly, $C^{\alpha} \in L^2_{\mathcal{F}_T}\left(\Omega, \mathbb{R}^{n\times n}\right)$ is a square integrable $\mathcal{F}_T$-measurable $\mathbb{R}^{n\times n}$-valued random matrix, $\Psi^{\alpha}(t) \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{n\times n}\right)$ is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^{n\times n}$-valued matrix random process and $Q_j^{\alpha}(t) \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{n\times m}\right)$ is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^{n\times n}$-valued matrix random process. $b_x^{\alpha}(t, x^{\alpha}, u)$ and $H_{xx}^{\alpha}(t, x^{\alpha}, u, \psi^{\alpha}, q^{\alpha})$ are the first and, correspondingly, the second derivatives of these functions by x^{α}. The function $H^{\alpha}(t, x, u, \psi, q)$ is defined as

$$
\boxed{H^{\alpha}(t, x, u, \psi, q) := b^{\alpha}(t, x, u)^{\mathsf{T}} \psi + \operatorname{tr}\left[q^{\mathsf{T}} \sigma^{\alpha}\right]} \tag{16.13}
$$

As it is seen from (16.12), if $C^{\alpha} = C^{\alpha\mathsf{T}}$ then for any $t \in [0, T]$ the random matrix $\Psi^{\alpha}(t)$ is symmetric (but not necessarily positive or negative definite). In (16.11) and (16.12), which are the backward stochastic differential equations with the $\{\mathcal{F}_t\}_{t\geq 0}$-adapted solutions, the unknown variables to be selected are the pair of terminal conditions c^{α}, C^{α} and the collection $\left(q^{\alpha}, Q_j^{\alpha}\ (j = 1, \ldots, l)\right)$ of $\{\mathcal{F}_t\}_{t\geq 0}$-adapted stochastic matrices. Note that the equations (16.4) and (16.11) can be rewritten in Hamiltonian form as

$$
\boxed{\begin{aligned}
dx^{\alpha}(t) = &\; H_{\psi}^{\alpha}\left(t, x^{\alpha}(t), u(t)\right)^{\mathsf{T}} \psi^{\alpha}(t), q^{\alpha}(t))dt \\
&+ \sigma^{\alpha}\left(t, x^{\alpha}(t), u(t)\right) dW(t) \\
x^{\alpha}(0) = x_0, &\quad t \in [0, T]
\end{aligned}} \tag{16.14}
$$

$$\boxed{\begin{aligned} &d\psi^{\alpha}(t) = -H_x^{\alpha}\left(t, x^{\alpha}(t), u(t)\right)^{\mathsf{T}} \psi^{\alpha}(t), q^{\alpha}(t))dt \\ &\qquad\qquad + q^{\alpha}(t)dW(t) \\ &\psi^{\alpha}(T) = c^{\alpha}, \quad t \in [0, T] \end{aligned}} \tag{16.15}$$

16.3.2 Main result

Now the main result of this chapter can be formulated.

Theorem 16.1. (robust stochastic maximum principle) *Let A1–A5 be fulfilled and* $\left(\bar{x}^{\alpha}(\cdot), \bar{u}(\cdot)\right)$ *be the* α*-robust optimal pairs* ($\alpha \in \mathcal{A}$). *The parametric uncertainty set* $\mathcal{A}$ *is a space with countable additive measure* $m(\alpha)$ *which is assumed to be given. Then for every* $\varepsilon > 0$ *there exist collections of terminal conditions* $c^{\alpha,(\varepsilon)}$, $C^{\alpha,(\varepsilon)}$, $\{\mathcal{F}_t\}_{t\geq 0}$ *-adapted stochastic matrices*

$$\left(q^{\alpha,(\varepsilon)}, Q_j^{\alpha,(\varepsilon)}\ (j = 1, \ldots, l)\right)$$

in (16.11) and (16.12), and nonnegative constants $\mu_{\alpha}^{(\varepsilon)}$ *and* $\nu_{\alpha j}^{(\varepsilon)}$ $(j = 1, \ldots, l)$ *such that the following conditions are fulfilled:*
*1. (**Complementary slackness condition**): for any* $\alpha \in \mathcal{A}$

$$\boxed{\begin{aligned} &(i)\ \left| E\left\{h^0\left(\bar{x}^{\alpha}(T)\right)\right\} - \max_{\alpha\in\mathcal{A}} E\left\{h^0\left(\bar{x}^{\alpha}(T)\right)\right\}\right| < \varepsilon \\ &\qquad \text{or } \mu_{\alpha}^{(\varepsilon)} = 0; \\ &(ii)\ \left| E\left\{h^j\left(\bar{x}^{\alpha}(T)\right)\right\}\right| < \varepsilon \text{ or } \nu_{\alpha j}^{(\varepsilon)} = 0 (j = 1, \ldots, l); \end{aligned}} \tag{16.16}$$

*2. (**Transversality condition**): for any* $\alpha \in \mathcal{A}$ *the inequality*

$$\left\| c^{\alpha,(\varepsilon)} + \mu_{\alpha}^{(\varepsilon)} h_x^0\left(\bar{x}^{\alpha}(T)\right) + \sum_{j=1}^{l} \nu_{\alpha j}^{(\varepsilon)} h_x^j\left(\bar{x}^{\alpha}(T)\right)\right\| < \varepsilon \quad \mathsf{P}\text{-}a.s. \tag{16.17}$$

$$\left\| C^{\alpha,(\varepsilon)} + \mu_{\alpha}^{(\varepsilon)} h_{xx}^0\left(\bar{x}^{\alpha}(T)\right) + \sum_{j=1}^{l} \nu_{\alpha j}^{(\varepsilon)} h_{xx}^j\left(\bar{x}^{\alpha}(T)\right)\right\| < \varepsilon \quad \mathsf{P}\text{-}a.s. \tag{16.18}$$

hold;
*3. (**Nontriviality condition**): there exists a set* $\mathcal{A}_0 \subset \mathcal{A}$ *with positive measure* $m(\mathcal{A}_0) > 0$ *such that for every* $\alpha \in \mathcal{A}_0$ *either* $c^{\alpha,(\varepsilon)} \overset{a.s.}{\neq} 0$ *or, at least, one of the numbers* $\mu_{\alpha}^{(\varepsilon)}$, $\nu_{\alpha j}^{(\varepsilon)}$ $(j = 1, \ldots, l)$ *is distinct from* 0, *that is, with probability one*

$$\boxed{\forall \alpha \in \mathcal{A}_0 \in \mathcal{A} : \left|c^{\alpha,(\varepsilon)}\right| + \mu_{\alpha}^{(\varepsilon)} + \sum_{j=1}^{l} \nu_{\alpha j}^{(\varepsilon)} > 0} \tag{16.19}$$

*4. (**Maximality condition**): the robust optimal control $\bar{u}(\cdot)$ for almost all $t \in [0, T]$ maximizes the generalized Hamiltonian function*

$$\boxed{\begin{aligned}&\mathcal{H}\left(t, \bar{x}^{\diamond}(t), u, \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right)\\ &\quad := \int_{\mathcal{A}} \mathcal{H}^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), \Psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) dm(\alpha)\end{aligned}} \tag{16.20}$$

where

$$\boxed{\begin{aligned}&H^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), \Psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right)\\ &\quad := H^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) - \frac{1}{2}\operatorname{tr}\left[\bar{\sigma}^{\alpha\mathsf{T}}\Psi^{\alpha,(\varepsilon)}(t)\,\bar{\sigma}^{\alpha}\right]\\ &\qquad + \frac{1}{2}\operatorname{tr}\left[\left(\sigma^{\alpha}\left(t, \bar{x}^{\alpha}(t), u\right) - \bar{\sigma}^{\alpha}\right)^{\mathsf{T}}\Psi^{\alpha,(\varepsilon)}(t)\left(\sigma^{\alpha}\left(t, \bar{x}^{\alpha}(t), u\right) - \bar{\sigma}^{\alpha}\right)\right]\end{aligned}} \tag{16.21}$$

the function $H^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right)$ is given by (16.13),

$$\boxed{\bar{\sigma}^{\alpha} := \sigma^{\alpha}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right)} \tag{16.22}$$

$$\begin{aligned}&\bar{x}^{\diamond}(t) := \left(\bar{x}^{1\mathsf{T}}(t), \ldots, \bar{x}^{N\mathsf{T}}(t)\right)^{\mathsf{T}}, \psi^{\diamond,(\varepsilon)}(t)\\ &\qquad := \left(\psi^{1,(\varepsilon)\mathsf{T}}(t), \ldots, \psi^{N,(\varepsilon)\mathsf{T}}(t)\right)^{\mathsf{T}}\\ &q^{\diamond,(\varepsilon)}(t) := \left(q^{1,(\varepsilon)}(t), \ldots, q^{N,(\varepsilon)}(t)\right), \Psi^{\diamond,(\varepsilon)}(t)\\ &\qquad := \left(\Psi^{1,(\varepsilon)}(t), \ldots, \Psi^{N,(\varepsilon)}(t)\right)\end{aligned}$$

and $\psi^{i,(\varepsilon)\mathsf{T}}(t)$, $\Psi^{i,(\varepsilon)}(t)$ verify (16.11) and (16.12) with the terminal conditions $c^{\alpha,(\varepsilon)}$ and $C^{\alpha,(\varepsilon)}$, respectively, i.e., for almost all $t \in [0, T]$

$$\boxed{\bar{u}(t) = \arg\max_{u \in U} H\left(t, \bar{x}^{\diamond}(t), u, \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right)} \tag{16.23}$$

16.4 Proof of Theorem 16.1

16.4.1 Formalism

Consider the random vector space $\mathbb{R}^{\diamond}$ with the coordinates $x^{\alpha,i} \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R})$ ($\alpha \in \mathcal{A}$, $i = 1, \ldots, n$). For each fixed $\alpha \in \mathcal{A}$ we may consider

$$x^{\alpha} := \left(x^{\alpha,1}, \ldots, x^{\alpha,ni}\right)^{\mathsf{T}}$$

as an element of a Hilbert (and, hence, self-conjugate) space $\mathbb{R}^\alpha$ with the usual scalar product given by

$$\left\langle x^\alpha, \tilde{x}^\alpha \right\rangle := \sqrt{\sum_{i=1}^{n} \mathsf{E}\left\{ x^{\alpha,i} \tilde{x}^{\alpha,i} \right\}}, \quad \left\| \tilde{x}^\alpha \right\| := \sqrt{\langle x^\alpha, x^\alpha \rangle}$$

However, in the whole space $\mathbb{R}^\diamond$ introduce the norm of the element $x^\diamond = \left(x^{\alpha,i}\right)$ in another way:

$$\left\| x^\diamond \right\| := m\text{ - ess}\sup_{\alpha \in \mathcal{A}} \sqrt{\sum_{i=1}^{n} \mathsf{E}\left\{ \left(x^{\alpha,i}\right)^2 \right\}}$$
$$= \sup_{\mathcal{A}_0 \subset \mathcal{A}: m(\mathcal{A}_0) > 0} \frac{1}{m(\mathcal{A}_0)} \int_P \sqrt{\sum_{i=1}^{n} \mathsf{E}\left\{ \left(x^{\alpha,i}\right)^2 \right\}} dm \tag{16.24}$$

Consider the set $\mathbb{R}^\diamond$ of all functions from $L^2_{\mathcal{F}_T}(\Omega, \mathbb{R})$ for any fixed $\alpha \in \mathcal{A}$, measurable on $\mathcal{A}$ and with values in $\mathbb{R}^n$, identifying every two functions which coincide almost everywhere. With the norm (16.24), $\mathbb{R}^\diamond$ is a Banach space. Now we describe its conjugate space $\mathbb{R}_\diamond$. Consider the set of all measurable functions $a(\alpha) \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R})$ defined on $\mathcal{A}$ with values in $\mathbb{R}^n$. It consists of all covariant random vectors $a_\diamond = \left(a_{\alpha,i}\right)$ $(\alpha \in \mathcal{A}, i = 1, \ldots, n)$ with the norm

$$\|a_\diamond\| := m\text{ - ess}\sup_{\alpha \in \mathcal{A}} \sqrt{\sum_{i=1}^{n} \mathsf{E}\left\{ \left(a_{\alpha,i}\right)^2 \right\}} \tag{16.25}$$

The set of all such functions $a(\alpha)$ is a linear normed space. In general, this normed space is not complete. The following example illustrates this fact.

Example 16.1. *Consider the case when A is the segment $[0, 1] \subset \mathbb{R}$ with the usual Lebesgue measure. Let $\varphi_k(\alpha)$ be the function on $[0, 1]$ that it is equal to 0 for $\alpha > \frac{1}{k}$ and is equal to k for $0 \leq \alpha \leq \frac{1}{k}$. Then $\int_{\mathcal{A}} \varphi_k(\alpha) d\alpha = 1$, and the sequence $\varphi_k(\alpha)$ $k = 1, 2, \ldots$ is a fundamental one in the norm (16.25). But their limit function $\lim_{k\to\infty} \varphi_k(\alpha)$ does not exist among measurable and summable functions. Such a limit is the Dirak function $\varphi^{(0)}(\alpha)$ which is equal to 0 for every $\alpha > 0$ and is equal to infinity at $\alpha = 0$ (with the normalization agreement that $\int_{\mathcal{A}} \varphi^{(0)}(\alpha) d\alpha = 1$).*

This example shows that the linear normed space of all measurable, summable functions with the norm (16.25) is, in general, incomplete. The complementation of this space is a Banach space, and we denote it by $\mathbb{R}_\diamond$. This is the conjugate space for $\mathbb{R}^\diamond$. The scalar product of $x^\diamond \in \mathbb{R}^\diamond$ and $a_\diamond \in \mathbb{R}_\diamond$ can be defined as

$$\left\langle a_{\diamond}, x^{\diamond} \right\rangle_E := \int\limits_{\mathcal{A}} \sum_{i=1}^{n} \mathsf{E} \left\{ a_{\alpha,i} x^{\alpha,i} \right\} dm$$

for which the Cauchy–Schwartz inequality evidently holds:

$$\left\langle a_{\diamond}, x^{\diamond} \right\rangle_E \leq \|a_{\diamond}\| \cdot \left\| x^{\diamond} \right\|$$

16.4.2 Proof of Properties 1–3

In this subsection consider the vector $x^{\diamond}(T)$ only.

The index $\alpha \in \mathcal{A}$ is said to be $\varepsilon \wedge h^0$*-active* if the given $\varepsilon > 0$

$$\mathsf{E}\left\{h^0\left(\bar{x}^{\alpha}(T)\right)\right\} > \max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{h^0\left(\bar{x}^{\alpha}(T)\right)\right\} - \varepsilon \tag{16.26}$$

and, it is $\varepsilon \wedge h^j$*-active* if

$$\mathsf{E}\left\{h^j\left(\bar{x}^{\alpha}(T)\right)\right\} > -\varepsilon \tag{16.27}$$

First, assume that there exists a set of a positive measure $G \subset \mathcal{A}$ and a set $\bar{\Omega} \subseteq \Omega$ $\left(\mathsf{P}\left\{\omega \in \bar{\Omega}\right\} > 0\right)$ such that for all $\varepsilon \wedge h^0$-active indices $\alpha \in \mathcal{A}$ we have that $\left\| h_x^0 \left(\bar{x}^{\alpha}(T)\right)\right\| < \varepsilon$ for all $\omega \in \bar{\Omega} \subseteq \Omega$ and almost everywhere on G. Then selecting (without violation of the transversality and nontriviality conditions)

$$\mu_{\alpha}^{(\varepsilon)} \neq 0, \quad \mu_{\tilde{\alpha}\neq\alpha}^{(\varepsilon)} = 0, \quad \nu_{\alpha j}^{(\varepsilon)} = 0 \quad (\forall \alpha \in \mathcal{A}, j = 1, \ldots, l)$$

it follows that

$$c^{\alpha,(\varepsilon)} = \psi^{\alpha,(\varepsilon)}(T) = 0, \quad C^{\alpha,(\varepsilon)} = \Psi^{\alpha,(\varepsilon)}(T) = 0$$

for almost all $\omega \in \bar{\Omega}$ and almost everywhere on G. In this situation, the only nonanticipative matrices $q^{\alpha,(\varepsilon)}(t) = 0$ and $Q_j^{\alpha,(\varepsilon)}(t) = 0$ are admissible, and for all $t \in [0, T]$, as a result,

$$H^{\alpha}(t, x, u, \psi, q) = 0, \quad \psi^{\alpha,(\varepsilon)}(t) = 0$$
$$\Psi^{\alpha,(\varepsilon)}(t) = 0$$

and for almost all $\omega \in \bar{\Omega}$ and almost everywhere on G. Thus, all conditions 1–4 of the theorem are satisfied automatically whether or not the control is robust optimal or not. So, it can be assumed that

$$\left\| h_x^0\left(\bar{x}^{\alpha}(T)\right)\right\| \geq \varepsilon \quad (\mathsf{P}\text{-}a.s.)$$

for all $\varepsilon \wedge h^0$-active indices $\alpha \in \mathcal{A}$. Similarly, it can be assumed that

$$\left\| h_x^0\left(\bar{x}^{\alpha}(T)\right)\right\| \geq \varepsilon \quad (\mathsf{P}\text{-}a.s.)$$

for all $\varepsilon \wedge h^j$-active indices $\alpha \in \mathcal{A}$.

Denote by $\Omega_1 \subseteq \mathbb{R}^{\diamond}$ the *controllability region*, that is, the set of all points $z^{\diamond} \in \mathbb{R}^{\diamond}$ such that there exists a feasible control $u(t) \in \mathcal{U}^s_{feas}[0,T]$ for which the trajectories $x^{\diamond}(t) = \left(x^{\alpha,i}(t)\right)$, corresponding to (16.4), satisfy $x^{\diamond}(T) = z^{\diamond}$ with probability 1:

$$\Omega_1 := \left\{ z^{\diamond} \in \mathbb{R}^{\diamond} : x^{\diamond}(T) \overset{a.s.}{=} z^{\diamond}, u(t) \in \mathcal{U}^s_{feas}[0,T], x^{\alpha}(0) = x_0 \right\} \tag{16.28}$$

Let $\Omega_{2,j} \subseteq \mathbb{R}^{\diamond}$ denote the set of all points $z^{\diamond} \in \mathbb{R}^{\diamond}$ satisfying the terminal condition (16.5) for some fixed index j and any $\alpha \in \mathcal{A}$, that is,

$$\Omega_{2j} := \left\{ z^{\diamond} \in \mathbb{R}^{\diamond} : \mathsf{E}\left\{ h^j\left(z^{\alpha}\right)\right\} \geq 0 \; \forall \alpha \in \mathcal{A} \right\} \tag{16.29}$$

Finally, denote by $\Omega_0^{(\varepsilon)} \subseteq \mathbb{R}^{\diamond}$ the set, containing the optimal point $\bar{x}^{\diamond}(T)$ (corresponding to the given robust optimal control $\bar{u}(\cdot)$) as well as all points $z^{\diamond} \in \mathbb{R}^{\diamond}$ satisfying for all $\alpha \in \mathcal{A}$

$$\mathsf{E}\left\{ h^0\left(z^{\alpha}\right)\right\} \leq \max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{ h^0\left(\bar{x}^{\alpha}(T)\right)\right\} - \varepsilon$$

that is, $\forall \alpha \in \mathcal{A}$

$$\Omega_0^{(\varepsilon)} := \left\{ \bar{x}^{\diamond}(T) \cup z^{\diamond} \in \mathbb{R}^{\diamond} : \mathsf{E}\left\{ h^0\left(z^{\alpha}\right)\right\} \leq \max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{ h^0\left(\bar{x}^{\alpha}(T)\right)\right\} - \varepsilon \right\} \tag{16.30}$$

In view of these definitions, if only the control $\bar{u}(\cdot)$ is robust optimal (locally), then

$$\Omega_0^{(\varepsilon)} \cap \Omega_1 \cap \Omega_{21} \cap \cdots \cap \Omega_{2l} = \left\{ \bar{x}^{\diamond}(T) \right\} \qquad \mathsf{P}\text{-}a.s. \tag{16.31}$$

Hence, if $K_0^{\diamond}, K_1^{\diamond}, K_{21}^{\diamond}, \ldots, K_{2l}^{\diamond}$ are the cones (the local tents) of the sets $\Omega_0^{(\varepsilon)}, \Omega_1, \Omega_{21}, \ldots, \Omega_{2l}$ at their common point $\bar{x}^{\diamond}(T)$, then these cones are *separable* and the Neustad Theorem 1 in Kushner (1972) is satisfied, that is, for any point $z^{\diamond} \in \mathbb{R}^{\diamond}$ there exist linear independent functionals $\mathbf{l}_s\left(\bar{x}^{\diamond}(T), z^{\diamond}\right)$ $(s = 0, 1, 2j;\ j = 1, \ldots, l)$ satisfying

$$\mathbf{l}_0\left(\bar{x}^{\diamond}(T), z^{\diamond}\right) + \mathbf{l}_1\left(\bar{x}^{\diamond}(T), z^{\diamond}\right) + \sum_{j=1}^{l} \mathbf{l}_{2s}\left(\bar{x}^{\diamond}(T), z^{\diamond}\right) \geq 0 \tag{16.32}$$

The implementation of the Riesz representation theorem for linear functionals (Yoshida, 1979) implies the existence of the covariant random vectors $v^s_{\diamond}\left(z^{\diamond}\right)$ $(s = 0, 1,\ 2j;\ j = 1, \ldots, l)$ belonging to the polar cones $K_{s\diamond}$, respectively, not equal to zero simultaneously and satisfying

$$\mathbf{l}_s\left(\bar{x}^{\diamond}(T), z^{\diamond}\right) = \left\langle v^s_{\diamond}\left(z^{\diamond}\right), z^{\diamond} - \bar{x}^{\diamond}(T)\right\rangle_E \tag{16.33}$$

The relations (16.32) and (16.33), and taking into account that they hold for any $z^{\diamond} \in \mathbb{R}^{\diamond}$, imply the property

$$v_{\diamond}^{0}\left(\bar{x}^{\diamond}(T)\right)+v_{\diamond}^{1}\left(\bar{x}^{\diamond}(T)\right)+\sum_{j=1}^{l} v_{\diamond}^{sj}\left(\bar{x}^{\diamond}(T)\right)=0 \qquad \mathsf{P}\text{-}a.s. \tag{16.34}$$

Consider then the possible structures of these vectors.
(a) Denote

$$\Omega_{0}^{\alpha}:=\left\{z^{\alpha} \in \mathbb{R}^{\alpha}: \left\{\mathsf{E}\left\{h^{0}\left(z^{\alpha}\right)\right\}>\max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{h^{0}\left(\bar{x}^{\alpha}(T)\right)\right\}-\varepsilon\right\} \cup\left\{\bar{x}^{\alpha}(T)\right\}\right\}$$

Taking into account that $h^{0}\left(z^{\alpha}\right)$ is $L_{\phi}\left(C^{2}\right)$-mapping and in view of the identity

$$\begin{aligned} h(x)-h(\bar{x}) &= h_{x}(\bar{x})^{\mathsf{T}}(x-\bar{x}) \\ &\quad+\int_{\theta=0}^{1} \operatorname{tr}\left[\theta h_{xx}(\theta \bar{x}+(1-\theta) x)(x-\bar{x})(x-\bar{x})^{\mathsf{T}}\right] d\theta \end{aligned} \tag{16.35}$$

which is valid for any twice differentiable function $h: \mathbb{R}^{n} \rightarrow \mathbb{R}$ and $x, \acute{x} \in \mathbb{R}^{n}$, it follows that

$$\begin{aligned} \mathsf{E}\left\{h^{0}\left(\bar{x}^{\alpha}(T)\right)\right\} &= \mathsf{E}\left\{h^{0}\left(z^{\alpha}\right)\right\}+\left\langle h_{x}^{0}\left(z^{\alpha}\right),\left(\bar{x}^{\alpha}(T)-z^{\alpha}\right)\right\rangle_{E} \\ &\quad+\mathsf{E}\left\{O\left(\left\|z^{\alpha}-\bar{x}^{\alpha}(T)\right\|^{2}\right)\right\} \end{aligned} \tag{16.36}$$

So, the corresponding cone K_{0}^{α} at the point $\bar{x}^{\diamond}(T)$ may be described as

$$K_{0}^{\alpha}:=\begin{cases}\left\{z^{\alpha} \in \mathbb{R}^{\alpha}: \left\langle h_{x}^{0}\left(z^{\alpha}\right),\left(\bar{x}^{\alpha}(T)-z^{\alpha}\right)\right\rangle_{E} \geq 0\right\} & \text{if } \alpha \text{ is } \varepsilon \wedge h^{0}\text{-active} \\ \mathbb{R}^{\alpha} \text{ if } \alpha \text{ is } \varepsilon \wedge h^{0}\text{-inactive}\end{cases}$$

Then the direct sum $K_{0}^{\diamond}:=\bigoplus_{\alpha \in \mathcal{A}} K_{0}^{\alpha}$ is a convex cone with apex point $\bar{x}^{\alpha}(T)$ and, at the same time, it is the tent $\Omega_{0}^{(\varepsilon)}$ at the same apex point. The polar cone $K_{0\diamond}$ can be presented as

$$K_{0\diamond}=\operatorname{conv}\left(\bigcup_{\alpha \in \mathcal{A}} K_{0\alpha}\right)$$

(here $K_{0\alpha}$ is a the polar cone to $K_{0}^{\alpha} \subseteq \mathbb{R}^{\alpha}$). Since

$$v_{\diamond}^{0}\left(z^{\diamond}\right)=\left(v_{\alpha}^{0}\left(z^{\alpha}\right)\right) \in K_{0\diamond}$$

then $K_{0\alpha}$ should have the form

$$v_{\alpha}^{0}\left(z^{\diamond}\right)=\mu_{\alpha}^{(\varepsilon)} h_{x}^{0}\left(z^{\diamond}\right) \tag{16.37}$$

where $\mu_{\alpha}^{(\varepsilon)} \geq 0$ and $\mu_{\alpha}^{(\varepsilon)}=0$ if α is $\varepsilon \wedge h^{0}$-inactive. So, the statement (1-i) (*Complementary slackness*) is proven.

(b) Now consider the set Ω_{2j}, containing all random vectors $z^{\diamond}$ admissible by the terminal condition (16.5) for some fixed index j and any $\alpha \in \mathcal{A}$. Defining for any α and the fixed index j the set

$$\Omega_{2j}^{\alpha} := \left\{ z^{\alpha} \in \mathbb{R}^{\alpha} : \mathsf{E}\left\{ h^{j}\left(z^{\alpha}\right)\right\} \geq -\varepsilon \right\}$$

in view of (16.36) applied for the function h^j, it follows that

$$K_{2j}^{\alpha} := \begin{cases} \left\{ z^{\alpha} \in \mathbb{R}^{\alpha} : \left\langle h_x^j\left(z^{\alpha}\right)^{\mathsf{T}}\left(z^{\alpha} - \bar{x}^{\alpha}(T)\right)\right\rangle_E \geq 0 \right\} & \text{if } \alpha \text{ is } \varepsilon \wedge h^j\text{-active} \\ \mathbb{R}^{\alpha} \text{ if } \alpha \text{ is } \varepsilon \wedge h^j\text{-inactive} \end{cases}$$

Let $\Omega_{2j} = \underset{\alpha\in\mathcal{A}}{\oplus} \Omega_{2j}^{\alpha}$ and $K_{2j}^{\diamond} = \underset{\alpha\in\mathcal{A}}{\oplus} K_{2j}^{\alpha}$. By analogy with the above,

$$K_{2j\diamond} = \operatorname{conv}\left(\bigcup_{\alpha\in\mathcal{A}} K_{2j\alpha}\right)$$

is the polar cone, and hence, $K_{2j\alpha}$ should consist of all

$$v_{\alpha}^{2j}\left(z^{\alpha}\right) = \nu_{\alpha j}^{(\varepsilon)} h_x^j\left(z^{\alpha}\right) \tag{16.38}$$

where $\nu_{\alpha j}^{(\varepsilon)} \geq 0$ and $\nu_{\alpha j}^{(\varepsilon)} = 0$ if α is $\varepsilon \wedge h^j$-inactive. So, the statement (1-ii) (*Complementary slackness*) is also proven.

(c) Consider the polar cone $K_{1\diamond}$. Let us introduce the so-called *needle-shape (or spike)* variation $u^{\delta}(t)$ $(\delta > 0)$ of the robust optimal control $\bar{u}(t)$ at the time region $[0, T]$ as follows:

$$u^{\delta}(t) := \begin{cases} \bar{u}(t) & \text{if } [0, T+\delta] \backslash T_{\delta_n} \\ u(t) \in \mathcal{U}_{feas}^{s}[0, T] & \text{if } t \in T_{\delta_n} \end{cases} \tag{16.39}$$

where $T_{\delta} \subseteq [0, T]$ is a measurable set with the Lebesgue measure $|T_{\delta}| = \delta$, $u(t)$ being any s-feasible control. Here it is assumed that $\bar{u}(t) = \bar{u}(T)$ for any $t \in [T, T+\delta]$. It is clear from this construction that $u^{\delta}(t) \in \mathcal{U}_{feas}^{s}[0, T]$ and, hence, the corresponding trajectories $x^{\diamond}(t) = \left(x^{\alpha,i}(t)\right)$, given by (16.4), also make sense. Denote by

$$\Delta^{\alpha} := \lim_{\delta\to 0} \delta^{-1}\left[x^{\alpha}(T) - \bar{x}^{\alpha}(T)\right]$$

the corresponding *displacement vector* (here the limit exists because of the differentiability of the vector $x^{\alpha}(t)$ at the point $t = T$). By the definition, Δ^{α} is a tangent vector of the controllability region Ω_1. Moreover, the vector

$$g^{\diamond}(\beta)\,|_{\beta=\pm 1} := \lim_{\delta\to 0} \delta^{-1}\left[\int_{s=T}^{T+\beta_{\delta}} b^{\diamond}(s, x(s), u(s))\,dt + \int_{s=T}^{T+\beta_{\delta}} \sigma^{\diamond}(s, x(s), u(s))\,dW(s)\right]$$

is also the tangent vector for Ω_1, since

$$x^{\diamond}(T+\beta_{\delta}) = x^{\diamond}(T) + \int_{s=T}^{T+\beta_{\delta}} b^{\alpha}(s, x(s), u(s))\, dt + \int_{s=T}^{T+\beta_{\delta}} \sigma^{\alpha}(s, x(s), u(s))\, dW(s)$$

Denoting by Q_1 the cone (linear combination of vectors with non-negative coefficients) generated by all displacement vectors Δ^{α} and the vectors $g^{\diamond}(\pm 1)$, it is concluded that $K_1^{\diamond} = \bar{x}^{\alpha}(T) + Q_1$. Hence

$$v_{\diamond}^{1}\left(z^{\alpha}\right) = c^{\diamond,(\varepsilon)} \in K_{1\diamond} \tag{16.40}$$

(d) Substituting (16.37), (16.38) and (16.40) into (16.34), *the transversality condition* (16.17) is obtained. Since at least one of the vectors

$$v_{\diamond}^{0}\left(z^{\alpha}\right), v_{\diamond}^{1}\left(z^{\alpha}\right), v_{\diamond}^{21}\left(z^{\alpha}\right), \ldots, v_{\diamond}^{2l}\left(z^{\alpha}\right)$$

should be distinct from zero at the point $z^{\alpha} = \bar{x}^{\alpha}(T)$, *the nontriviality condition* is obtained too. *The transversality condition* (16.18) can be satisfied by the corresponding selection of the matrices $C^{\alpha,(\varepsilon)}$. The statement 3 is also proven.

16.4.3 Proof of Property 4 (maximality condition)

This part of the proof seems to be more delicate and needs some additional constructions. In view of (16.33), (16.34), (16.37), (16.38) and (16.40), for $z = x^{\alpha}(T)$ the inequality (16.32) can be represented as follows:

$$\begin{aligned} 0 \le F_{\delta}\left(u^{\delta}(\cdot)\right) &:= \mathbf{l}_0\left(\bar{x}^{\diamond}(T), x^{\alpha}(T)\right) \\ &\quad + \mathbf{l}_1\left(\bar{x}^{\diamond}(T), x^{\alpha}(T)\right) + \sum_{j=1}^{l} \mathbf{l}_{2s}\left(\bar{x}^{\diamond}(T), x^{\alpha}(T)\right) \\ &= \sum_{\alpha\in\mathcal{A}} \left[\mu_{\alpha}^{(\varepsilon)} \left\langle h_x^0\left(x^{\alpha}(T)\right), x^{\alpha}(T) - \bar{x}^{\alpha}(T) \right\rangle_E \right. \\ &\quad + \left\langle c^{\alpha,(\varepsilon)}, x^{\alpha}(T) - \bar{x}^{\alpha}(T) \right\rangle_E \\ &\quad \left. + \sum_{j=1}^{l} \nu_{\alpha j}^{(\varepsilon)} \left\langle h_x^j\left(x^{\alpha}(T)\right), x^{\alpha}(T) - \bar{x}^{\alpha}(T) \right\rangle_E \right] \end{aligned} \tag{16.41}$$

valid for any s-feasible control $u^{\delta}(t)$.

As has been shown in Zhou (1991) and Yong and Zhou (1999), any $u^{\delta}(t) \in \mathcal{U}^{s}_{feas}[0,T]$ provides the following trajectory variation:

$$x^{\alpha}(t) - \bar{x}^{\alpha}(t) = y^{\delta\alpha}(t) + z^{\delta\alpha}(t) + o^{\delta\alpha}_{\omega}(t) \tag{16.42}$$

where $y^{\delta\alpha}(t)$, $z^{\delta\alpha}(t)$ and $o^{\delta\alpha}(t)$ are $\{\mathcal{F}_t\}_{t\geq 0}$-adapted stochastic vector processes satisfying (for the simplification of the calculations given below, the argument dependence is omitted) the following equations:

$$\begin{aligned} dy^{\delta\alpha} &= b^{\alpha}_{x} y^{\delta\alpha} dt + \sum_{j=1}^{m} \left[\sigma^{\alpha j}_{x} y^{\delta\alpha} + \Delta\sigma^{\alpha j} \chi_{T_\delta} \right] dW^{j} \\ y^{\delta\alpha}(0) &= 0 \end{aligned} \tag{16.43}$$

where

$$\begin{aligned} b^{\alpha}_{x} &:= b^{\alpha}_{x}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right), \; \sigma^{\alpha j}_{x} := \sigma^{\alpha j}_{x}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) \\ \Delta\sigma^{\alpha j} &:= \left[\sigma^{\alpha j}\left(t, \bar{x}^{\alpha}(t), u^{\varepsilon}(t)\right) - \sigma^{\alpha j}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) \right] \end{aligned} \tag{16.44}$$

(χ_{T_δ} is the characteristic function of the set T_δ),

$$\begin{aligned} dz^{\delta\alpha} &= \left[b^{\alpha}_{x} z^{\delta\alpha} + \frac{1}{2}\mathcal{B}^{\alpha}(t) + \Delta b^{\alpha} \chi_{T_\delta} \right] dt \\ &\quad + \sum_{j=1}^{m} \left[\sigma^{\alpha j}_{x} z^{\delta\alpha} + \frac{1}{2}\Xi^{\alpha j}(t) + \Delta\sigma^{\alpha j}_{x}(t)\, \chi_{T_\delta} \right] dW^{j} \\ z^{\delta\alpha}(0) &= 0 \end{aligned} \tag{16.45}$$

where

$$\begin{aligned} \mathcal{B}^{\alpha}(t) &:= \begin{pmatrix} \operatorname{tr}\left[b^{\alpha 1}_{xx}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) Y^{\delta\alpha}(t) \right] \\ \cdots \\ \operatorname{tr}\left[b^{\alpha n}_{xx}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) Y^{\delta\alpha}(t) \right] \end{pmatrix} \\ \Delta b^{\alpha} &:= b^{\alpha}\left(t, \bar{x}^{\alpha}(t), u^{\delta}(t)\right) - b^{\alpha}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) \\ \sigma^{\alpha j}_{x} &:= \sigma^{\alpha j}_{x}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) \end{aligned} \tag{16.46}$$

$$\begin{aligned} \Xi^{\alpha j}(t) &:= \begin{pmatrix} \operatorname{tr}\left[\sigma^{\alpha 1 j}_{xx}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) Y^{\delta\alpha}(t) \right] \\ \cdots \\ \operatorname{tr}\left[\sigma^{\alpha n j}_{xx}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) Y^{\delta\alpha}(t) \right] \end{pmatrix} (j = 1, \ldots, m) \\ \Delta\sigma^{\alpha j}_{x} &:= \sigma^{\alpha j}_{x}\left(t, \bar{x}^{\alpha}(t), u^{\delta}(t)\right) - \sigma^{\alpha j}_{x}\left(t, \bar{x}^{\alpha}(t), \bar{u}(t)\right) \\ Y^{\varepsilon\alpha}(t) &:= y^{\varepsilon\alpha}(t)\, y^{\varepsilon\alpha\mathsf{T}}(t) \end{aligned} \tag{16.47}$$

and

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathsf{E}\left\{\|x^{\alpha}(t)-\bar{x}^{\alpha}(t)\|^{2k}\right\}=O\left(\delta^{k}\right)\\
&\sup_{t\in[0,T]} \mathsf{E}\left\{\left\|y^{\delta\alpha}(t)\right\|^{2k}\right\}=O\left(\delta^{k}\right)\\
&\sup_{t\in[0,T]} \mathsf{E}\left\{\left\|z^{\delta\alpha}(t)\right\|^{2k}\right\}=O\left(\delta^{2k}\right)\\
&\sup_{t\in[0,T]} \mathsf{E}\left\|o_{\omega}^{\delta\alpha}(t)\right\|^{2k}=o\left(\delta^{2k}\right)
\end{aligned} \tag{16.48}$$

hold for any $\alpha\in\mathcal{A}$ and $k\geq 1$. The structures (16.43), (16.44), (16.45), (16.46) and (16.47) and the properties (16.48) are guaranteed by the assumptions **A1**–**A4**.

Taking into account these properties and the identity

$$h_x(x)=h_x(\bar{x})+\int_{\theta=0}^{1} h_{xx}(\bar{x}+\theta(x-\bar{x}))(x-\bar{x})\,d\theta \tag{16.49}$$

valid for any $L_\phi\left(C^2\right)$-mapping $h(x)$, and substituting (16.42) into (16.41), it follows that

$$\begin{aligned}
0\leq F_\delta\left(u^\delta(\cdot)\right)=\int_{\alpha\in\mathcal{A}}&\left[\mu_\alpha^{(\varepsilon)}\left\langle h_x^0\left(\bar{x}^\alpha(T)\right),y^{\delta\alpha}(T)+z^{\delta\alpha}(T)\right\rangle_E\right.\\
&+\left\langle c^{\alpha,(\varepsilon)},y^{\delta\alpha}(T)+z^{\delta\alpha}(T)\right\rangle_E+\nu_{\alpha j}^{(\varepsilon)}\left\langle h_x^j\left(\bar{x}^\alpha(T)\right),y^{\delta\alpha}(T)+z^{\delta\alpha}(T)\right\rangle_E\\
&+\mu_\alpha^{(\varepsilon)}\left\langle h_{xx}^0\left(\bar{x}^\alpha(T)\right)y^{\delta\alpha}(T),y^{\delta\alpha}(T)\right\rangle_E\\
&\left.+\nu_{\alpha j}^{(\varepsilon)}\left\langle h_{xx}^j\left(\bar{x}^\alpha(T)\right)y^{\delta\alpha}(T),y^{\delta\alpha}(T)\right\rangle_E\right]dm+o(\delta)
\end{aligned} \tag{16.50}$$

In view of the transversality conditions, the last expression (16.50) can be represented as follows:

$$0\leq F_\delta\left(u^\delta(\cdot)\right)=-\int_{\alpha\in\mathcal{A}} E\left\{\mathrm{tr}\left[\Psi^{\alpha,(\varepsilon)}(T)\,Y^{\delta\alpha}(t)\right]\right\}dm+o(\delta) \tag{16.51}$$

The following fact (see Lemma 4.6 in Yong and Zhou (1999) for the quadratic matrix case) is used.

Lemma 16.1. *Let*

$$Y(\cdot),\Psi_j(\cdot)\in L^2_{\mathcal{F}}\left(0,T;\mathbb{R}^{n\times r}\right),\ P(\cdot)\in L^2_{\mathcal{F}}\left(0,T;\mathbb{R}^{r\times n}\right)$$

satisfy

$$\begin{aligned}
dY(t)&=\Phi(t)\,Y(t)+\sum_{j=1}^{m}\Psi_j(t)\,dW^j\\
dP(t)&=\Theta(t)\,P(t)+\sum_{j=1}^{m}Q_j(t)\,dW^j
\end{aligned}$$

with

$$\Phi(\cdot) \in L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^{n\times n}\right), \Psi_j(\cdot) \in L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^{n\times r}\right)$$
$$Q_j(\cdot) \in L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^{r\times n}\right), \Theta(\cdot) \in L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^{r\times r}\right)$$

Then

$$\begin{aligned}&\mathsf{E}\{\operatorname{tr}[P(T)Y(T)] - \operatorname{tr}[P(0)Y(0)]\}\\ &\quad = \mathsf{E}\left\{\int_{t=0}^{T}\left(\operatorname{tr}[\Theta(t)Y(t)] + \operatorname{tr}[P(t)\Phi(t)] + \sum_{j=1}^{m} Q_j(t)\Psi_j(t)\right)dt\right\}\end{aligned} \tag{16.52}$$

The proof is based on the direct application of Ito's formula.

(a) *The evaluation of the term* $\mathsf{E}\left\{\psi^{\alpha,(\varepsilon)}(T)^{\intercal} y^{\delta\alpha}(T)\right\}$. Applying directly (16.52) and taking into account that $y^{\delta\alpha}(0) = 0$, it follows that

$$\begin{aligned}\mathsf{E}\left\{\psi^{\alpha,(\varepsilon)}(T)^{\intercal} y^{\delta\alpha}(T)\right\} &= \mathsf{E}\left\{\operatorname{tr}\left[y^{\delta\alpha}(T)\psi^{\alpha,(\varepsilon)}(T)^{\intercal}\right]\right\}\\ &= \mathsf{E}\left\{\int_{t=0}^{T}\operatorname{tr}\left[\sum_{j=1}^{m} q^{\alpha j,(\varepsilon)}(t)^{\intercal}\Delta\sigma^{\alpha j}\right]\chi_{T_\delta}dt\right\}\\ &= \mathsf{E}\left\{\int_{t=0}^{T}\operatorname{tr}\left[q^{\alpha,(\varepsilon)}(t)^{\intercal}\Delta\sigma^{\alpha}\right]\chi_{T_\delta}dt\right\}\end{aligned} \tag{16.53}$$

(b) *The evaluation of the term* $\mathsf{E}\left\{\psi^{\alpha,(\varepsilon)}(T)^{\intercal} z^{\delta\alpha}(T)\right\}$. In a similar way, applying directly (16.52) and taking into account that $z^{\delta\alpha}(0) = 0$, it follows that

$$\begin{aligned}&\mathsf{E}\left\{\psi^{\alpha,(\varepsilon)}(T)^{\intercal} z^{\delta\alpha}(T)\right\} = \mathsf{E}\left\{\operatorname{tr}\left[z^{\delta\alpha}(T)\psi^{\alpha,(\varepsilon)}(T)^{\intercal}\right]\right\}\\ &= \mathsf{E}\left\{\int_{t=0}^{T}\operatorname{tr}\left[\left(\frac{1}{2}\mathcal{B}^{\alpha}\psi^{\alpha,(\varepsilon)}(t)^{\intercal} + \frac{1}{2}\sum_{j=1}^{m} q^{\alpha j,(\varepsilon)\intercal}\Xi^{\alpha j}\right)\right.\right.\\ &\quad\left.\left. + \left(\Delta b^{\alpha}\psi^{\alpha,(\varepsilon)\intercal} + \sum_{j=1}^{m} q^{\alpha j,(\varepsilon)\intercal}\Delta\sigma_x^{\alpha j}(t)\, y^{\delta\alpha}\right)\chi_{T_\delta}\right]dt\right\}\end{aligned}$$

The equalities

$$\operatorname{tr}\left[\mathcal{B}^{\alpha}(t)\psi^{\alpha,(\varepsilon)}(T)^{\intercal} + \sum_{j=1}^{m} q^{\alpha j,(\varepsilon)}(t)^{\intercal}\Xi^{\alpha j}(t)\right] = \operatorname{tr}\left[H^{\alpha}_{xx}(t)Y^{\delta\alpha}(t)\right]$$
$$\mathsf{E}\left\{\int_{t=0}^{T}\operatorname{tr}\left[\sum_{j=1}^{m} q^{\alpha j,(\varepsilon)}(t)^{\intercal}\Delta\sigma_x^{\alpha j}(t)\, y^{\delta\alpha}(t)\right]\chi_{T_\delta}dt\right\} = o(\delta)$$

imply

$$\begin{aligned}
&\mathsf{E}\left\{\psi^{\alpha,(\varepsilon)}(T)^{\mathsf{T}} z^{\delta\alpha}(T)\right\} \\
&\quad = \mathsf{E}\left\{\int_{t=0}^{T} \operatorname{tr}\left[\frac{1}{2} H_{xx}^{\alpha}(t) Y^{\delta\alpha}(t) + \Delta b^{\alpha}(t)\,\psi^{\alpha,(\varepsilon)}(t)^{\mathsf{T}} \chi_{T_\delta}\right] dt\right\} + o(\delta)
\end{aligned} \tag{16.54}$$

(c) *The evaluation of the term* $\frac{1}{2}\mathsf{E}\left\{\operatorname{tr}\left[\Psi^{\alpha,(\varepsilon)}(T)\,Y^{\delta\alpha}(T)\right]\right\}$. Using (16.43) and applying the Itô formula to $Y^{\delta\alpha}(t) = y^{\delta\alpha}(t) y^{\delta\alpha}(t)^{\mathsf{T}}$, it follows that (for details see Yong and Zhou (1999))

$$\begin{aligned}
dY^{\delta\alpha}(t) &= \left[b_x^{\alpha} Y^{\delta\alpha} + Y^{\delta\alpha} b_x^{\alpha\mathsf{T}} + \sum_{j=1}^{m}\left(\sigma_x^{\alpha j} Y^{\delta\alpha} \sigma_x^{\alpha j\mathsf{T}} + B_{2j}^{\alpha} + B_{2j}^{\alpha\mathsf{T}}\right)\right] dt \\
&\quad + \sum_{j=1}^{m}\left(\sigma_x^{\alpha j} Y^{\delta\alpha} + Y^{\delta\alpha}\sigma_x^{\alpha j\mathsf{T}} + \left(\Delta\sigma^{\alpha j} y^{\delta\alpha\mathsf{T}} + y^{\delta\alpha}\Delta\sigma^{\alpha j\mathsf{T}}\right)\chi_{T_\delta}\right)\Bigg] dW^{j} \\
Y^{\delta\alpha}(0) &= 0
\end{aligned} \tag{16.55}$$

where

$$B_{2j}^{\alpha} := \left(\Delta\sigma^{\alpha j}\Delta\sigma^{\alpha j\mathsf{T}} + \sigma_x^{\alpha j} y^{\delta\alpha}\Delta\sigma^{\alpha j\mathsf{T}}\right)\chi_{T_\delta}$$

Again, applying directly (16.52) and taking into account that $Y^{\delta\alpha}(0) = 0$ and

$$\mathsf{E}\left\{\int_{t=0}^{T}\sum_{j=1}^{m} Q_j^{\alpha,(\varepsilon)}(t)\left(\Delta\sigma^{\alpha j} y^{\delta\alpha\mathsf{T}} + y^{\delta\alpha}\Delta\sigma^{\alpha j\mathsf{T}}\right)\chi_{T_\delta} dt\right\} = o(\delta)$$

it follows that

$$\begin{aligned}
&\mathsf{E}\left\{\operatorname{tr}\left[\Psi^{\alpha,(\varepsilon)}(T)\,Y^{\delta\alpha}(T)\right]\right\} \\
&\quad = \mathsf{E}\int_{t=0}^{T}\left(-\operatorname{tr}\left[H_{xx}^{\alpha} Y^{\delta\alpha}(t)\right] + \operatorname{tr}\left[\Delta\sigma^{\alpha\mathsf{T}}\Psi^{\alpha,(\varepsilon)}\Delta\sigma^{\alpha}\right]\chi_{T_\delta}\right) dt + o(\delta)
\end{aligned} \tag{16.56}$$

In view of definition (16.21)

$$\begin{aligned}
\delta\mathcal{H} &:= \mathcal{H}\left(t, \bar{x}^{\diamond}(t), u^{\delta}(t), \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right) \\
&\qquad - \mathcal{H}\left(t, \bar{x}^{\diamond}(t), \bar{u}(t), \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right) \\
&= \int_{\alpha\in\mathcal{A}}\left(\Delta b^{\alpha\mathsf{T}}\psi^{(\varepsilon)} + \operatorname{tr}\left[q^{\alpha,(\varepsilon)\mathsf{T}}\Delta\sigma^{\alpha}\right] + \frac{1}{2}\operatorname{tr}\left[\Delta\sigma^{\alpha\mathsf{T}}\Psi^{\alpha,(\varepsilon)}\Delta\sigma^{\alpha}\right]\right) dm
\end{aligned} \tag{16.57}$$

Using (16.53), (16.54), (16.56) and (16.57), it follows that

$$\mathsf{E}\left\{\int_{t=0}^{T}\delta\mathcal{H}(t)\,\chi_{T_{\delta_n}}dt\right\}=\mathsf{E}\left\{\int_{t=0}^{T}\int_{\alpha\in\mathcal{A}}\left(\Delta b^{\alpha\mathsf{T}}\psi^{(\varepsilon)}+\mathrm{tr}\left[q^{\alpha,(\varepsilon)\mathsf{T}}\Delta\sigma^{\alpha}\right]\right.\right.$$
$$\left.\left.+\frac{1}{2}\mathrm{tr}\left[\Delta\sigma^{\alpha\mathsf{T}}\Psi^{\alpha,(\varepsilon)}\Delta\sigma^{\alpha}\right]\right)dm\,\chi_{T_{\delta_n}}dt\right\}$$
$$=\left\langle\psi^{\diamond,(\varepsilon)}(T),y^{\delta\alpha}(T)+z^{\delta\alpha}(T)\right\rangle_E$$
$$+\frac{1}{2}\int_{\alpha\in\mathcal{A}}\mathsf{E}\left\{\mathrm{tr}\left[\Psi^{\alpha,(\varepsilon)}(T)Y^{\delta\alpha}(T)\right]\right\}dm+o(\delta_n)\quad(16.58)$$

Since

$$y^{\delta\alpha}(T)+z^{\delta\alpha}(T)=\delta\Delta^{\alpha}+o^{\delta\alpha}(T)$$

where $\Delta^{\alpha}\in K_1^{\alpha}$ is a displacement vector, and $\psi^{\alpha,(\varepsilon)}(T)=c^{\alpha,(\varepsilon)}\in K_{1\alpha}$, then

$$\left\langle\psi^{\diamond,(\varepsilon)}(T),y^{\delta\alpha}(T)+z^{\delta\alpha}(T)\right\rangle_E=\delta\left\langle c^{\alpha,(\varepsilon)},\Delta^{\alpha}\right\rangle_E+o(\delta)\leq 0\quad(16.59)$$

for sufficiently small $\delta>0$ and any fixed $\varepsilon>0$. In view of (16.51) and (16.59), the right-hand side of (16.58) can be estimated as

$$\mathsf{E}\left\{\int_{t=0}^{T}\delta\mathcal{H}(t)\,\chi_{T_{\delta_n}}dt\right\}=\delta\left\langle c^{\diamond,(\varepsilon)},\Delta^{\diamond}\right\rangle_E$$
$$+\frac{1}{2}\int_{\alpha\in\mathcal{A}}\mathsf{E}\left\{\mathrm{tr}\left[\Psi^{\alpha,(\varepsilon)}(T)\,Y^{\varepsilon\alpha}(T)\right]\right\}dm+o(\delta)\leq o(\delta_n)$$

Dividing by δ_n, it follows that

$$\delta_n^{-1}\mathsf{E}\left\{\int_{t=0}^{T}\delta\mathcal{H}(t)\,\chi_{T_\delta}dt\right\}\leq o(1)\quad(16.60)$$

Using Lemma 1 from Kushner (1972) for

$$T_\delta=[t_0-\delta_n\beta_1,t_0+\delta_n\beta_2]\,(\beta_1,\beta_2\geq 0;\,\beta_1+\beta_2>0)$$

and $\{\delta_n\}$ so that $\delta_n\to 0$, and in view of (16.60), it follows that

$$\delta_n^{-1}\mathsf{E}\left\{\int_{t=0}^{T}\delta\mathcal{H}(t)\,\chi_{T_{\delta_n}}dt\right\}\to(\beta_1+\beta_2)\,\mathsf{E}\{\delta\mathcal{H}(t_0)\}\leq 0\quad(16.61)$$

for almost all $t_0\in[0,T]$. Here if $t_0=0$ then $\beta_1=0$ and if $t_0=T$ then $\beta_2=0$, but if

$t_0 \in (0, T)$ then $\beta_1, \beta_2 > 0$. The inequality (16.61) implies

$$\mathsf{E}\{\delta\mathcal{H}(t)\} \leq 0 \tag{16.62}$$

from which (16.23) follows directly. Indeed, assume that there exist the control $\check{u}(t) \in \mathcal{U}^s_{feas}[0, T]$ and a time $t_0 \in (0, T)$ (not belonging to a set of null measure) such that

$$\mathsf{P}\{\omega \in \Omega_0(\rho)\} \geq p > 0 \tag{16.63}$$

where $\Omega_0(\rho) := \{\omega \in \Omega : \delta\mathcal{H}(t_0) > \rho > 0\}$. Then (16.62) can be rewritten as

$$\begin{aligned} 0 \geq \mathsf{E}\{\delta\mathcal{H}(t)\} &= \mathsf{E}\{\chi(\omega \in \Omega_0(\rho))\,\delta\mathcal{H}(t)\} \\ &\quad + \mathsf{E}\{\chi(\omega \notin \Omega_0(\rho))\,\delta\mathcal{H}(t)\} \\ &\geq \rho\mathsf{P}\{\omega \in \Omega_0(\rho)\} + \mathsf{E}\{\chi(\omega \notin \Omega_0(\rho))\,\delta\mathcal{H}(t)\} \\ &\geq \rho p + \mathsf{E}\{\chi(\omega \notin \Omega_0(\rho))\,\delta\mathcal{H}(t)\} \end{aligned}$$

Since this inequality should be also valid for the control $\hat{u}(t)$ satisfying

$$\hat{u}(t) = \begin{cases} \check{u}(t) & \text{for almost all } \omega \in \Omega_0(\rho) \\ \bar{u}(t) & \text{for almost all } \omega \neq \Omega_0(\rho) \end{cases}$$

there is the contradiction

$$0 \geq \mathsf{E}\{\delta\mathcal{H}(t)\} \geq \rho p + \mathsf{E}\{\chi(\omega \notin \Omega_0(\rho))\,\delta\mathcal{H}(t)\} = \rho p > 0$$

This completes the proof. □

16.5 Discussion

16.5.1 The important comment on Hamiltonian structure

The Hamiltonian function $\mathcal{H}$ used for the construction of the robust optimal control $\bar{u}(t)$ is equal to (see (16.20)) the Lebesgue integral over the uncertainty set of the standard stochastic Hamiltonians $\mathcal{H}^\alpha$ corresponding to each fixed value of the uncertain parameter.

16.5.2 RSMP for the control-independent diffusion term

From the Hamiltonian structure (16.21) it follows that if $\sigma^{\alpha j}(t, \bar{x}^\alpha(t), u(t))$ does not depend on $u(t)$, then

$$\begin{aligned} &\arg\max_{u \in U} \mathcal{H}\left(t, \bar{x}^\diamond(t), u, \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right) \\ &\quad = \arg\max_{u \in U} \int_{\mathcal{A}} \mathcal{H}^\alpha\left(t, \bar{x}^\alpha(t), u, \psi^{\alpha,(\varepsilon)}(t), \Psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) dm(\alpha) \\ &\quad = \arg\max_{u \in U} \int_{\mathcal{A}} H^\alpha\left(t, \bar{x}^\alpha(t), u, \psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) dm(\alpha) \end{aligned} \tag{16.64}$$

So, it follows that the *2nd order adjoint process* does not participate in the robust optimal constructions.

16.5.3 The standard stochastic maximum principle

If the stochastic plant is completely known, that is, there is no parametric uncertainty ($\mathcal{A} = \alpha_0, dm(\alpha) = \delta(\alpha - \alpha_0)\, d\alpha$), then from (16.64)

$$\begin{aligned}
&\arg\max_{u \in U} \mathcal{H}\left(t, \bar{x}^{\diamond}(t), u, \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right) \\
&\quad = \arg\max_{u \in U} \int_{\mathcal{A}} \mathcal{H}^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), \Psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) dm(\alpha) \\
&\quad = \arg\max_{u \in U} \mathcal{H}^{\alpha_0}\left(t, \bar{x}^{\alpha_0}(t), u, \psi^{\alpha_0,(\varepsilon)}(t), \Psi^{\alpha_0,(\varepsilon)}(t), q^{\alpha_0,(\varepsilon)}(t)\right)
\end{aligned} \tag{16.65}$$

and if $\varepsilon \to 0$, it follows that, in this case, RSMP converts to *stochastic maximum principle* (see Fleming and Rishel (1975), Zhou (1991) and Yong and Zhou (1999)).

16.5.4 Deterministic systems

In the deterministic case, when there are no stochastics

$$\left(\sigma^{\alpha}\left(t, \bar{x}^{\alpha}(t), u(t)\right) \equiv 0\right)$$

the robust maximum principle for min-max problems (in Mayer form) stated in Boltyanskii and Poznyak (1999) is obtained directly, that is, for $\varepsilon \to 0$ it follows that

$$\begin{aligned}
&\arg\max_{u \in U} \mathcal{H}\left(t, \bar{x}^{\diamond}(t), u, \psi^{\diamond,(\varepsilon)}(t), \Psi^{\diamond,(\varepsilon)}(t), q^{\diamond,(\varepsilon)}(t)\right) \\
&\quad = \arg\max_{u \in U} \int_{\mathcal{A}} b^{\alpha}(t, \bar{x}(t), u)^{\mathsf{T}} \psi^{\alpha}(t)\, dm(\alpha)
\end{aligned} \tag{16.66}$$

When, in addition, there are no parametric uncertainties ($\mathcal{A} = \alpha_0, dm(\alpha) = \delta(\alpha - \alpha_0) d\alpha$), the *classical maximum principle* for optimal control problems (in Mayer form), is obtained (Pontryagin et al., 1969), that is,

$$\begin{aligned}
&\arg\max_{u \in U} \mathcal{H}\left(t, \bar{x}^{\diamond}(t), u, \psi^{\diamond,(0)}(t), \Psi^{\diamond,(0)}(t), q^{\diamond,(0)}(t)\right) \\
&\quad = \arg\max_{u \in U} \mathcal{H}(t, \bar{x}(t), u, \psi(t), \Psi(t), q(t)) \\
&\quad = \arg\max_{u \in U} b(t, \bar{x}(t), u)^{\mathsf{T}} \psi(t)
\end{aligned} \tag{16.67}$$

16.5.5 Comment on possible non-fixed horizon extension

Consider the case when the function $h^0(x)$ is positive. Let us introduce a new variable x^{n+1} (associated with time t) with the equation

$$\dot{x}^{n+1} \equiv 1 \tag{16.68}$$

and consider the variable vector $\bar{x} = (x^1, \ldots, x^n, x^{n+1}) \in \mathbb{R}^{n+1}$. For the plant (16.1), combined with (16.68), the initial conditions are as follows:

$$x(t_0) = x_0 \in \mathbb{R}^n, \quad x^{n+1}(t_0) = 0 \text{ (for all } \alpha \in \mathcal{A})$$

Furthermore, we determine the terminal set $\mathcal{M}$ for the plant (16.1) and (16.68) by the inequality

$$\mathcal{M} := \left\{ x \in \mathbb{R}^{n+1} : h^{l+1}(x) = \tau - x^{n+1} \leq 0 \right\}$$

assuming that the numbers t_0, τ are fixed ($t_0 < \tau$). Let now $u(t)$, $\bar{x}(t)$, $0 \leq t \leq T$, be an admissible control that satisfies the terminal condition. Then $T \geq \tau$, since otherwise the terminal condition $x(t_1) \in \mathcal{M}$ would not be satisfied. The function $h^0(x)$ is defined only on $\mathbb{R}^n$, but we prolong it in to $\mathbb{R}^{n+1}$, setting

$$h^0(\bar{x}) = \begin{cases} h^0(x) & \text{for } x^{n+1} \leq \tau \\ h^0(x) + (x^{n+1} - \tau)^2 & \text{for } x^{n+1} > \tau \end{cases}$$

If now $T > \tau$, then (for every $\alpha \in \mathcal{A}$)

$$h^0(x(t_1)) = h^0(x(\tau)) + (t_1 - \tau)^2 > h^0(x(\tau))$$

Thus F^0 may attain its minimum only for $T = \tau$, that is, we have the problem with fixing time $T = \tau$. By this, we may make the following conclusion.

Conclusion 16.1. ***The theorem above gives the robust maximum principle only for the problem with a fixed horizon.*** *The non-fixed horizon case demands a special construction and implies another formulation of RMP.*

16.5.6 The case of absolutely continuous measures for uncertainty set

Consider now the case of an absolutely continuous measure $m(\mathcal{A}_0)$; that is, consider the situation when there exists a summable (the Lebesgue integral

$$\int_{R^s} p(x) \left(dx^1 \vee \cdots \vee dx^n \right)$$

is finite and s-fold) nonnegative function $p(x)$, given on $\mathbb{R}^s$ and named **the density of a measure** $m(\mathcal{A}_0)$, such that for every measurable subset $\mathcal{A}_0 \subset \mathbb{R}^s$ we have

$$m(\mathcal{A}_0) = \int\limits_{\mathcal{A}_0} p(x)\, dx, \quad dx := dx^1 \vee \cdots \vee dx^n$$

By this initial agreement, $\mathbb{R}^s$ is a space with the countable additive measure. Now it is possible to consider controlled object (16.1) with the set of uncertainty $\mathcal{A} = \mathbb{R}^s$. In this case

$$\int\limits_{\mathcal{A}_0} f(x)\, dm = \int\limits_{\mathcal{A}_0} f(x) p(x) dx \tag{16.69}$$

The statements of the robust maximum principle for this special case are obtained from the main theorem with evident variation. It is possible also to consider a particular case when $p(x)$ is defined only on a ball $\mathcal{A} \subset \mathbb{R}^s$ (or on another subset of $\mathbb{R}^s$) and integral (16.69) is defined only for $\mathcal{A}_0 \subset \mathcal{A}$.

16.5.7 Uniform density case

If no *a priori* information on some or other parameter values and the distance on a compact $\mathcal{A} \subset \mathbb{R}^s$ is defined in the natural way as $\|\alpha_1 - \alpha_2\|$, then the *maximum condition* (16.23) can be formulated (and proved) as follows:

$$
\begin{aligned}
u(t) \in\ & \operatorname*{Arg\,max}_{u\in U} H^{\diamond}(\psi(t), x(t), u) \\
&= \operatorname*{Arg\,max}_{u\in U} \int_{\mathcal{A}} \mathcal{H}^{\alpha}\left(t, \bar{x}^{\alpha}(t), u, \psi^{\alpha,(\varepsilon)}(t), \Psi^{\alpha,(\varepsilon)}(t), q^{\alpha,(\varepsilon)}(t)\right) d\alpha \\
&\quad \text{almost everywhere on } [t_0, t_1]
\end{aligned}
\tag{16.70}
$$

which represents, evidently, a partial case of the general condition (16.23) with a uniform absolutely continuous measure, that is, when

$$
dm(\alpha) = p(\alpha)d\alpha = \frac{1}{m(\mathcal{A})} d\alpha
$$

with $p(\alpha) = m^{-1}(\mathcal{A})$.

16.5.8 Can the complementary slackness inequalities be replaced by equalities?

It is naturally to ask: is it possible, in a general case, to replace the inequalities by the equalities as it was done above or not? Below we present an example that gives the negative answer. Consider the case of the absolutely continuous measure for $s = 1$ $\left(R^s = R^1\right)$ with the density $p(x) = e^{-x^2}$. Furthermore, take, for the simplicity, $n = 1$. Consider the family of the simple controlled plants given by

$$
\dot{x}^{\alpha,1} = f^{\alpha}(x, u) = -\frac{\alpha^2}{1+\alpha^2} + u
$$

with $t_0 = 0$, $t_1 = \frac{1}{2}$, $x^{\alpha,1}(0) = 1$, $\alpha \in [-1, 1]$, $U = [-1, 1]$ and no noise at all. The terminal set $\mathcal{M}$ is defined by the inequality $h^1(x) \le 0$ with $h^1(x) = x$. Finally, we take the cost function as $h^0(x) = 1 - x$. It is evident (applying the main theorem) that the optimal control is as follows: $u(t) \equiv -1$, $0 \le t \le \frac{1}{2}$ and $F^0 = 1$. But the complementary slackness condition in the form of (16.16) implies that $\mu_{\alpha}^{(0)} = \nu_{\alpha}^{(0)} = 0$ for all α and any $\varepsilon = 0$. Consequently the transversality condition gives $\psi^{(0)}(t) \equiv 0$. But this contradicts the nontriviality condition. Thus

Claim 16.1. *The inequalities in the main theorem **cannot be replaced** by equalities (16.16).*

16.6 Finite uncertainty set

If the uncertainty set $\mathcal{A}$ is *finite*, the robust maximum principle, proved above, gives the result contained in Poznyak et al. (2002a,b). In this case, the integrals may be replaced by finite sums and the number ε in formulation of the main theorem is superfluous and may be omitted, which is why in the complementary slackness condition we have the equalities.

Now, because of special importance, we will present the particular version of Theorem 16.1 when the parametric set $\mathcal{A}$ is finite.

16.6.1 Main result

Theorem 16.2. **(Poznyak et al., 2002b)** *Let the assumptions A1–A5 be fulfilled and*

$$\left(\bar{x}(\cdot), \bar{u}(\cdot)\right)$$

be the robust optimal dynamics. Then there exist collections of terminal conditions c^{α}, C^{α}, $\{\mathcal{F}_t\}_{t\geq 0}$-adapted stochastic matrices

$$\left(q^{\alpha}, Q_j^{\alpha}\,(j=1,\ldots,l)\right)$$

in (16.11) and (16.12), and nonnegative constants μ_α and $\nu_{\alpha j}$ $(j=1,\ldots,l)$ such that the following conditions are satisfied:

*1. (**Complementary slackness condition**): For any $\alpha \in \mathcal{A}$*

$$\boxed{\begin{array}{l}(i)\ \ \mu_\alpha\left[\mathsf{E}\left\{h^0\left(x^\alpha(T)\right)\right\} - \max\limits_{\alpha\in\mathcal{A}} E\left\{h^0\left(x^\alpha(T)\right)\right\}\right] = 0\\ (ii)\ \ \nu_{\alpha j}\mathsf{E}\left\{h^j\left(x^\alpha(T)\right)\right\} = 0\,(j=1,\ldots,l)\end{array}} \tag{16.71}$$

*2. (**Transversality condition**): For any $\alpha \in \mathcal{A}$ with probability one*

$$\boxed{\begin{array}{c}c^\alpha + \mu_\alpha h_x^0\left(x^\alpha(T)\right) + \sum\limits_{j=1}^{l}\nu_{\alpha j}h_x^j\left(x^\alpha(T)\right) = 0\\ C^\alpha + \mu_\alpha h_{xx}^0\left(x^\alpha(T)\right) + \sum\limits_{j=1}^{l}\nu_{\alpha j}h_{xx}^j\left(x^\alpha(T)\right) = 0\end{array}} \tag{16.72}$$

*3. (**Nontriviality condition**): There exists $\alpha \in \mathcal{A}$ such that $c^\alpha \neq 0$ or, at least, one of the numbers μ_α, $\nu_{\alpha j}$ $(j=1,\ldots,l)$ is distinct from 0, that is,*

$$\boxed{\exists\alpha\in\mathcal{A} : \left|c^\alpha\right| + \mu_\alpha + \sum_{j=1}^{l}\nu_{\alpha j} > 0} \tag{16.73}$$

*4. (**Maximality condition**): The robust optimal control $\bar{u}(\cdot)$ for almost all $t \in [0,T]$ maximizes the Hamiltonian function*

$$\boxed{\begin{aligned}&\mathcal{H}\left(t,\bar{x}^{\diamond}(t),u,\psi^{\diamond}(t),\Psi^{\diamond}(t),q^{\diamond}(t)\right)\\&\quad:=\sum_{\alpha\in\mathcal{A}}\mathcal{H}^{\alpha}\left(t,\bar{x}^{\alpha}(t),u,\psi^{\alpha}(t),\Psi^{\alpha}(t),q^{\alpha}(t)\right)\end{aligned}} \tag{16.74}$$

where

$$\boxed{\begin{aligned}&\mathcal{H}^{\alpha}\left(t,\bar{x}^{\alpha}(t),u,\psi^{\alpha}(t),\Psi^{\alpha}(t),q^{\alpha}(t)\right)\\&\quad:=H^{\alpha}\left(t,\bar{x}^{\alpha}(t),u,\psi^{\alpha}(t),q^{\alpha}(t)\right)-\frac{1}{2}\mathrm{tr}\left[\bar{\sigma}^{\alpha\intercal}\Psi^{\alpha}(t)\bar{\sigma}^{\alpha}\right]\\&\qquad+\frac{1}{2}\mathrm{tr}\left[\left(\sigma^{\alpha}\left(t,\bar{x}^{\alpha}(t),u\right)-\bar{\sigma}^{\alpha}\right)^{\intercal}\Psi^{\alpha}(t)\left(\sigma^{\alpha}\left(t,\bar{x}^{\alpha}(t),u\right)-\bar{\sigma}^{\alpha}\right)\right]\end{aligned}} \tag{16.75}$$

with

$$\begin{aligned}&\bar{\sigma}^{\alpha}:=\sigma^{\alpha}\left(t,\bar{x}^{\alpha}(t),\bar{u}(t)\right)\\&\bar{x}^{\diamond}(t):=\left(\bar{x}^{1\intercal}(t),\ldots,\bar{x}^{N\intercal}(t)\right)^{\intercal},\ \psi^{\diamond}(t):=\left(\psi^{1\intercal}(t),\ldots,\psi^{N\intercal}(t)\right)^{\intercal}\\&q^{\diamond}(t):=\left(q^{1}(t),\ldots,q^{N}(t)\right),\ \Psi^{\diamond}(t):=\left(\Psi^{1}(t),\ldots,\Psi^{N}(t)\right)\end{aligned}$$

i.e., for almost all $t\in[0,T]$

$$\boxed{\bar{u}(t)=\arg\max_{u\in U}\mathcal{H}\left(t,\bar{x}^{\diamond}(t),u,\psi^{\diamond}(t),\Psi^{\diamond}(t),q^{\diamond}(t)\right)} \tag{16.76}$$

16.6.2 Min-max production planning

Consider the stochastic process

$$\boxed{z(t)=z(0)+\int_{s=0}^{t}\xi^{\alpha}(s)\,ds+\int_{s=0}^{t}\sigma^{\alpha}(s)\,dW(s)} \tag{16.77}$$

which is treated (see Zhou (1991)) as *'the market demand process'* at time t where

- $\xi^{\alpha}(s)$ is the expected *demand rate* at the given environment conditions $\alpha\in\mathcal{A}$;
- the term $\int_{s=0}^{t}\sigma^{\alpha}(s)\,dW(s)$ represents the demand fluctuation due to environmental uncertainties.

The set $\mathcal{A}=\{\alpha_1,\alpha_2,\alpha_3\}$ may contain the elements α_i $(i=1,2,3)$ corresponding to

- '*very-stable market environment*' $(\alpha=\alpha_1)$;
- '*normal market environment*' $(\alpha=\alpha_2>\alpha_1)$;
- '*very-unstable market environment*' $(\alpha=\alpha_3>\alpha_2)$.

To meet the demand the factory serving this market should adjust its production rate all the time $t \in [0, T]$ (T is a planned working period) to accommodate any possible changes in the current market situation.

Let $y(t)$ be the inventory product level kept in the buffer of the capacity y^+. Then this system *'inventory-demands'*, in view of (16.77), can be written as

$$\boxed{\begin{array}{l} dy(t) = [u(t) - z(t)]\,dt,\ y(0) = y_0 \\ dz(t) = \xi^\alpha(t)\,dt + \sigma^\alpha(t)\,dW(t),\ z(0) = z_0 \end{array}} \tag{16.78}$$

where $u(t)$ is the control (or the production rate) at time t subject to the constraint

$$0 \le u(t) \le u^+ \tag{16.79}$$

The control processes $u(t)$, introduced in (16.78), should be nonanticipative, i.e., should be dependent on the past information only. All processes in (16.78) are assumed to be $\{\mathcal{F}_t\}_{t\ge 0}$-adapted $\mathbb{R}$-valued random processes. To avoid misleading, it is assumed that the maximal possible demands during the time $[0, T]$ can not exceed the *maximal production level*, i.e.,

$$\boxed{\begin{array}{rl} u^+T \ge & \max\limits_{\alpha\in\mathcal{A}} E\left\{\int\limits_{t=0}^{T} z(t)\,dt\right\} \\ = & \max\limits_{\alpha\in\mathcal{A}}\left[z_0T + E\left\{\int\limits_{t=0}^{T}\int\limits_{st=0}^{t}\xi^\alpha(s)\,ds\right\}\right] \end{array}} \tag{16.80}$$

Introduce the following cost function $h^0(y)$ defined by

$$\boxed{\begin{array}{c} h^0(y) = \dfrac{\lambda_1}{2}\left[y - y^+\right]_+^2 + \dfrac{\lambda_2}{2}[-y]_+^2 \\ [z]_+ := \begin{cases} z & \text{if } z > 0 \\ 0 & \text{if } z \le 0 \end{cases} \end{array}} \tag{16.81}$$

where the term $\left[y - y^+\right]_+^2$ corresponds to losses, related to *'extra production storage'*, the term $[-y]_+^2$ reflects *'losses due to a deficit'* and λ_1, λ_2 are two nonnegative weighting parameters.

This problem can be rewritten in standard form as follows:

$$\boxed{\begin{array}{c} \max\limits_{\alpha\in\mathcal{A}} \mathsf{E}\left\{h^0(x_1^\alpha(T))\right\} \to \min\limits_{u(\cdot)\in U_{ad}^s[0,T]} \\ U = \left\{u : 0 \le u \le u^+\right\} \end{array}} \tag{16.82}$$

and

$$\boxed{\begin{array}{l} d\begin{pmatrix} x_1^\alpha(t) \\ x_2^\alpha(t) \end{pmatrix} = \begin{pmatrix} u(t) - x_2^\alpha(t) \\ \xi^\alpha(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma^\alpha(t) \end{pmatrix} dW(t) \\ x_1^\alpha(0) = y_0, \quad x_2^\alpha(0) = z_0 \end{array}} \tag{16.83}$$

where for any fixed $\alpha \in \mathcal{A}$

$$x_1^\alpha(t) = y(t), \quad x_2^\alpha(t) = z(t)$$

In this problem formulation there are no terminal constraints.

In view of the technique suggested above, and taking into account that the diffusion term does not depend on control, it follows that

$$\begin{array}{l} d\begin{pmatrix} \psi_1^\alpha(t) \\ \psi_2^\alpha(t) \end{pmatrix} = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \psi_1^\alpha(t) \\ \psi_2^\alpha(t) \end{pmatrix} dt + \begin{pmatrix} q_1^\alpha(t) \\ q_2^\alpha(t) \end{pmatrix} dW(t) \\ \begin{pmatrix} \psi_1^\alpha(T) \\ \psi_2^\alpha(T) \end{pmatrix} = -\mu_\alpha \begin{pmatrix} \lambda_1 \left[x_1^\alpha(T) - y^+\right]_+ - \lambda_2\left[-x_1^\alpha(T)\right]_+ \\ 0 \end{pmatrix} \end{array}$$

From these equations the following equality is obtained:

$$\begin{array}{ll} q_1^\alpha(t) = 0, & \psi_1^\alpha(t) = \psi_1^\alpha(T) = -\mu_\alpha\left(\lambda_1\left[x_1^\alpha(T) - y^+\right]_+ - \lambda_2\left[-x_1^\alpha(T)\right]_+\right) \\ q_2^\alpha(t) = 0, & \psi_2^\alpha(t) = \psi_1^\alpha(T)\,[T - t] \end{array}$$

So, we have

$$\mathcal{H} = \sum_{\alpha\in\mathcal{A}} \left(\psi_1^\alpha(t)\left[u(t) - x_2^\alpha(t)\right] + \psi_1^\alpha(t)\,\xi^\alpha(t)\right)$$

and the maximality condition leads to

$$\begin{array}{l} \bar{u}(t) = \arg\max_{u\in U} \sum_{\alpha\in\mathcal{A}} \psi_1^\alpha(t)\,u(t) = u^+ \operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}} \psi_1^\alpha(t)\right] \\ \operatorname{sgn}[v] := \begin{Bmatrix} v & \text{if } v > 0 \\ 0 & \text{if } v \le 0 \end{Bmatrix} \end{array} \tag{16.84}$$

Under this control it follows that

$$\begin{array}{l} x_1^\alpha(T) = y_0 + Tu^+ \operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}} \psi_1^\alpha(T)\right] - Z_T^\alpha \\ Z_T^\alpha := \int\limits_{t=0}^{T} x_2^\alpha(t)\,dt = \int\limits_{t=0}^{T}\int\limits_{s=0}^{t} \xi^\alpha(s)ds + \int\limits_{t=0}^{T}\int\limits_{s=0}^{t} \sigma^\alpha(s)dW(s) \end{array} \tag{16.85}$$

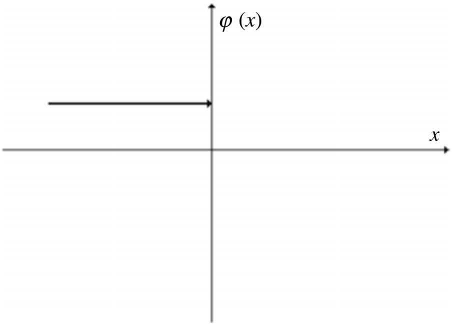

Fig. 16.1. $\varphi(x)$-function.

and for any $\alpha \in \mathcal{A}$

$$
\begin{aligned}
J^{\alpha} &:= \mathsf{E}\left\{h^{0}(x_{1}^{\alpha}(T))\right\} \\
&= \mathsf{E}\left\{h^{0}(y_{0} + Tu^{+}\operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}} \psi_{1}^{\alpha}(T)\right] - Z_{T}^{\alpha})\right\}
\end{aligned}
\tag{16.86}
$$

Since at least one active index exists, it follows that $\sum\limits_{\alpha\in\mathcal{A}} \mu_{\alpha} > 0$ and for any $\alpha \in \mathcal{A}$

$$
\begin{aligned}
x_{1}^{\alpha}(T) &= y_{0} + Tu^{+}\varphi(x) - Z_{T}^{\alpha} \\
\varphi(x) &:= \operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}} \mu_{\alpha}\left(\lambda_{2}\left[-x_{1}^{\alpha}(T)\right]_{+} - \lambda_{1}\left[x_{1}^{\alpha}(T) - y^{+}\right]_{+}\right)\right] \\
&= \operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}} \nu_{\alpha}\left(\lambda_{2}\left[-x_{1}^{\alpha}(T)\right]_{+} - \lambda_{1}\left[x_{1}^{\alpha}(T) - y^{+}\right]_{+}\right)\right] = \operatorname{sgn}[x] \\
x &:= \sum_{\alpha\in\mathcal{A}} \nu_{\alpha}\left(\lambda_{2}\left[-x_{1}^{\alpha}(T)\right]_{+} - \lambda_{1}\left[x_{1}^{\alpha}(T) - y^{+}\right]_{+}\right)
\end{aligned}
\tag{16.87}
$$

where $\nu_{\alpha} := \mu_{\alpha}/\sum\limits_{\alpha\in\mathcal{A}} \mu_{\alpha}$ is the component of the vector $\nu = (\nu_{1}, \ldots, \nu_{N})$ $(N = 3)$ satisfying

$$
\nu \in S_{N} := \left\{\nu = (\nu_{1}, \ldots, \nu_{N}) \mid \nu_{\alpha} \geq 0, \sum_{\alpha=1}^{N} \nu_{\alpha} = 1\right\}
$$

Multiplying both sides by μ_{α}, summing then over $\alpha \in \mathcal{A}$ and dividing by $\sum\limits_{\alpha\in\mathcal{A}} \mu_{\alpha}$, equation (16.87) can be transformed to (see Figs. 16.1–16.3)

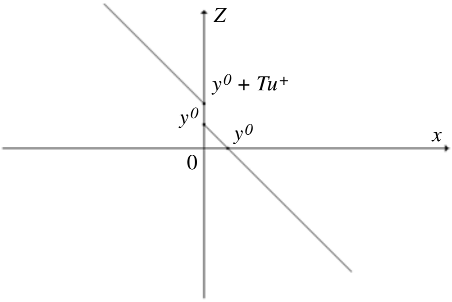

Fig. 16.2. $Z(x)$ mapping.

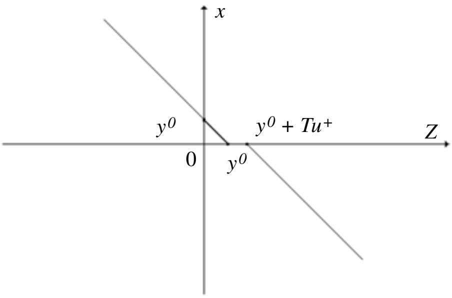

Fig. 16.3. $x(Z)$ mapping.

$$\begin{cases} Z = y_0 + Tu^+\varphi(x) - x \\ x = (y_0 - Z)\left[1 - \chi_{[y_0;y_0+Tu^+]}(x)\right] + Tu^+\varphi\left(Z - y_0 - Tu^+\right) \\ x := \sum_{\alpha\in\mathcal{A}} \nu_\alpha \left(\lambda_2\left[-x_1^\alpha(T)\right]_+ - \lambda_1\left[x_1^\alpha(T) - y^+\right]_+\right) \\ Z := \sum_{\alpha\in\mathcal{A}} \upsilon_\alpha Z_T^\alpha \end{cases} \tag{16.88}$$

where $\chi_{[y_0;y_0+Tu^+]}(x)$ is the characteristic function of the interval $[y_0;\ y_0 + Tu^+]$.

Below we present the procedure for calculating J^α from (16.86). First, let us find the density $p_x(v;\nu)$ of the distribution function of the random variable x related to $x = x(Z)$ from (16.88) with Z having density equal to $p_Z(v;\nu)$:

$$p_x(v;\nu) = \frac{d}{dv}\mathbf{P}\{x \le v\} = \frac{d}{dv}\int_{s=-\infty}^{\infty} \operatorname{sgn}[v - x(s)]\, p_Z(s;\nu)\, ds$$

$$= \int_{s=-\infty}^{\infty} \delta(v - x(s))\, p_Z(s;\nu)\, ds = \int_{s=-\infty}^{y_0-0} \delta(v - x(s))\, p_Z(s;\nu)\, ds$$

$$+ \int_{s=y_0}^{y_0+Tu^+} \delta(v - x(s))\, p_Z(s;\nu)\, ds$$

$$+ \int_{s=y_0+Tu^++0}^{\infty} \delta(v - x(s))\, p_Z(s;\nu)\, ds$$

$$= \int_{s=-\infty}^{y_0-0} \delta(v - [y_0 - s])\, p_Z(s;\nu)\, ds + \int_{s=y_0}^{y_0+Tu^+} \delta(v)\, p_Z(s;\nu)\, ds$$

$$+ \int_{s=y_0+Tu^++0}^{\infty} \delta\left(v - \left[y_0 + Tu^+ - s\right]\right) p_Z(s;\nu)\, ds$$

$$= \int_{s=-\infty}^{\infty} \chi(s < y_0)\, \delta(s - [y_0 - v])\, p_Z(s;\nu)\, ds + \delta(v) \int_{s=y_0}^{y_0+Tu^+} p_Z(s;\nu)\, ds$$

$$+ \int_{s=-\infty}^{\infty} \chi\left(s > y_0 + Tu^+\right) \delta\left(s - \left[y_0 + Tu^+ - v\right]\right) p_Z(s;\nu)\, ds$$

Hence

$$p_x(v;\nu) = \chi(v < 0)\, p_Z(y_0 - v;\nu) + \delta(v) \int_{s=y_0}^{y_0+Tu^+} p_Z(s;\nu)\, ds$$
$$+ \chi(v > 0)\, p_Z\left(y_0 + Tu^+ - v;\nu\right) \tag{16.89}$$

Note that, in view of (16.85) and (16.88), Z_T^α and Z have the following Gaussian distributions:

$$\begin{aligned} p_{Z_T^\alpha}(s) &= \mathcal{N}\left(\mathsf{E}\int_{t=0}^{T}\int_{s=0}^{t} \xi^\alpha(s) ds, \int_{t=0}^{T}\int_{s=0}^{t} E\left(\sigma^\alpha(s)\right)^2 ds\right) \\ p_Z(s;\nu) &= \mathcal{N}\left(\sum_{\alpha\in\mathcal{A}} \nu_\alpha \mathsf{E}\int_{t=0}^{T}\int_{s=0}^{t} \xi^\alpha(s) ds, \sum_{\alpha\in\mathcal{A}} \nu_\alpha^2 \int_{t=0}^{T}\int_{s=0}^{t} \mathsf{E}\left(\sigma^\alpha(s)\right)^2 ds\right) \end{aligned} \tag{16.90}$$

Then for each α we calculate

$$J^\alpha(\nu) := \mathsf{E}\left\{h^0(x_1^\alpha(T))\right\} = \int_{v=-\infty}^{\infty} h^0(v)\, p_{x_1^\alpha(T)}(v;\nu)\, dv \tag{16.91}$$

as a function of the vector ν. The integral in (16.91) can be calculated numerically for any $\nu \in S_N$ and the *worst case cost function* in this problem is

$$\min_{u(\cdot)\in U_{ad}^{s}[0,T]} \max_{\alpha\in\mathcal{A}} J^{\alpha} = \min_{\nu\in S_N} \max_{\alpha\in\mathcal{A}} J^{\alpha}(\nu)$$

The expression for the *robust optimal control* $\bar{u}(t)$ is evaluated from (16.84) as follows:

$$\bar{u}(t) = u^{+}\operatorname{sgn}\left[x^{*}\right] \tag{16.92}$$

where the random variable x^* has the distribution $p_x(v; \nu^*)$ given by (16.89) with

$$\nu^{*} := \arg\min_{\nu\in S_N} \max_{\alpha\in\mathcal{A}} J^{\alpha}(\nu) \tag{16.93}$$

Finally, $\bar{u}(t)$ is the binomial $\{\mathcal{F}_t\}_{t\geq 0}$-adapted random process since it depends on only the first two moments

$$\mathsf{E}\left\{\int_{t=0}^{T}\int_{s=0}^{t} \xi^{\alpha}(s)ds\right\}, \quad \int_{t=0}^{T}\int_{s=0}^{t} \mathsf{E}\left\{\left(\sigma^{\alpha}(s)\right)^{2}\right\} ds$$

of the entering random processes and it is given by

$$\boxed{\begin{aligned} &\bar{u}(t) = \bar{u}(0) = \begin{cases} u^{+} & \text{with the probability } P^{*} \\ 0 & \text{with the probability } 1 - P^{*} \end{cases} \\ &P^{*} = \int_{v=-\infty}^{0-} p_x\left(v; \nu^{*}\right) dv \end{aligned}} \tag{16.94}$$

The derived robust optimal control (16.94) is unique if the optimization problem (16.93) has a unique solution.

16.6.3 Min-max reinsurance–dividend management

Consider the following $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}$-valued random processes

$$\boxed{\begin{aligned} &dy(t) = \left[a(t)\tilde{\mu}^{\alpha} - \delta^{\alpha} - c(t)\right] dt - a(t)\sigma^{\alpha} dW(t) \\ &y(0) = y_0 \end{aligned}} \tag{16.95}$$

where, according to Taksar and Zhou (1998),

- $y(t)$ is the value of the liquid assets of a company at time t;
- $c(t)$ is the dividend rate paid out to the shareholders at time t;
- $\tilde{\mu}^{\alpha}$ is the difference between premium rate and expected payment on claims per unit time (‘safety loading’);

- δ^{α} is the rate of the debt repayment;
- $[1 - a(t)]$ is the reinsurance fraction;
- $\sigma^{\alpha} := \sqrt{\lambda_{\alpha} E\left\{\eta^2\right\}}$ (λ_{α} is the intensity of Poisson process, η is the size of claim).

The controls are

$$u_1(t) := a(t) \in [0, 1] \quad \text{and} \quad u_2(t) := c(t) \in \left[0, c^+\right]$$

The finite parametric set $\mathcal{A}$ describes the possible different environmental situations. The payoff-cost function is as follows:

$$\boxed{J = \min_{\alpha \in \mathcal{A}} E\left\{\int_{t=0}^{T} e^{-\gamma t} c(t) dt\right\} \to \max_{a(\cdot), c(\cdot)}, \quad \gamma \in \left[0, c^+/k\right]} \tag{16.96}$$

At time T it is natural to satisfy

$$\boxed{k\mathsf{E}\left\{y(T)\right\} \geq \mathsf{E}\left\{\int_{t=0}^{T} e^{-\gamma t} c(t) dt\right\} \geq k_0, \quad k_0 > 0} \tag{16.97}$$

This problem can be rewritten in the standard form in the following way: for any fixed parameter $\alpha \in \mathcal{A}$ define

$$x_1^{\alpha}(t) := y(t), \quad x_2^{\alpha}(t) := \int_{s=0}^{t} e^{-\gamma s} u_2(s) ds$$

and then the problem is as follows:

$$\begin{aligned}
&\max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{h^0\left(x^{\alpha}(T)\right)\right\} \to \min_{u(\cdot) \in U_{ad}^{s}[0,T]} \\
&\mathsf{E}\left\{h^1\left(x^{\alpha}(T)\right)\right\} \geq 0 \\
&h^0(x) = -x_2, \ h^1(x) = kx_1 - x_2, h^2(x) = x_2 - k_0 \\
&U = \left\{u \in \mathbb{R}^2 : 0 \leq u_1 \leq 1, 0 \leq u_2 \leq c^+\right\}
\end{aligned}$$

with the dynamic models given by

$$\begin{aligned}
&d\begin{pmatrix} x_1^{\alpha}(t) \\ x_2^{\alpha}(t) \end{pmatrix} = \begin{pmatrix} u_1(t)\tilde{\mu}^{\alpha} - \delta^{\alpha} - u_2(t) \\ e^{-\gamma t} u_2(t) \end{pmatrix} dt + \begin{pmatrix} -u_1(t)\sigma^{\alpha} \\ 0 \end{pmatrix} dW(t) \\
&x_1^{\alpha}(0) = y_0, \quad x_2^{\alpha}(0) = 0
\end{aligned}$$

Following the technique suggested above, we obtain

$$\begin{aligned}
&d\begin{pmatrix} \psi_1^{\alpha}(t) \\ \psi_2^{\alpha}(t) \end{pmatrix} = \begin{pmatrix} q_1^{\alpha}(t) \\ q_2^{\alpha}(t) \end{pmatrix} dW(t) \\
&\begin{pmatrix} \psi_1^{\alpha}(T) \\ \psi_2^{\alpha}(T) \end{pmatrix} = \begin{pmatrix} -k\nu_{\alpha 1} \\ \mu_{\alpha} + \nu_{\alpha 1} - \nu_{\alpha 2} \end{pmatrix}
\end{aligned}$$

$$d\Psi^{\alpha}(t) = Q^{\alpha}(t)\, dW(t)$$
$$\Psi^{\alpha}(T) = 0$$

and, hence,

$$q_1^{\alpha}(t) = q_2^{\alpha}(t) = Q^{\alpha}(t)$$
$$\psi_1^{\alpha}(t) = \psi_1^{\alpha}(T) = -k\nu_{\alpha}$$
$$\psi_2^{\alpha}(t) = \psi_2^{\alpha}(T) = \mu_{\alpha} + \nu_{12} - \nu_{\alpha 2}$$

Then

$$\begin{aligned}\mathcal{H} &= \sum_{\alpha\in\mathcal{A}} \left(\psi_1^{\alpha}(t)\left[u_1(t)\tilde{\mu}^{\alpha} - \delta^{\alpha} - u_2(t)\right] + \psi_2^{\alpha}(t)\, e^{-\gamma t} u_2(t)\right) \\ &= \sum_{\alpha\in\mathcal{A}} \left(-k\nu_{\alpha 1}\left[u_1(t)\tilde{\mu}^{\alpha} - \delta^{\alpha} - u_2(t)\right] + \left[\mu_{\alpha} + \nu_{\alpha 1} - \nu_{\alpha 2}\right] e^{-\gamma t} u_2(t)\right)\end{aligned}$$

and the robust optimal control, maximizing this Hamiltonian, is

$$\bar{u}(t) = \begin{pmatrix} \operatorname{sgn}\left[-\sum_{\alpha\in\mathcal{A}} \nu_{\alpha 1}\tilde{\mu}^{\alpha}\right] \\ c^{+}\operatorname{sgn}\left[\phi(t)\right] \end{pmatrix}$$

$$\phi(t) = \sum_{\alpha\in\mathcal{A}} \left(e^{-\gamma t}\left[\mu_{\alpha} + \nu_{\alpha 1} - \nu_{\alpha 2}\right] + k\nu_{\alpha 1}\right)$$

There are two cases to be considered:

- the first one, corresponding to the switching of $\bar{u}_2(t)$ from 0 to c^{+} and;
- the second – the switching of $\bar{u}_2(t)$ from c^{+} to 0.

1. The switching of $\bar{u}_2(t)$ from 0 to c^{+}. With this control, the expectation of the state is

$$\begin{aligned}\mathsf{E}\left\{\begin{pmatrix} x_1^{\alpha}(T) \\ x_2^{\alpha}(T)\end{pmatrix}\right\} &= \begin{pmatrix} y_0 \\ 0\end{pmatrix} + \mathsf{E}\left\{\int_{t=0}^{T} \begin{pmatrix} u_1(t)\tilde{\mu}^{\alpha} - \delta^{\alpha} - u_2(t) \\ e^{-\gamma t} u_2(t)\end{pmatrix} dt\right\} \\ &= \begin{pmatrix} y_0 \\ 0\end{pmatrix} + \begin{pmatrix} \left(\operatorname{sgn}\left[-\sum_{\alpha\in\mathcal{A}} \nu_{\alpha 1}\tilde{\mu}^{\alpha}\right]\tilde{\mu}^{\alpha} - \delta^{\alpha}\right) T - c^{+}\tau \\ c^{+}\gamma^{-1}\left[1 - e^{-\gamma\tau}\right]\end{pmatrix}\end{aligned}$$

where

$$\tau := \inf\{t \in [0, T] : \phi(t) = 0\}$$

The robust optimal control corresponds to the selection of the minimizing parameters $\mu_{\alpha}, \nu_{\alpha 1}, \nu_{\alpha 2}$:

$$\begin{aligned}&\arg \min_{\{\mu_{\alpha},\nu_{\alpha 1},\nu_{\alpha 2}\}} \max_{\alpha\in\mathcal{A}} \mathsf{E}\left\{-x_2^{\alpha}(T)\right\} \\ &\quad = \arg \min_{\{\mu_{\alpha},\nu_{\alpha 1},\nu_{\alpha 2}\}} \max_{\alpha\in\mathcal{A}} \left(c^{+}\gamma^{-1}\left[e^{-\gamma\tau} - 1\right]\right) \qquad (16.98)\end{aligned}$$

According to the existing constraints for any $\alpha \in \mathcal{A}$ there is the inequality

$$k\mathsf{E}\left\{x_1^{\alpha}(T)\right\} \geq \mathsf{E}\left\{x_2^{\alpha}(T)\right\} \geq k_0$$

or, in another form, since $\mathsf{E}\left\{x_2^{\alpha}(T)\right\}$ does not depend on $\alpha \in \mathcal{A}$,

$$k\left[y_0 + T\rho - c^+\tau\right] \geq c^+\gamma^{-1}\left[1 - e^{-\gamma\tau}\right] \geq k_0$$

where

$$\rho := \min_{\alpha \in \mathcal{A}} \left(\operatorname{sgn}\left[-\sum_{\alpha \in \mathcal{A}} \nu_{\alpha 1} \tilde{\mu}^{\alpha} \right] \tilde{\mu}^{\alpha} - \delta^{\alpha} \right)$$

From these two constraints it follows that

$$\tau_1 \leq \tau \leq \tau_2$$

where

$$\tau_1 = -\gamma^{-1} \ln\left(1 - k_0\gamma/c^+\right)$$

and τ_2 is the solution of

$$k\left(y_0 + \rho T - c^+\tau\right) = c^+\gamma^{-1}\left[1 - e^{-\gamma\tau}\right]$$

The goal now is

$$\begin{aligned}
&\min_{\{\mu_\alpha, \nu_{\alpha 1}, \nu_{\alpha 2}\}} \min_{\tau_1 \leq \tau \leq \tau_2} \max_{\alpha \in \mathcal{A}} E\left\{-x_2^{\alpha}(T)\right\} \\
&\quad = \min_{\{\mu_\alpha, \nu_{\alpha 1}, \nu_{\alpha 2}\}} \min_{\tau_1 \leq \tau \leq \tau_2} \left(-c^+\gamma^{-1}\left[1 - e^{-\gamma\tau}\right]\right) \\
&\quad = c^+\gamma^{-1} \min_{\{\mu_\alpha, \nu_{\alpha 1}, \nu_{\alpha 2}\}} \min_{\tau_1 \leq \tau \leq \tau_2} \left(e^{-\gamma\tau} - 1\right)
\end{aligned}$$

It may be done by the variation of the unknown nonnegative parameters μ_α, $\nu_{\alpha 1}$ and $\nu_{\alpha 2}$ involved in the switching function $\phi(t)$. The optimal parameter selection should satisfy the equality

$$\tau_1 \leq \tau\left(\mu_\alpha^*, \nu_{\alpha 1}^*, \nu_{\alpha 2}^*\right) = \max_{\{\mu_\alpha, \nu_{\alpha 1}, \nu_{\alpha 2}\}} \tau_2$$

Finally, *the robust optimal control* is equal to

$$\boxed{\bar{u}(t) = \begin{pmatrix} \operatorname{sgn}\left[\sum_{\alpha \in \mathcal{A}} \nu_{\alpha 1}^* \left[-\tilde{\mu}^{\alpha}\right]_+\right] \\ c^+ \operatorname{sgn}\left[t - \tau\left(\mu_\alpha^*, \nu_{\alpha 1}^*, \nu_{\alpha 2}^*\right)\right] \end{pmatrix}} \tag{16.99}$$

and the corresponding *worst case cost function* (or the best payoff) is

$$\boxed{\begin{aligned} J_{0\to 1}^{*} &:= \min_{u(\cdot)\in U_{ad}^{s}[0,T]} \max_{\alpha\in\mathcal{A}} \mathsf{E}\left\{h^{0}\left(x^{\alpha}\left(T\right)\right)\right\} \\ &= c^{+}\gamma^{-1}\left[e^{-\gamma\tau\left(\mu_{\alpha}^{*},\nu_{\alpha1}^{*},\nu_{\alpha2}^{*}\right)}-1\right] \end{aligned}} \tag{16.100}$$

2. The switching of $\bar{u}_2(t)$ *from* c^{+} *to* 0. Analogously, the expectation of the state is

$$\begin{aligned} \mathsf{E}\left\{\begin{pmatrix} x_1^{\alpha}(T) \\ x_2^{\alpha}(T) \end{pmatrix}\right\} &= \begin{pmatrix} y_0 \\ 0 \end{pmatrix} + \mathsf{E}\left\{\int_{t=0}^{T}\begin{pmatrix} u_1(t)\tilde{\mu}^{\alpha}-\delta^{\alpha}-u_2(t) \\ e^{-\gamma t}u_2(t) \end{pmatrix}dt\right\} \\ &= \begin{pmatrix} y_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \left(\operatorname{sgn}\left[-\sum_{\alpha\in\mathcal{A}}\nu_{\alpha1}\tilde{\mu}^{\alpha}\right]\tilde{\mu}^{\alpha}-\delta^{\alpha}\right)T-c^{+}\left[T-\tau\right] \\ c^{+}\gamma^{-1}\left[e^{-\gamma\tau}-e^{-\gamma T}\right] \end{pmatrix} \end{aligned}$$

By the same reasons, the minimizing parameters $\tilde{\mu}_{\alpha}, \nu_{\alpha1}, \nu_{\alpha2}$ satisfy (16.98) and the constraints can be rewritten as follows:

$$k\left(y_0+\rho T-c^{+}\left[T-\tau\right]\right) \geq c^{+}\gamma^{-1}\left[e^{-\gamma\tau}-e^{-\gamma T}\right] \geq k_0$$

From these two constraints it follows that

$$\tau_4 \leq \tau \leq \tau_3$$

where

$$\tau_3 = -\gamma^{-1}\ln\left(k_0\gamma/c^{+}+e^{-\gamma T}\right)$$

and τ_4 is the solution of

$$k\left(y_0+\rho T-c^{+}\left[T-\tau\right]\right) = c^{+}\gamma^{-1}\left[e^{-\gamma\tau}-e^{-\gamma T}\right]$$

(if there is no solution then $\tau_4 := T$). Our goal is

$$\begin{aligned} &\min_{\{\mu_{\alpha},\nu_{\alpha1},\nu_{\alpha2}\}} \min_{\tau_4\leq\tau\leq\tau_3} \max_{\alpha\in\mathcal{A}} \mathsf{E}\left\{-x_2^{\alpha}(T)\right\} \\ &\quad = \min_{\{\mu_{\alpha},\nu_{\alpha1},\nu_{\alpha2}\}} \min_{\tau_4\leq\tau\leq\tau_3} c^{+}\gamma^{-1}\left[e^{-\gamma T}-e^{-\gamma\tau}\right] \end{aligned}$$

The *robust optimal control* is equal to

$$\boxed{\bar{u}(t) = \begin{pmatrix} \operatorname{sgn}\left[\sum_{\alpha\in\mathcal{A}}\nu_{\alpha1}^{*}\left[-\tilde{\mu}^{\alpha}\right]_{+}\right] \\ c^{+}\operatorname{sgn}\left[\tau\left(\mu_{\alpha}^{*},\nu_{\alpha1}^{*},\nu_{\alpha2}^{*}\right)-t\right] \end{pmatrix}} \tag{16.101}$$

where the optimal parameter selection $\left(\mu_{\alpha}^{*},\nu_{\alpha1}^{*},\nu_{\alpha2}^{*}\right)$ should satisfy the inequality

$$\tau\left(\mu_{\alpha}^{*},\nu_{\alpha1}^{*},\nu_{\alpha2}^{*}\right) = \min_{\{\mu_{\alpha},\nu_{\alpha1},\nu_{\alpha2}\}}\tau_4 \leq \tau_3$$

and the corresponding *worst case cost function* (or the best payoff) is

$$\boxed{\begin{aligned} J^*_{1\to 0} &:= \min_{u(\cdot)\in U^s_{ad}[0,T]} \max_{\alpha\in\mathcal{A}} \mathsf{E}\left\{h^0\left(x^\alpha\left(T\right)\right)\right\} \\ &= c^+\gamma^{-1}\left[e^{-\gamma T} - e^{-\gamma\tau\left(\mu^*_\alpha,\nu^*_{\alpha 1},\nu^*_{\alpha 2}\right)}\right] \end{aligned}} \tag{16.102}$$

At the last step compare $J^*_{0\to 1}$ with $J^*_{1\to 0}$ and select the case with the minimal worst cost function, i.e.,

$$\boxed{J^* = J^*_{0\to 1} \wedge J^*_{1\to 0}} \tag{16.103}$$

with the corresponding switching rule ($0 \to 1$ or $1 \to 0$) and the robust optimal control $\bar{u}\left(t\right)$ given by (16.99) or (16.101).

16.7 Min-Max LQ-control

In this section we will present robust stochastic control designing in detail for the class of linear quadratic min-max stochastic problems.

16.7.1 Stochastic uncertain linear system

Let $\left(\Omega, \mathcal{F}, \left\{\mathcal{F}_t\right\}_{t\geq 0}, \mathsf{P}\right)$ be a given filtered probability space where an m-dimensional standard Brownian motion

$$\left(W\left(t\right) = \left(W^1\left(t\right), \ldots, W^m\left(t\right)\right), t \geq 0\right)$$

(with $W\left(0\right) = 0$) is defined. $\left\{\mathcal{F}_t\right\}_{t\geq 0}$ is assumed to be the natural filtration generated by $\left(W\left(t\right), t \geq 0\right)$ and augmented by the P-null sets from $\mathcal{F}$. Consider the stochastic linear controlled continuous-time system with the dynamics $x\left(t\right) \in \mathbb{R}^n$ given by

$$\boxed{\begin{aligned} &dx\left(t\right) = \left[A^\alpha\left(t\right)x\left(t\right) + B^\alpha\left(t\right)u\left(t\right) + b^\alpha\left(t\right)\right]dt \\ &\qquad\qquad + \sum_{i=1}^m \left[C^\alpha_i\left(t\right)x\left(t\right) + D^\alpha_i\left(t\right)u\left(t\right) + \sigma^\alpha_i\left(t\right)\right]dW^i\left(t\right) \\ &x\left(0\right) = x_0, \quad t \in [0, T] \ \left(T > 0\right) \end{aligned}} \tag{16.104}$$

In the above, $u\left(t\right) \in \mathbb{R}^k$ is a stochastic control at time t, and

$$\begin{aligned} &A^\alpha, C^\alpha_j : [0, T] \to \mathbb{R}^{n\times n} \\ &B^\alpha, D^\alpha_j\left(t\right) : [0, T] \to \mathbb{R} \\ &b^\alpha, \sigma^\alpha_j : [0, T] \to \mathbb{R}^n \end{aligned}$$

are the known deterministic Borel *measurable* functions of suitable sizes.

α is a parameter taking values from the finite set $\mathcal{A} = \{\alpha_1, \ldots, \alpha_N\}$. The initial state x_0 is assumed to be a square-integrable random vector with the *a priori* known mean m_0 and covariance matrix X_0.

The only sources of uncertainty in this system description are

- the system random noise $W(t)$;
- the unknown parameter $\alpha \in \mathcal{A}$.

It is assumed that *the past information is available* for the controller.

To emphasize the dependence of the random trajectories on the parameter $\alpha \in \mathcal{A}$ equation (16.2) is rewritten as

$$\boxed{\begin{aligned} dx^{\alpha}(t) &= \left[A^{\alpha}(t)x^{\alpha}(t) + B^{\alpha}(t)u(t) + b^{\alpha}(t)\right]dt \\ &\quad + \sum_{i=1}^{m}\left[C_i^{\alpha}(t)x^{\alpha}(t) + D_i^{\alpha}(t)u(t) + \sigma_i^{\alpha}(t)\right]dW^i(t) \\ x(0) &= x_0, \quad t \in [0, T]\ (T > 0) \end{aligned}} \tag{16.105}$$

16.7.2 Feasible and admissible control

The following definitions will be used throughout.

A stochastic control $u(\cdot)$ is called **feasible** in the stochastic sense (or, s-feasible) for the system (16.105) if

1. $u(\cdot) \in U[0, T]$
   $:= \left\{u : [0, T] \times \Omega \to \mathbb{R}^k \mid u(\cdot) \text{ is } \{\mathcal{F}_t\}_{t \geq 0}\text{-adapted}\right\}$
2. $x^{\alpha}(t)$ is the unique solution of (16.105) in the sense that for any $x^{\alpha}(t)$ and $\hat{x}^{\alpha}(t)$, satisfying (16.105),

   $$\mathsf{P}\left\{\omega \in \Omega : x^{\alpha}(t) = \hat{x}^{\alpha}(t)\right\} = 1$$

The pair $(x^{\alpha}(t); u(\cdot))$, where $x^{\alpha}(t)$ is the solution of (16.105) corresponding to this $u(\cdot)$, is called an s**-feasible pair**. The measurability of all deterministic functions in (16.105) guarantees that any $u(\cdot) \in U[0, T]$ is s-feasible. Since any additional constraints are absent it follows that any s-feasible control is **admissible** (or, s-admissible). The set of all s-admissible controls is denoted by $U^s_{adm}[0, T]$.

16.7.3 Min-max stochastic control problem setting

For any s-admissible control $u(\cdot) \in U^s_{adm}[0, T]$ and for any $\alpha \in \mathcal{A}$ define *the α-cost function*

$$\boxed{\begin{aligned} J^{\alpha}(u(\cdot)) := \mathsf{E}&\left\{\frac{1}{2}x^{\alpha}(T)^{\mathsf{T}} G x^{\alpha}(T)\right\} \\ &+ \mathsf{E}\int_{t=0}^{T}\left[\frac{1}{2}x^{\alpha}(t)^{\mathsf{T}}\bar{Q}(t)x^{\alpha}(t) + u(t)^{\mathsf{T}}S(t)x^{\alpha}(t)\right. \\ &\left. + \frac{1}{2}u(t)^{\mathsf{T}}R(t)u(t)\,dt\right] \end{aligned}} \tag{16.106}$$

where for all $t \in [0, T]$

$$\bar{Q}(t) = \bar{Q}^{\intercal}(t) \geq 0, S(t), R(t) = R^{\intercal}(t) > 0$$

are the known Borel measurable $\mathbb{R}^{n\times n}$, $\mathbb{R}^{n\times k}$, $\mathbb{R}^{k\times k}$ valued deterministic matrices, respectively, and G is the given $\mathbb{R}^{n\times n}$ deterministic matrix.

Since the value of the parameter α is unknown, define the *worst (highest) cost* as follows:

$$J(u(\cdot)) = \max_{\alpha\in\mathcal{A}} J^{\alpha}(u(\cdot)) \tag{16.107}$$

The stochastic control $\bar{u}(\cdot)$ is **robust optimal** (in min-max sense) if
1. it is *admissible*, that is,

$$\bar{u}(\cdot) \in U^{s}_{adm}[0, T]$$

and
2. it provides *the minimal worst* cost, that is,

$$\bar{u}(\cdot) = \arg \min_{u(\cdot)\in U^{s}_{adm}[0,T]} \max_{\alpha\in\mathcal{A}} J^{\alpha}(u(\cdot))$$

If the solution $\bar{x}(t)$ corresponds to this robust optimal control $\bar{u}(t)$ then $\left(\bar{x}(\cdot), \bar{u}(\cdot)\right)$ is called the *robust optimal dynamics*.

Thus the robust (with respect to the unknown parameter) optimal stochastic control problem (in Bolza form) consists in finding the robust optimal control $\bar{u}(t)$ according to the definition given above, that is,

$$\boxed{J\left(\bar{u}(\cdot)\right) = \min_{u(\cdot)\in U^{s}_{adm}[0,T]} \max_{\alpha\in\mathcal{A}} J^{\alpha}(u(\cdot))} \tag{16.108}$$

16.7.4 The problem presentation in Mayer form

To apply directly Theorem 16.2 let us present this problem in the so-called *Mayer form* introducing the new variable $x^{\alpha}_{n+1}(t)$ as follows:

$$x^{\alpha}_{n+1}(t) := \frac{1}{2}\int_{s=0}^{t} \left[x^{\alpha}(s)^{\intercal}\bar{Q}(s)x^{\alpha}(s) + u(s)^{\intercal}S(s)x^{\alpha}(s) + u(s)^{\intercal}R(s)u(s)\,ds\right.$$

which satisfies

$$\begin{aligned}
dx^{\alpha}_{n+1}(t) &= b_{n+1}(t, x^{\alpha}(t), u(t)) \\
&:= +x^{\alpha}(t)^{\intercal}\bar{Q}(t)x^{\alpha}(t)/2 + u(t)^{\intercal}S(t)x(t) \\
&\quad + u(t)^{\intercal}R(t)u(t)/2 + \sigma^{\intercal}_{n+1}(t)\,dW(t)
\end{aligned}$$

$$x^{n+1}(0) = 0,\ \sigma^{\intercal}_{n+1}(t) \equiv 0$$

16.7.5 Adjoint equations

According to (16.11) and (16.12), we have

- the *1st order vector adjoint equations*:

$$
\begin{aligned}
d\psi^{\alpha}(t) &= -\left[A^{\alpha}(t)^{\mathsf{T}}\psi^{\alpha}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\mathsf{T}} q_i^{\alpha}(t)\right. \\
&\quad \left. + \left(\bar{Q}(t)x^{\alpha}(t) + S(t)^{\mathsf{T}}u(t)\right)\psi_{n+1}^{\alpha}(t)\right]dt \\
&\quad + \sum_{i=1}^{m} q_i^{\alpha}(t)dW^i(t),\ t \in [0,T] \\
\psi^{\alpha}(T) &= c^{\alpha} \\
d\psi_{n+1}^{\alpha}(t) &= q_{n+1}^{\alpha}(t)^{\mathsf{T}}dW(t),\ t \in [0,T] \\
\psi_{n+1}^{\alpha}(T) &= c_{n+1}^{\alpha}
\end{aligned}
\tag{16.109}
$$

and

- the *2nd order matrix adjoint equations*:

$$
\begin{aligned}
d\Psi^{\alpha}(t) &= -\left[A^{\alpha}(t)^{\mathsf{T}}\Psi^{\alpha}(t)\right. \\
&\quad + \Psi^{\alpha}(t)A^{\alpha}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\mathsf{T}}\Psi^{\alpha}(t)C_i^{\alpha}(t) \\
&\quad + \sum_{i=1}^{m}\left(C_i^{\alpha}(t)^{\mathsf{T}}Q_i^{\alpha}(t) + Q_i^{\alpha}(t)C_i^{\alpha}(t)\right) \\
&\quad \left. + \psi_{n+1}^{\alpha}(t)\bar{Q}(t)\right]dt + \sum_{i=1}^{m} Q_i^{\alpha}(t)dW^i(t) \\
\Psi^{\alpha}(T) &= C_{\psi}^{\alpha}, \quad t \in [0,T] \\
d\Psi_{n+1}^{\alpha}(t) &= Q_{n+1}^{\alpha}(t)dW(t), \quad t \in [0,T] \\
\Psi_{n+1}^{\alpha}(T) &= C_{\psi,n+1}^{\alpha}
\end{aligned}
\tag{16.110}
$$

Here

- $c^{\alpha} \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R}^n)$ is a square integrable $\mathcal{F}_T$-measurable $\mathbb{R}^n$-valued random vector;
- $c_{n+1}^{\alpha} \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R})$, $\psi^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R}^n)$ is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^n$-valued vector random process;
- $\psi_{n+1}^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R})$, $q_i^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R}^n)$ and $q_{n+1}^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R}^m)$. Similarly,
- $C_i^{\alpha} \in L^2_{\mathcal{F}_T}\left(\Omega, \mathbb{R}^{n\times n}\right)$, $C_{n+1}^{\alpha} \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{R})$, $\Psi^{\alpha}(t) \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{n\times n}\right)$; $\Psi_{n+1}^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R})$, $Q_j^{\alpha}(t) \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{n\times m}\right)$ and $Q_{n+1}^{\alpha}(t) \in L^2_{\mathcal{F}_t}(\Omega, \mathbb{R}^m)$.

16.7.6 Hamiltonian form

The Hamiltonian function H^{α} is defined as

$$\begin{aligned}
&H^{\alpha} = H^{\alpha}\left(t, x, u, \psi, q, \psi_{n+1}\right) := \operatorname{tr}\left[\mathbf{q}^{\alpha\top}\boldsymbol{\sigma}^{\alpha}\right] \\
&\qquad + \left[A^{\alpha}(t)\,x^{\alpha} + B^{\alpha}(t)\,u + b^{\alpha}(t)\right]^{\top}\psi^{\alpha} + b_{n+1}^{\alpha}(t, x, u)\,\psi_{n+1}^{\alpha} \\
&\boldsymbol{\sigma}^{\alpha} := \left(\sigma_1^{\alpha}, \ldots, \sigma_m^{\alpha}\right), \mathbf{q}^{\alpha} := \left(q_1^{\alpha}, \ldots, q_m^{\alpha}\right) \\
&\sigma_i^{\alpha} := C_i^{\alpha}(t)\,x^{\alpha}(t) + D_i^{\alpha}(t)\,u(t) + \sigma_i^{\alpha}(t) \\
&W^{\top} := \left(W_1^{\top}, \ldots, W_m^{\top}\right)
\end{aligned} \tag{16.111}$$

Note that the equations in (16.105) and (16.109) can be rewritten in Hamiltonian form as

$$\begin{aligned}
&dx^{\alpha}(t) = H_{\psi}^{\alpha}dt + \boldsymbol{\sigma}^{\alpha}dW(t) \\
&x^{\alpha}(0) = x_0, \quad t \in [0, T] \\
&dx_{n+1}^{\alpha}(t) = H_{\psi_{n+1}}^{\alpha}dt \\
&x^{\alpha}(0) = x_0, \quad t \in [0, T]
\end{aligned} \tag{16.112}$$

$$\begin{aligned}
&d\psi^{\alpha}(t) = -H_x^{\alpha}dt + \mathbf{q}^{\alpha}(t)dW(t) \\
&\psi^{\alpha}(T) = c^{\alpha}, \quad t \in [0, T] \\
&d\psi_{n+1}^{\alpha}(t) = -H_{x_{n+1}}^{\alpha}dt + q_{n+1}^{\alpha}(t)\,dW(t) \\
&\psi_{n+1}^{\alpha}(T) = c_{n+1}^{\alpha}, \quad t \in [0, T]
\end{aligned} \tag{16.113}$$

Rewrite the cost function $J^{\alpha}(u(\cdot))$ as

$$\begin{aligned}
&J^{\alpha}(u(\cdot)) = \mathsf{E}\left\{h^0\left(x^{\alpha}(T), x_{n+1}^{\alpha}(T)\right)\right\} \\
&h^0\left(x^{\alpha}(T), x_{n+1}^{\alpha}(T)\right) := \mathsf{E}\left\{\frac{1}{2}x^{\alpha}(T)^{\top}Gx^{\alpha}(T)\right\} + \mathsf{E}\left\{x_{n+1}^{\alpha}(T)\right\}
\end{aligned}$$

16.7.7 Basic theorem

Now we are ready to formulate the main result of this section.

Theorem 16.3. *Let* $\left(\bar{x}(\cdot), \bar{u}(\cdot)\right)$ *be the robust optimal dynamics. Then there exist collections of terminal conditions*

$$c^{\alpha}, c_{n+1}^{\alpha}, C^{\alpha}, C_{n+1}^{\alpha}$$

$\{\mathcal{F}_t\}_{t\geq 0}$*-adapted stochastic matrices* $\left(\mathbf{q}^{\alpha}, Q_j^{\alpha}\ (j = 1, \ldots, l)\right)$ *and vectors* $\left(q_{n+1}^{\alpha}, Q_{n+1}^{\alpha}\right)$ *in (16.109) and (16.110), and nonnegative constants* μ_{α} *such that the following conditions are satisfied:*

1. *(**Complementary slackness condition**): for any* $\alpha \in \mathcal{A}$

$$\boxed{\mu_\alpha \left[\mathsf{E}\left\{ h^0\left(x^\alpha(T), x_{n+1}^\alpha(T)\right)\right\} - \max_{\alpha \in \mathcal{A}} \mathsf{E}\left\{ h^0\left(x^\alpha(T), x_{n+1}^\alpha(T)\right)\right\}\right] = 0} \quad (16.114)$$

2. *(**Transversality condition**): for any* $\alpha \in \mathcal{A}$ *with probability one*

$$\boxed{\begin{array}{c} c^\alpha + \mu_\alpha h_x^0\left(x^\alpha(T), x_{n+1}^\alpha(T)\right) = 0 \\ c_{n+1}^\alpha + \mu_\alpha = 0 \end{array}} \quad (16.115)$$

$$\boxed{\begin{array}{c} C_{\psi,n+1}^\alpha + \mu_\alpha h_{xx}^0\left(x^\alpha(T)\right) = 0 \\ C_{\psi,n+1}^\alpha = 0 \end{array}} \quad (16.116)$$

3. *(**Nontriviality condition**): there exists* $\alpha \in \mathcal{A}$ *such that* $c^\alpha, c_{n+1}^\alpha \neq 0$ *or, at least,* μ_α *is distinct from* 0*, that is,*

$$\exists \alpha : \left|c^\alpha\right| + \left|c_{n+1}^\alpha\right| + \mu_\alpha > 0 \quad (16.117)$$

4. *(**Maximality condition**): the robust optimal control* $\bar{u}(\cdot)$ *for almost all* $t \in [0, T]$ *maximizes the Hamiltonian function*

$$\boxed{\bar{H} = \sum_{\alpha \in \mathcal{A}} \bar{H}^\alpha\left(t, \bar{x}^\alpha(t), u, \psi^\alpha(t), \Psi^\alpha(t), \mathbf{q}^\alpha(t)\right)} \quad (16.118)$$

where

$$\boxed{\begin{aligned} &\bar{H}^\alpha\left(t, \bar{x}^\alpha, u, \psi^\alpha, \Psi^\alpha, \mathbf{q}^\alpha\right) \\ &\quad := H^\alpha\left(t, \bar{x}^\alpha, u, \psi^\alpha, \mathbf{q}^\alpha\right) - \frac{1}{2}\mathrm{tr}\left[\bar{\boldsymbol{\sigma}}^{\alpha\mathsf{T}}\Psi^\alpha\bar{\boldsymbol{\sigma}}^\alpha\right] \\ &\qquad + \frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{\sigma}^\alpha\left(t, \bar{x}^\alpha, u\right) - \bar{\boldsymbol{\sigma}}^\alpha\right)^\mathsf{T}\Psi^\alpha\left(\boldsymbol{\sigma}^\alpha\left(t, \bar{x}^\alpha, u\right) - \bar{\boldsymbol{\sigma}}^\alpha\right)\right] \end{aligned}} \quad (16.119)$$

and the function $H^\alpha\left(t, \bar{x}^\alpha, u, \psi^\alpha, \mathbf{q}^\alpha\right)$ *is given by (16.111),*

$$\bar{\boldsymbol{\sigma}}^\alpha = \boldsymbol{\sigma}^\alpha\left(t, \bar{x}^\alpha(t), \bar{u}(t)\right) \quad (16.120)$$

that is, for almost all $t \in [0, T]$

$$\boxed{\bar{u}(t) = \arg\max_{u \in U} \bar{H}} \quad (16.121)$$

By the transversality condition it follows that the only $\{\mathcal{F}_t\}_{t\geq 0}$-adapted variables allowed are

$$q_{n+1}^{\alpha}(t) \equiv 0,\ Q_{n+1}^{\alpha}(t) \equiv 0 \qquad \mathsf{P}\text{-}a.s.$$

and, as a result, we derive

$$\begin{aligned}
&c^{\alpha} = -\mu_{\alpha} G x^{\alpha}(T), \quad C^{\alpha} = -\mu_{\alpha} G \\
&\psi_{n+1}^{\alpha}(t) = \psi_{n+1}^{\alpha}(T) = c_{n+1}^{\alpha} = -\mu_{\alpha} \\
&\Psi_{n+1}^{\alpha}(t) = \Psi_{n+1}^{\alpha}(T) = 0
\end{aligned}$$

So, the adjoint equations become

$$\begin{aligned}
d\psi^{\alpha}(t) = &-\left[A^{\alpha}(t)^{\mathsf{T}} \psi^{\alpha}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\mathsf{T}} q_i^{\alpha}(t)\right. \\
&\left. - \mu_{\alpha}\left(\bar{Q}(t) x^{\alpha}(t) + S(t)^{\mathsf{T}} u(t)\right)\right] dt \\
&+ \sum_{i=1}^{m} \quad q_i^{\alpha}(t) dW^i(t), \quad t \in [0, T] \\
\psi^{\alpha}(T) = &-\mu_{\alpha} G x^{\alpha}(T) \\
\psi_{n+1}^{\alpha}(t) = &\ \psi_{n+1}^{\alpha}(T) = c_{n+1}^{\alpha} = -\mu_{\alpha}
\end{aligned} \tag{16.122}$$

$$\begin{aligned}
d\Psi^{\alpha}(t) = &-\left[A^{\alpha}(t)^{\mathsf{T}} \Psi^{\alpha}(t)\right. \\
&+ \Psi^{\alpha}(t) A^{\alpha}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\mathsf{T}} \Psi^{\alpha}(t) C_i^{\alpha}(t) \\
&+ \sum_{i=1}^{m} \left(C_i^{\alpha}(t)^{\mathsf{T}} Q_i^{\alpha}(t) + Q_i^{\alpha}(t) C_i^{\alpha}(t)\right) \\
&\left. - \mu_{\alpha} \bar{Q}(t)\right] dt + \sum_{i=1}^{m} Q_i^{\alpha}(t)\, dW^i(t) \\
\Psi^{\alpha}(T) = &-\mu_{\alpha} G, \quad t \in [0, T] \\
\Psi_{n+1}^{\alpha}(t) = &\ \Psi^{\alpha}(T) = 0
\end{aligned} \tag{16.123}$$

The Hamiltonian $\bar{H}$ is quadratic in u and, hence, the maximum exists if for almost all $t \in [0, T]$ with probability one

$$\nabla_u^2 \bar{H} = -\sum_{\alpha \in A} \mu_{\alpha} R + \sum_{\alpha \in A} \sum_{i=1}^{m} D_i^{\alpha}(t)^{\mathsf{T}} \Psi^{\alpha}(t) D_i^{\alpha}(t) \leq 0 \tag{16.124}$$

and the maximizing vector $\bar{u}(t)$ satisfies

$$\sum_{\alpha \in A} \mu_{\alpha} R(t) \bar{u}(t) = \sum_{\alpha \in A} \left[B^{\alpha}(t)^{\mathsf{T}} \psi^{\alpha}(t) - \mu_{\alpha} S \bar{x}^{\alpha}(t) + \sum_{i=1}^{m} D_i^{\alpha}(t)^{\mathsf{T}} q_i^{\alpha}(t)\right] \tag{16.125}$$

16.7.8 Normalized form for the adjoint equations

Since at least one active index exists it follows that

$$\sum_{\alpha \in \mathcal{A}} \mu(\alpha) > 0$$

If $\mu(\alpha) = 0$, then with the probability one

$$\begin{aligned} & q_i^{\alpha}(t) = 0, \ \ Q_i^{\alpha}(t) = 0 \\ & \dot{\psi}^{\alpha}(t) = \psi^{\alpha}(t) = 0 \\ & \dot{\Psi}^{\alpha}(t) = \Psi^{\alpha}(t) = 0 \end{aligned}$$

Therefore, the following normalized adjoint variable $\tilde{\psi}_{\alpha}(t)$ can be introduced:

$$\begin{aligned} & \tilde{\psi}_{\alpha,i}(t) = \begin{cases} \psi_{\alpha,i}(t)\,\mu^{-1}(\alpha) & \text{if} \quad \mu(\alpha) > 0 \\ 0 & \text{if} \quad \mu(\alpha) = 0 \end{cases}, \quad i = 1, \ldots, n+1 \\ & \tilde{\Psi}_{\alpha,i}(t) = \begin{cases} \Psi_{\alpha,i}(t)\,\mu^{-1}(\alpha) & \text{if} \quad \mu(\alpha) > 0 \\ 0 & \text{if} \quad \mu(\alpha) = 0 \end{cases}, \quad i = 1, \ldots, n+1 \\ & \psi^{\alpha\intercal} := \left(\psi_{\alpha,1}, \ldots, \psi_{\alpha,n}\right), \quad \Psi_{\alpha} := \left(\Psi_{\alpha,1}, \ldots, \Psi_{\alpha,n}\right)^{\intercal} \end{aligned} \tag{16.126}$$

satisfying

$$\begin{aligned} d\tilde{\psi}^{\alpha}(t) = -\Bigg[& A^{\alpha}(t)^{\intercal}\,\tilde{\psi}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\intercal}\,\tilde{q}_i^{\alpha}(t) \\ & - \left(\bar{Q}(t)\,x^{\alpha}(t) + S(t)^{\intercal}\,u(t)\right)\Bigg] dt \\ & + \sum_{i=1}^{m} \tilde{q}_i^{\alpha}(t) dW^i(t), \quad t \in [0, T] \\ d\tilde{\psi}_{\alpha,n+1}(t) = 0 & \end{aligned} \tag{16.127}$$

and

$$\begin{aligned} d\tilde{\Psi}^{\alpha}(t) = -\Bigg[& A^{\alpha}(t)^{\intercal}\,\tilde{\Psi}^{\alpha}(t) \\ & + \tilde{\Psi}^{\alpha}(t)\,A^{\alpha}(t) + \sum_{i=1}^{m} C_i^{\alpha}(t)^{\intercal}\,\tilde{\Psi}^{\alpha}(t)\,C_i^{\alpha}(t) \\ & + \sum_{i=1}^{m}\left(C_i^{\alpha}(t)^{\intercal}\,\tilde{Q}_i^{\alpha}(t) + \tilde{Q}_i^{\alpha}(t)\,C_i^{\alpha}(t)\right) \\ & - \bar{Q}(t)\Bigg] dt + \sum_{i=1i}^{m} \tilde{Q}^{\alpha}(t)\,dW^i(t) \\ d\tilde{\Psi}_{n+1}^{\alpha}(t) = 0 & \end{aligned} \tag{16.128}$$

with the transversality conditions given by

$$
\begin{aligned}
&\tilde{\psi}_{\alpha}(t_1) = -Gx^{\alpha}(t_1) \\
&\tilde{\psi}_{\alpha,n+1}(t_1) = c^{\alpha}_{n+1} = -1 \\
&\tilde{\Psi}^{\alpha}(T) = -G, \quad \tilde{\Psi}^{\alpha}_{n+1}(T) = 0
\end{aligned}
\tag{16.129}
$$

Here

$$
\tilde{q}^{\alpha}_i(t) = \begin{cases} q^{\alpha}_i(t)\,\mu^{-1}(\alpha) & \text{if} \quad \mu(\alpha) > 0 \\ 0 & \text{if} \quad \mu(\alpha) = 0 \end{cases}, \quad i = 1, \ldots, n+1
$$

$$
\tilde{\Psi}_{\alpha,i}(t) = \begin{cases} \psi_{\alpha,i}(t)\,\mu^{-1}(\alpha) & \text{if} \quad \mu(\alpha) > 0 \\ 0 & \text{if} \quad \mu(\alpha) = 0 \end{cases}, \quad i = 1, \ldots, n+1
$$

The robust optimal control (16.125) becomes (if $R > 0$) as follows:

$$
\begin{aligned}
\bar{u}(t) &= \frac{\sum\limits_{\alpha \in A} \mu_{\alpha}}{\sum\limits_{\alpha \in A} \mu_{\alpha} R(t)} \left[B^{\alpha}(t)^{\intercal} \tilde{\psi}^{\alpha}(t) - S\,\bar{x}^{\alpha}(t) + \sum_{i=1}^{m} D^{\alpha}_i(t)^{\intercal} \tilde{q}^{\alpha}_i(t) \right] \\
&= R^{-1}(t) \sum_{\alpha \in A} \lambda_{\alpha} \left[B^{\alpha}(t)^{\intercal} \tilde{\psi}^{\alpha}(t) - S\,\bar{x}^{\alpha}(t) + \sum_{i=1}^{m} D^{\alpha}_i(t)^{\intercal} \tilde{q}^{\alpha}_i(t) \right]
\end{aligned}
\tag{16.130}
$$

where the vector $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_N)^{\intercal}$ belongs to the simplex S^N defined as

$$
S^N := \left\{ \boldsymbol{\lambda} \in R^{N=|\mathcal{A}|} : \lambda_{\alpha} = \frac{\mu(\alpha)}{\sum\limits_{\alpha=1}^{N} \mu(\alpha)} \geq 0, \sum_{\alpha=1}^{N} \lambda_{\alpha} = 1 \right\}
\tag{16.131}
$$

Remark 16.1. *Since the control action can vary in the whole space $\mathbb{R}^k$, that is, there are no constraints, from the Hamiltonian structure (16.119) it follows that the robust control (16.130) does not depend on the second adjoint variables $\tilde{\Psi}^{\alpha}(t)$. This means that these variables can be omitted below. If the control u is restricted to a compact $U \subset \mathbb{R}^k$, then the robust optimal control obligatory is a function of the second adjoint variables.*

16.7.9 The extended form for the closed-loop system

For simplicity, the time argument in the expressions below is omitted.

Introduce the block-diagonal $\mathbb{R}^{nN\times nN}$ valued matrices $\mathbf{A}$, $\mathbf{Q}$, $\mathbf{G}$, $\mathbf{\Lambda}$ and the extended matrix $\mathbf{B}$ as follows:

$$\mathbf{A} := \begin{bmatrix} A^1 & 0\cdot\cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot\cdot 0 & A^N \end{bmatrix}, \quad \mathbf{Q} := \begin{bmatrix} \bar{Q} & 0\cdot\cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot\cdot 0 & \bar{Q} \end{bmatrix}$$
$$\mathbf{G} := \begin{bmatrix} G & 0\cdot\cdot & 0 \\ 0 & \cdot & 0 \\ 0 & \cdot\cdot 0 & G \end{bmatrix}, \quad \mathbf{C}_i := \begin{bmatrix} C_i^1 & 0\cdot\cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot\cdot 0 & C_i^N \end{bmatrix} \tag{16.132}$$
$$\mathbf{\Lambda} := \begin{bmatrix} \lambda_1 I_{n\times n} & 0\cdot\cdot & 0 \\ 0 & \cdot & 0 \\ 0 & \cdot\cdot 0 & \lambda_N I_{n\times n} \end{bmatrix}$$
$$\mathbf{B}^\top := \begin{bmatrix} B^{1\top} & \cdot\cdot & B^{N\top} \end{bmatrix} \in \mathbb{R}^{r\times nN}$$
$$\mathbf{D}_i^\top := \begin{bmatrix} D_i^{1\top} & \cdot\cdot & D_i^{N\top} \end{bmatrix} \in \mathbb{R}^{r\times nN}$$
$$\mathbf{S} := \begin{bmatrix} S^1 & \cdot\cdot & S^N \end{bmatrix} \in \mathbb{R}^{r\times nN}$$
$$\boldsymbol{\theta}_i := \begin{bmatrix} \sigma_i^{1\top} & \cdot\cdot & \sigma_i^{N\top} \end{bmatrix}^\top \in \mathbb{R}^{nN}$$

In view of (16.132), the dynamic equations (16.105), (16.127), (16.128) and the corresponding robust optimal control (16.130) can be represented as follows:

$$\boxed{\begin{aligned} d\mathbf{x} &= (\mathbf{A}\mathbf{x} + \mathbf{B}u + \mathbf{b})\,dt + \sum_{i=1}^{m} (\mathbf{C}_i\mathbf{x} + \mathbf{D}_i u + \mathbf{\Theta}_i)\,dW^i \\ d\boldsymbol{\psi} &= \left(-\mathbf{A}^\top\boldsymbol{\psi} - \sum_{i=1}^{m} \mathbf{C}_i^\top \mathbf{q}_i + \mathbf{Q}\mathbf{x} + \mathbf{S}^\top u\right) dt + \sum_{i=1}^{m} \mathbf{q}_i\,dW^i \\ \mathbf{x}(0) &= \begin{bmatrix} x_0^\top & x_0^\top \cdot\cdot & x_0^\top \end{bmatrix}^\top, \quad \boldsymbol{\psi}(T) = -\mathbf{G}\mathbf{x}(T) \\ u &= R^{-1}\left(\mathbf{B}^\top\mathbf{\Lambda}\boldsymbol{\psi} - \mathbf{S}\mathbf{\Lambda}\boldsymbol{x} + \sum_{i=1}^{m} \mathbf{D}_i^\top\mathbf{\Lambda}\mathbf{q}_i\right) \end{aligned}} \tag{16.133}$$

where

$$\mathbf{x}^\top := \left(x^{1\top}, \ldots, x^{N\top}\right) \in R^{1\times nN}$$
$$\boldsymbol{\psi}^\top := \left(\tilde{\boldsymbol{\psi}}_1^\top, \ldots, \tilde{\boldsymbol{\psi}}_N^\top\right) \in R^{1\times nN}$$
$$\mathbf{b}^\top := \left(b^{1\top}, \ldots, b^{N\top}\right) \in R^{1\times nN}$$
$$\mathbf{q}_i := \begin{bmatrix} \tilde{q}_i^{1\top} & \cdot\cdot & \tilde{q}_i^{N\top} \end{bmatrix}^\top \in R^{nN\times 1}$$

and

$$\mathbf{X}_0 := \mathsf{E}\left\{\mathbf{x}(0)\,\mathbf{x}^\top(0)\right\} = \begin{bmatrix} X_0 & \cdot & X_0 \\ \cdot & \cdot & \cdot \\ X_0 & \cdot & X_0 \end{bmatrix}$$

$$\mathbf{m}_0^\top := \mathsf{E}\{\mathbf{x}(0)\}^\top = \begin{bmatrix} m_0^\top & m_0^\top \cdot\cdot & m_0^\top \end{bmatrix}^\top$$

The problem now is to find the parametric matrix $\mathbf{\Lambda}$ given in the N-dimensional simplex and $\{\mathcal{F}_t\}_{t\geq 0}$-adapted extended stochastic vectors $\mathbf{q}_i \in L^2_{\mathcal{F}_t}\left(\Omega, \mathbb{R}^{nN}\right)$ minimizing the performance index (16.107).

16.7.10 Riccati equation and robust optimal control

Theorem 16.4. *The robust optimal control (16.130) achieving (16.108) is equal to*

$$\boxed{u = -R_\Lambda^{-1}\left(\mathbf{W}_\lambda \mathbf{x} + \mathbf{v}_\lambda\right)} \tag{16.134}$$

where

$$\boxed{\begin{aligned} \mathbf{v}_\lambda &:= \mathbf{B}^\top \mathbf{p}_\lambda + \sum_{i=1}^{m} \mathbf{D}_\mathbf{i}^\top \mathbf{P}_\lambda \boldsymbol{\theta}_i \\ \mathbf{W}_\lambda &:= \mathbf{B}^\top \mathbf{P}_\lambda + \mathbf{S}\mathbf{\Lambda} + \sum_{i=1}^{m} \mathbf{D}_i^\top \mathbf{P}_\lambda \mathbf{C}_i \end{aligned}} \tag{16.135}$$

and the matrix $\mathbf{P}_\lambda = \mathbf{P}_\lambda{}^T \in \mathbb{R}^{nN\times nN}$ *is the solution of the **differential matrix Riccati equation***

$$\boxed{\begin{aligned} &-\dot{\mathbf{P}}_\lambda = \mathbf{P}_\lambda \mathbf{A} + \mathbf{A}^\top \mathbf{P}_\lambda + \mathbf{\Lambda}\mathbf{Q} + \left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{m}} \mathbf{C}_\mathbf{i}^\top \mathbf{P}_\lambda \mathbf{C}_\mathbf{i}\right) - \mathbf{W}_\lambda^\top R_\Lambda^{-1} \mathbf{W}_\lambda \\ &\mathbf{P}_\lambda(t_1) = \mathbf{\Lambda}\mathbf{G} \end{aligned}} \tag{16.136}$$

$$\boxed{R_\Lambda^{-1} := \left[R + \sum_{i=1}^{m} \mathbf{D}_i^\top \mathbf{P}_\lambda \mathbf{D}_i\right]^{-1}} \tag{16.137}$$

*the **shifting vector*** $\mathbf{p}_\lambda$ *satisfies*

$$\boxed{\begin{aligned} -\dot{\mathbf{p}}_\lambda &= \mathbf{P}_\lambda \mathbf{b} + \mathbf{A}^\top \mathbf{p}_\lambda + \sum_{i=1}^{m} \mathbf{C}_i^\top \mathbf{P}_\lambda \boldsymbol{\theta}_i \\ &\quad - \mathbf{W}_\lambda^\top R_\Lambda^{-1}\left(\mathbf{B}^\top \mathbf{p}_\lambda + \sum_{i=1}^{m} \mathbf{D}_i^\top \mathbf{P}_\lambda \theta_i\right) \\ \mathbf{p}_\lambda(t_1) &= 0 \end{aligned}} \tag{16.138}$$

the matrix $\mathbf{\Lambda} = \mathbf{\Lambda}(\lambda^*)$ *is defined by (16.132) with the weight vector* $\lambda = \lambda^*$ *solving the following finite dimensional optimization problem*

$$\boxed{\lambda^* = \arg\min_{\lambda\in S^N} J_{t_1}(\lambda)} \tag{16.139}$$

$$
\begin{aligned}
J_{t_1}(\lambda) = {} & \frac{1}{2}\operatorname{tr}\{\mathbf{X}_0\mathbf{P}_\lambda(0)\} + \mathbf{m}_0^\mathsf{T}\mathbf{p}_\lambda(0) \\
& + \frac{1}{2}\int_{t=0}^{t_1}\left[\sum_{i=1}^{m}\boldsymbol{\theta}_i^\mathsf{T}\mathbf{P}_\lambda\boldsymbol{\theta}_i + 2\mathbf{p}^\mathsf{T}\mathbf{b} - \mathbf{v}^\mathsf{T}R_\Lambda^{-1}\mathbf{v}\right]dt \\
& + \frac{1}{2}\max_{i=1,N}\mathsf{E}\left\{\int_0^{t_1}\left(x^{i\mathsf{T}}Qx^i + 2u^\mathsf{T}Sx^i\right)dt + x^{i\mathsf{T}}(t_1)\,Gx^i(t_1)\right\} \\
& - \frac{1}{2}\sum_{i=1}^{N}\lambda_i\mathsf{E}\left\{\int_0^{t_1}\left(x^{i\mathsf{T}}Qx^i + 2u^\mathsf{T}Sx^i\right)dt + x^{i\mathsf{T}}(t_1)\,Gx^i(t_1)\right\}
\end{aligned}
\tag{16.140}
$$

and

$$
\boxed{J(\bar{u}(\cdot)) = \min_{u(\cdot)\in U_{adm}^s[0,T]}\max_{\alpha\in\mathcal{A}} J^\alpha(u(\cdot)) = J_{t_1}\left(\lambda^*\right)}
\tag{16.141}
$$

Proof. Since the robust optimal control (16.133) is proportional to $\boldsymbol{\Lambda\psi}$, it is natural to find $\boldsymbol{\psi}$ satisfying

$$
\boldsymbol{\Lambda\psi}(t) = -\mathbf{P}_\lambda(t)\,\mathbf{x} - \mathbf{p}_\lambda(t)
\tag{16.142}
$$

where $\mathbf{P}_\lambda(t)$ and $\mathbf{p}_\lambda(t)$ are a differentiable deterministic matrix and vector, respectively. The commutation of the operators

$$
\boldsymbol{\Lambda}^k\mathbf{A} = \mathbf{A}\boldsymbol{\Lambda}^k, \quad \boldsymbol{\Lambda}^k\mathbf{Q} = \mathbf{Q}\boldsymbol{\Lambda}^k, \quad \boldsymbol{\Lambda}^k\mathbf{C}_i = \mathbf{C}_i\boldsymbol{\Lambda}^k\ (k \geq 0)
$$

implies

$$
\begin{aligned}
\boldsymbol{\Lambda}d\boldsymbol{\psi} &= -\dot{\mathbf{P}}_{\boldsymbol{\lambda}}\mathbf{x}dt \\
&\quad - \mathbf{P}_\lambda\left[(\mathbf{Ax} + \mathbf{B}u + \mathbf{b})\,dt + \sum_{i=1}^{m}(\mathbf{C}_i\mathbf{x} + \mathbf{D}_i u + \boldsymbol{\theta}_i)\,dW^i\right] - d\mathbf{p}_\lambda \\
&= \boldsymbol{\Lambda}\left[\left(-\mathbf{A}^\mathsf{T}\boldsymbol{\psi} - \sum_{i=1}^{m}\mathbf{C}_i^\mathsf{T}\mathbf{q}_i + \mathbf{Qx} + \mathbf{S}^\mathsf{T}u\right)dt + \sum_{i=1}^{m}\mathbf{q}_i\,dW^i\right]
\end{aligned}
\tag{16.143}
$$

from which it follows that

$$
\boldsymbol{\Lambda}\mathbf{q}_i = -\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{C}_i\mathbf{x} + \mathbf{D}_i u + \boldsymbol{\theta}_i)
\tag{16.144}
$$

The substitution of (16.142) and (16.144) into (16.133) leads to

$$
u = -R^{-1}\left(\mathbf{B}^\mathsf{T}\left[\mathbf{P}_{\boldsymbol{\lambda}}\mathbf{x} + \mathbf{p}_\lambda\right] + \mathbf{S\Lambda x} + \sum_{i=1}^{m}\mathbf{D}_i^\mathsf{T}\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{C}_i\mathbf{x} + \mathbf{D}_i u + \boldsymbol{\theta}_i)\right)
$$

which is equivalent to (16.134). Then, (16.143), (16.144) and the matrix commutations

$$
\boldsymbol{\Lambda}\mathbf{A}^\mathsf{T} = \mathbf{A}^\mathsf{T}\boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda}\mathbf{C}_\mathbf{i}^\mathsf{T} = \mathbf{C}_i^\mathsf{T}\boldsymbol{\Lambda}
$$

imply

$$-\left(\dot{\mathbf{P}}_{\lambda}+\mathbf{P}_{\lambda}\mathbf{A}+\mathbf{A}^{\top}\mathbf{P}_{\lambda}+\mathbf{\Lambda Q}+\left(\sum_{\mathbf{i=1}}^{\mathbf{m}}\mathbf{C}_{\mathbf{i}}^{\top}\mathbf{P}_{\lambda}\mathbf{C}_{\mathbf{i}}\right)-\mathbf{W}_{\lambda}^{\top}R_{\Lambda}^{-1}\mathbf{W}_{\lambda}\right)\mathbf{x}$$
$$=\mathbf{P}_{\lambda}\mathbf{b}+\dot{\mathbf{p}}_{\lambda}+\mathbf{A}^{\top}\mathbf{p}_{\lambda}+\sum_{i=1}^{m}\mathbf{C}_{i}^{\top}\mathbf{P}_{\lambda}\boldsymbol{\theta}_{i}-\mathbf{W}_{\lambda}^{\top}R_{\Lambda}^{-1}\left(\mathbf{B}^{\top}\mathbf{p}_{\lambda}+\sum_{i=1}^{m}\mathbf{D}_{i}^{\top}\mathbf{P}_{\lambda}\boldsymbol{\theta}_{i}\right)$$

This equation is satisfied identically under the conditions (16.136) and (16.138) of this theorem. This implies:

$$\begin{aligned}J_{t_1}(\lambda) &:= \max_{\alpha\in\mathcal{A}} J^{\alpha} = \max_{\nu\in S^N}\sum_{i=1}^{N}\nu_i J^i\\
&= \frac{1}{2}\sum_{i=1}^{N}\nu_i \mathsf{E}\left\{\int_0^{t_1}\left[u^{\top}Ru+x^{i\top}\bar{Q}x^{i}+2u^{\top}Sx^{i}\right]dt+x^{i\top}(t_1)Gx^{i}(t_1)\right\}\\
&= \frac{1}{2}\mathsf{E}\int_0^{t_1}u^{\top}Rudt+\frac{1}{2}\max_{\nu\in S^N}\mathsf{E}\left\{\int_0^{t_1}\left(\mathbf{x}^{\top}\mathbf{Q}_{\nu}\mathbf{x}+\mathbf{2}u^{\top}\mathbf{S}_{\nu}\mathbf{x}\right)dt+\mathbf{x}^{\top}(t_1)\mathbf{G}_{\nu}\mathbf{x}(t_1)\right\}\end{aligned}\quad(16.145)$$

where

$$\mathbf{Q}_{\nu}:=\begin{bmatrix}\nu_1\bar{Q} & 0 & \cdot & 0\\ 0 & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & 0\\ 0 & \cdot & 0 & \nu_N\bar{Q}\end{bmatrix},\quad \mathbf{G}_{\nu}:=\begin{bmatrix}\nu_1 G & 0 & \cdot & 0\\ 0 & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & 0\\ 0 & \cdot & 0 & \nu_N G\end{bmatrix}$$
$$\mathbf{S}_{\nu}:=\begin{bmatrix}\nu_1 S & 0 & \cdot & 0\\ 0 & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & 0\\ 0 & \cdot & 0 & \nu_N S\end{bmatrix}$$

and, hence, in view of (16.133), (16.143) and (16.144), it follows that

$$\mathsf{E}\int_0^{t_1}u^{\top}Rudt=\mathsf{E}\int_0^{t_1}u^{\top}\left(\mathbf{B}^{\top}\mathbf{\Lambda}\boldsymbol{\psi}-\mathbf{S\Lambda x}+\sum_{i=1}^{m}\mathbf{D}_i^{\top}\mathbf{\Lambda}\mathbf{q}_i\right)dt$$
$$=\mathsf{E}\int_0^{t_1}\left(u^{\top}\mathbf{B}^{\top}\mathbf{\Lambda}\boldsymbol{\psi}+u^{\top}\sum_{i=1}^{m}\mathbf{D}_i^{\top}\mathbf{\Lambda}\mathbf{q}_i\right)dt-\mathsf{E}\int_0^{t_1}u^{\top}\mathbf{S\Lambda x}dt$$
$$=\mathsf{E}\int_0^{t_1}\left(\left[u^{\top}\mathbf{B}^{\top}+\mathbf{b}^{\top}+\mathbf{x}^{\top}\mathbf{A}^{\top}\right]\mathbf{\Lambda}\boldsymbol{\psi}-u^{\top}\sum_{i=1}^{m}\mathbf{D}_i^{\top}\mathbf{P}_{\lambda}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_i u+\boldsymbol{\theta}_i\right)\right)dt$$

$$-E\int_0^{t_1}\left[\mathbf{b}^\top+\mathbf{x}^\top\mathbf{A}^\top\right]\boldsymbol{\Lambda}\boldsymbol{\psi}\,dt-E\int_0^{t_1}u^\top\mathbf{S}\boldsymbol{\Lambda}\mathbf{x}dt$$

$$=E\int_0^{t_1}d\mathbf{x}^\top\boldsymbol{\Lambda}\boldsymbol{\psi}-E\int_0^{t_1}u^\top\sum_{i=1}^{m}\mathbf{D}_i^\top\mathbf{P}_{\boldsymbol{\lambda}}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)dt-E\int_0^{t_1}\mathbf{b}^\top\boldsymbol{\Lambda}\boldsymbol{\psi}\,dt$$

$$-\mathsf{E}\int_0^{t_1}u^\top\mathbf{S}\boldsymbol{\Lambda}\mathbf{x}dt+\mathsf{E}\int_0^{t_1}\mathbf{x}^\top\boldsymbol{\Lambda}\left[-\mathbf{A}^\top\boldsymbol{\psi}-\sum_{i=1}^{m}\mathbf{C}_i^\top\mathbf{q}_i+\mathbf{Q}\mathbf{x}+\mathbf{S}^\top u\right]dt$$

$$+\mathsf{E}\int_0^{t_1}\mathbf{x}^\top\boldsymbol{\Lambda}\left[\sum_{i=1}^{m}\mathbf{C}_i^\top\mathbf{q}_i-\mathbf{Q}\mathbf{x}-\mathbf{S}^\top u\right]dt=\mathsf{E}\int_0^{t_1}\left(d\mathbf{x}^\top\boldsymbol{\Lambda}\boldsymbol{\psi}+\mathbf{x}^\top\boldsymbol{\Lambda}d\boldsymbol{\psi}\right)$$

$$-\mathsf{E}\int_0^{t_1}u^\top\sum_{i=1}^{m}\mathbf{D}_i^\top\mathbf{P}_{\boldsymbol{\lambda}}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)dt-\mathsf{E}\int_0^{t_1}\mathbf{b}^\top\boldsymbol{\Lambda}\boldsymbol{\psi}\,dt$$

$$-\mathsf{E}\int_0^{t_1}\mathbf{x}^\top\sum_{i=1}^{m}\mathbf{C}_i^\top\mathbf{P}_{\boldsymbol{\lambda}}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)dt-\mathsf{E}\int_0^{t_1}\mathbf{x}^\top\boldsymbol{\Lambda}\mathbf{Q}\mathbf{x}dt-2\mathsf{E}\int_0^{t_1}u^\top\mathbf{S}\boldsymbol{\Lambda}\mathbf{x}dt$$

The application of the Itô formula and the use of (16.136) and (16.138) imply

$$\mathsf{E}\int_0^{t_1}u^\top Ru\,dt=\mathsf{E}\int_0^{t_1}d\left(\mathbf{x}^\top\left[-\mathbf{P}_\lambda\mathbf{x}-\mathbf{p}_\lambda\right]\right)$$

$$+\mathsf{E}\int_0^{t_1}\sum_{i=1}^{m}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)^\top\mathbf{P}_\lambda\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)dt$$

$$+\mathsf{E}\int_0^{t_1}\mathbf{b}^\top\left[\mathbf{P}_\lambda\mathbf{x}+\mathbf{p}_\lambda\right]dt$$

$$-\mathsf{E}\int_0^{t_1}\sum_{i=1}^{m}\left(u^\top\mathbf{D}_i^\top+\mathbf{x}^\top\mathbf{C}_i^\top\right)\mathbf{P}_{\boldsymbol{\lambda}}\left(\mathbf{C}_i\mathbf{x}+\mathbf{D}_iu+\boldsymbol{\theta}_i\right)dt$$

$$-\mathsf{E}\int_0^{t_1}\mathbf{x}^\top\boldsymbol{\Lambda}\mathbf{Q}\mathbf{x}dt-2\mathsf{E}\int_0^{t_1}u^\top\mathbf{S}\boldsymbol{\Lambda}\mathbf{x}dt$$

$$
\begin{aligned}
&= \mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{P}_{\lambda}(0)\,\mathbf{x}(0) - \mathsf{E}\mathbf{x}^{\mathsf{T}}(t_1)\,\mathbf{\Lambda}\mathbf{G}\mathbf{x}(t_1) \\
&\quad + \mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{p}_{\lambda}(0) + \mathsf{E}\int_0^{t_1}\sum_{i=1}^{m}\boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\left(\mathbf{C}_i\mathbf{x} + \mathbf{D}_i u + \boldsymbol{\theta}_i\right)dt \\
&\quad + \mathsf{E}\int_0^{t_1}\mathbf{b}^{\mathsf{T}}\left[\mathbf{P}_{\lambda}\mathbf{x} + \mathbf{p}_{\lambda}\right]dt - \mathsf{E}\int_0^{t_1}\mathbf{x}^{\mathsf{T}}\mathbf{\Lambda}\mathbf{Q}\mathbf{x}dt - 2\mathsf{E}\int_0^{t_1}u^{\mathsf{T}}\mathbf{S}\mathbf{\Lambda}\mathbf{x}dt \qquad (16.146)
\end{aligned}
$$

The use of (16.133) and (16.138) leads to the identity

$$
\begin{aligned}
&-\mathsf{E}\left\{\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{p}_{\lambda}(0)\right\} = \mathsf{E}\left\{\mathbf{x}^{\mathsf{T}}(t_1)\,\mathbf{p}(t_1) - \mathbf{x}^{\mathsf{T}}(0)\,\mathbf{p}_{\lambda}(0)\right\} \\
&= \mathsf{E}\int_{t=0}^{t_1} d\left(\mathbf{x}^{\mathsf{T}}\mathbf{p}_{\lambda}\right) = \mathsf{E}\int_{t=0}^{t_1}\Bigg[\mathbf{p}_{\lambda}^{\mathsf{T}}\left(\mathbf{A}\mathbf{x} + \mathbf{B}u + \mathbf{b}\right) \\
&\quad + \mathbf{x}^{\mathsf{T}}\left(-\mathbf{P}_{\lambda}\mathbf{b} - \mathbf{A}^{\mathsf{T}}\mathbf{p}_{\lambda} - \sum_{i=1}^{m}\mathbf{C}_i^{\mathsf{T}}\mathbf{P}_{\boldsymbol{\lambda}}\boldsymbol{\theta}_i + \mathbf{W}_{\lambda}^{\mathsf{T}}R_{\Lambda}^{-1}\left(\mathbf{B}^{\mathsf{T}}\mathbf{p}_{\lambda} + \sum_{i=1}^{m}\mathbf{D}_i^{\mathsf{T}}\mathbf{P}_{\boldsymbol{\lambda}}\boldsymbol{\theta}_i\right)\right)\Bigg]dt \\
&= \mathsf{E}\int_{t=0}^{t_1}\Bigg[\mathbf{p}_{\lambda}^{\mathsf{T}}\left(\mathbf{b} - \mathbf{B}R_{\Lambda}^{-1}\mathbf{v}_{\lambda}\right) \\
&\quad - \mathbf{x}^{\mathsf{T}}\left(\mathbf{P}_{\lambda}\mathbf{b} + \sum_{i=1}^{m}\mathbf{C}_i^{\mathsf{T}}\mathbf{P}_{\boldsymbol{\lambda}}\boldsymbol{\theta}_i - \mathbf{W}_{\lambda}^{\mathsf{T}}R_{\Lambda}^{-1}\sum_{i=1}^{m}\mathbf{D}_i^{\mathsf{T}}\mathbf{P}_{\boldsymbol{\lambda}}\boldsymbol{\theta}_i\right)\Bigg]dt
\end{aligned}
$$

The substitution of this identity into (16.146) implies

$$
\begin{aligned}
\mathsf{E}\int_0^{t_1}u^{\mathsf{T}}Ru\,dt &= \mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{P}_{\lambda}(0)\,\mathbf{x}(0) \\
&\quad - \mathsf{E}\mathbf{x}^{\mathsf{T}}(t_1)\,\mathbf{\Lambda}\mathbf{G}\mathbf{x}(t_1) + \mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{p}_{\lambda}(0) \\
&\quad + \mathsf{E}\int_0^{t_1}\Bigg[\left(\sum_{i=1}^{m}\boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\mathbf{C}_i - \left(\sum_{i=1}^{m}\boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\mathbf{D}_i\right)R_{\Lambda}^{-1}\mathbf{W}_{\lambda} + \mathbf{b}^{\mathsf{T}}\mathbf{P}_{\lambda}\right)\mathbf{x} \\
&\quad - \sum_{i=1}^{m}\boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\mathbf{D}_i R_{\Lambda}^{-1}\mathbf{v}_{\lambda} + \sum_{i=1}^{m}\boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\boldsymbol{\theta}_i dt\Bigg] \\
&\quad + \mathsf{E}\int_0^{t_1}\mathbf{b}^{\mathsf{T}}\mathbf{p}_{\lambda}dt - \mathsf{E}\int_0^{t_1}\mathbf{x}^{\mathsf{T}}\mathbf{\Lambda}\mathbf{Q}\mathbf{x}dt - 2\mathsf{E}\int_0^{t_1}u^{\mathsf{T}}\mathbf{S}\mathbf{\Lambda}\mathbf{x}dt
\end{aligned}
$$

$$= \mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{P}_{\lambda}(0)\,\mathbf{x}(0) - \mathsf{E}\mathbf{x}^{\mathsf{T}}(t_1)\,\mathbf{\Lambda}\mathbf{G}\mathbf{x}(t_1) + 2\mathsf{E}\mathbf{x}^{\mathsf{T}}(0)\,\mathbf{p}_{\lambda}(0)$$
$$+ 2\int_{0}^{t_1} \mathbf{b}^{\mathsf{T}}\mathbf{p}_{\lambda}dt - \int_{t=0}^{t_1} \mathbf{v}_{\lambda}^{\mathsf{T}} R_{\Lambda}^{-1}\mathbf{v}_{\lambda}dt + \int_{t=0}^{t_1} \sum_{i=1}^{m} \boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\boldsymbol{\theta}_i dt$$
$$- \mathsf{E}\int_{0}^{t_1} \mathbf{x}^{\mathsf{T}}\mathbf{\Lambda}\mathbf{Q}\mathbf{x}dt - 2\mathsf{E}\int_{0}^{t_1} u^{\mathsf{T}}\mathbf{S}\mathbf{\Lambda}\mathbf{x}dt$$

which is equivalent to the following expression

$$\begin{aligned}
\mathsf{E}\int_{0}^{t_1} u^{\mathsf{T}} R u dt = {} & \operatorname{tr}\{\mathbf{X}_0\mathbf{P}_{\lambda}(0)\} \\
& - \mathsf{E}\mathbf{x}^{\mathsf{T}}(t_1)\,\mathbf{\Lambda}\mathbf{G}\mathbf{x}(t_1) + 2\mathbf{m}^{\mathsf{T}}\mathbf{p}_{\lambda}(0) \\
& - \int_{t=0}^{t_1} \mathbf{v}_{\lambda} R_{\Lambda}^{-1}\mathbf{v}_{\lambda}dt + \int_{0}^{t_1} \sum_{i=1}^{m} \boldsymbol{\theta}_i^{\mathsf{T}}\mathbf{P}_{\lambda}\boldsymbol{\theta}_i dt + 2\int_{0}^{t_1} \mathbf{b}^{\mathsf{T}}\mathbf{p}_{\lambda}dt \\
& - \mathsf{E}\int_{0}^{t_1} \mathbf{x}^{\mathsf{T}}\mathbf{\Lambda}\mathbf{Q}\mathbf{x}dt - 2\mathsf{E}\int_{0}^{t_1} u^{\mathsf{T}}\mathbf{S}\mathbf{\Lambda}\mathbf{x}dt
\end{aligned} \tag{16.147}$$

The relation (16.146) together with (16.145) leads to (16.139). The theorem is proven. □

16.7.11 Linear stationary systems with infinite horizon

Consider the class of linear stationary controllable systems (16.105) without exogenous input:

$$A^{\alpha}(t) \equiv A^{\alpha}, \quad B^{\alpha}(t) \equiv B^{\alpha}, \quad \sigma_i^{\alpha}(t) \equiv \sigma_i^{\alpha}$$
$$b(t) = 0, \quad C_i^{\alpha}(t) \equiv 0, \quad D_i^{\alpha}(t) \equiv 0$$

and containing the only integral term ($G = 0$) with $S(t) \equiv 0$, that is,

$$\boxed{\begin{gathered}
dx(t) = \left[A^{\alpha}x(t) + B^{\alpha}u(t)\right]dt + \sum_{i=1}^{m} \sigma_i^{\alpha} dW^i(t) \\
x(0) = x_0, \quad t \in [0, T]\,(T > 0) \\
J^{\alpha}(u(\cdot)) = \frac{1}{2}\mathsf{E}\int_{t=0}^{T} \left[x^{\alpha}(t)^{\mathsf{T}}\bar{Q}(t)\,x^{\alpha}(t)\right. \\
\left. + u(t)^{\mathsf{T}} R(t)\,u(t)\,dt\right]
\end{gathered}} \tag{16.148}$$

Then, from (16.138) and (16.140) it follows that $\mathbf{p}(t) \equiv 0$, $\mathbf{v}(t) \equiv 0$ and, hence,

$$
\begin{aligned}
J_{t_1}(\lambda) = & \frac{1}{2} \operatorname{tr}\{\mathbf{X}_0 \mathbf{P}_\lambda(0)\} \\
& + \frac{1}{2} \int_{t=0}^{t_1} \operatorname{tr}\left\{\left(\sum_{i=1}^{m} \boldsymbol{\theta}_i \boldsymbol{\theta}_i^\intercal\right) \mathbf{P}_\lambda\right\} dt \\
& + \frac{1}{2} \max_{i=1,N} \mathsf{E}\left\{\int_0^{t_1}\left[x^{i\intercal} Q x^i - \sum_{j=1}^{N} \lambda_j x^{j\intercal} Q x^j dt\right] dt\right\}
\end{aligned} \tag{16.149}
$$

and

$$
\min_{u(\cdot) \in U_{adm}^s[0,T]} \max_{\alpha \in \mathcal{A}} J^\alpha(u(\cdot)) = \min_{\lambda \in S^N} J_{t_1}(\lambda)
$$

Nothing changes, if instead of $J^\alpha(u(\cdot))$ we will deal initially with the so-called '*time-averaged*' cost function

$$
\min_{u(\cdot) \in U_{adm}^s[0,T]} \max_{\alpha \in \mathcal{A}} \frac{1}{t_1} J^\alpha(u(\cdot)) = \frac{1}{t_1} J_{t_1}\left(\lambda^*\right)
$$

making the formal substitution

$$
Q \to \frac{1}{t_1} Q, \quad R \to \frac{1}{t_1} R, \quad G \to \frac{1}{t_1} G
$$

Indeed, this transforms (16.136) to the following equation

$$
\begin{aligned}
& -\dot{\mathbf{P}}_\lambda = \mathbf{P}_\lambda \mathbf{A} + \mathbf{A}^\intercal \mathbf{P}_\lambda - \mathbf{P}_\lambda \mathbf{B}\left(\frac{1}{t_1} R\right)^{-1} \mathbf{B}^\intercal \mathbf{P}_\lambda + \frac{1}{t_1} \Lambda \mathbf{Q} \\
& \mathbf{P}_\lambda(t_1) = \frac{1}{t_1} \Lambda \mathbf{G} = \mathbf{0}
\end{aligned}
$$

or

$$
\begin{aligned}
& \frac{d}{dt} \tilde{\mathbf{P}}_\lambda + \tilde{\mathbf{P}}_\lambda \mathbf{A} + \mathbf{A}^\intercal \tilde{\mathbf{P}}_\lambda - \tilde{\mathbf{P}}_\lambda \mathbf{B} R^{-1} \mathbf{B}^\intercal \tilde{\mathbf{P}}_\lambda + \boldsymbol{\Lambda} \mathbf{Q} = 0 \\
& \tilde{\mathbf{P}} := t_1 \mathbf{P}\left(t \in [0, t_1]\right), \quad \tilde{\mathbf{P}}(t_1) = \mathbf{0}
\end{aligned} \tag{16.150}
$$

and, hence, the robust optimal control (16.134) remains the same

$$
u = -\left(\frac{1}{t_1} R\right)^{-1} \mathbf{B}^\intercal \mathbf{P}_\lambda \mathbf{x} = -R^{-1} \mathbf{B}^\intercal \tilde{\mathbf{P}}_\lambda \mathbf{x}
$$

For any $t \geq 0$ and some $\varepsilon > 0$ let us define another matrix function, say $\bar{\mathbf{P}}_\lambda$, as follows:

$$
\bar{\mathbf{P}}_\lambda := \begin{cases} \tilde{\mathbf{P}}_\lambda & \text{if} \quad t \in [0, t_1] \\ \tilde{\mathbf{P}}_\lambda\left[1 - \sin\left(\frac{\pi}{2\varepsilon}(t - t_1)\right)\right] + \tilde{\mathbf{P}}_{st} \sin\left(\frac{\pi}{2\varepsilon}(t - t_1)\right) & \text{if} \quad t \in (t_1, t_1 + \varepsilon] \\ \tilde{\mathbf{P}}_{st} & \text{if} \quad t > t_1 + \varepsilon \end{cases}
$$

where $\tilde{\mathbf{P}}_{st}$ is the solution to the algebraic Riccati equation

$$\tilde{\mathbf{P}}_{st}\mathbf{A} + \mathbf{A}^{\intercal}\tilde{\mathbf{P}}_{st} - \tilde{\mathbf{P}}_{st}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\intercal}\tilde{\mathbf{P}}_{st} + \mathbf{\Lambda Q} = 0 \tag{16.151}$$

This matrix function $\bar{\mathbf{P}}_{\lambda}$ is differentiable for all $t \in [0.\infty)$. If the algebraic Riccati equation (16.151) has a positive definite solution $\tilde{\mathbf{P}}_{st}$ (when the pair $\left(\mathbf{A}, R^{1/2}\right)$ is controllable and the pair $\left(\mathbf{Q}^{1/2}, \mathbf{A}\right)$ is observable, see, for example, Poznyak (2008)) for any $\lambda \in S^N$, then $\bar{\mathbf{P}}_{\lambda}(t) \underset{t\to\infty}{\rightarrow} \tilde{\mathbf{P}}_{st}$ for any $\tilde{\mathbf{P}}(t_1)$. Tending t_1 to ∞ leads to the following result.

Corollary 16.1. *The robust optimal control $\bar{u}(\cdot)$ solving the min-max problem*

$$\boxed{\begin{gathered} J(\bar{u}(\cdot)) := \min_{u(\cdot)\in U^s_{adm}[0,\infty]} \max_{\alpha\in\mathcal{A}} \limsup_{t_1\to\infty} \\ \mathsf{E}\left\{\frac{1}{2t_1}\int_{t=0}^{t_1}\left(\left\|x^{\alpha}(t)\right\|^2_{Q} + \|u(t)\|^2_R\right)dt\right\}\end{gathered}} \tag{16.152}$$

is given by

$$\boxed{u = -R^{-1}\mathbf{B}^{\intercal}\tilde{\mathbf{P}}_{st}(\Lambda)\,\mathbf{x}} \tag{16.153}$$

where the matrix $\mathbf{\Lambda} = \mathbf{\Lambda}(\lambda^*)$ is defined by (16.132) with the weight vector $\lambda = \lambda^*$ solving the following finite dimensional optimization problem

$$\boxed{\begin{gathered} \lambda^* = \arg\min_{\lambda\in S^N} J_{\infty}(\lambda) \\ J_{\infty}(\lambda) = \frac{1}{2}\mathrm{tr}\left\{\left(\sum_{i=1}^{m}\boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\intercal}\right)\tilde{\mathbf{P}}_{st}(\Lambda)\right\} + \max_{i=\overline{1,N}} \mathrm{tr}\left\{X^i Q - \sum_{j=1}^{N}\lambda_j X^j Q\right\} \\ X^i := \limsup_{t\to\infty}\frac{1}{2t_1}\int_0^{t_1}\mathsf{E}\left\{x^i(t)\,x^{i\intercal}(t)\right\}dt \end{gathered}} \tag{16.154}$$

and

$$J(\bar{u}(\cdot)) = J_{\infty}\left(\lambda^*\right) \tag{16.155}$$

The application of the robust optimal control (16.153) provides for the corresponding closed loop system, the so-called '**ergodicity**' property that implies the existence of the limit (not only upper limit) for the averaged cost function $t_1^{-1}J_{t_1}(\lambda^*)$ when $t_1 \to \infty$, that is,

$$X^i = \lim_{t\to\infty}\mathsf{E}\left\{x^i(t)\,x^{i\intercal}(t)\right\}$$

and t

$$t_1^{-1} J_{t_1}\left(\lambda^*\right) \underset{t\to\infty}{\to} J_\infty(\lambda)$$

Below we show that the matrices X^i may be calculated as the solution of the algebraic Lyapunov matrix equation.

16.7.12 Numerical example

Let us consider one dimensional (double-structured) plant given by

$$\begin{aligned} dx^\alpha(t) &= \left[a^\alpha x^\alpha(t) + b^\alpha u(t)\right] dt + \sigma^\alpha dW \\ x^\alpha(0) &= x_0, \quad \alpha = 1, 2 \end{aligned}$$

and the performance index defined as

$$h^\alpha = \limsup_{t_1\to\infty} \frac{1}{2t_1} \int_{t=0}^{t_1} \mathsf{E}\left\{q\left[x^\alpha(t)\right]^2 + r\left[u(t)\right]^2\right\} dt, \quad q \ge 0, r > 0$$

To solve the min-max problem

$$\max_{\alpha=1,2} h^\alpha \to \min_{u(\cdot)}$$

according to Corollary given above, let us apply the robust control (16.130)

$$\begin{aligned} u &= -r^{-1}\mathbf{B}^\intercal \tilde{\mathbf{P}}_{st}(\Lambda)\,\mathbf{x} \\ &= -r^{-1}\begin{bmatrix} b^1 \\ b^2 \end{bmatrix}^\intercal \begin{bmatrix} \tilde{\mathbf{P}}_{11}(\boldsymbol{\lambda}^*) & \tilde{\mathbf{P}}_{12}(\boldsymbol{\lambda}^*) \\ \tilde{\mathbf{P}}_{21}(\boldsymbol{\lambda}^*) & \tilde{\mathbf{P}}_{22}(\boldsymbol{\lambda}^*) \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \end{aligned} \tag{16.156}$$

where

$$\begin{aligned} \lambda^* &= \arg\min_{\lambda\in S^2} J_\infty(\lambda) \\ J_\infty(\lambda) &= \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\theta}\boldsymbol{\theta}^\intercal \tilde{\mathbf{P}}_{st}(\Lambda)\right\} + q\max_{i=1,2} \mathrm{tr}\left\{X^i - \sum_{j=1}^{2}\lambda_j X^j\right\} \\ &= \frac{1}{2}\begin{pmatrix}\sigma^1\\ \sigma^2\end{pmatrix}^\intercal \tilde{\mathbf{P}}_{st}(\boldsymbol{\Lambda}) \begin{pmatrix}\sigma^1\\ \sigma^2\end{pmatrix} + q\max_{i=1,2} \mathrm{tr}\left\{X^i - \sum_{j=1}^{2}\lambda_j X^j\right\} \\ \lambda_1 + \lambda_2 &= 1, \quad \lambda_i \ge 0 \end{aligned}$$

Here, if $\lambda_i > 0$, the matrix $\mathbf{X} = \begin{bmatrix} X^1 & \cdot \\ \cdot & X^2 \end{bmatrix}$ is the solution of the following Lyapunov matrix equation

$$\mathbf{X}\left(\mathbf{A} - \mathbf{B}\mathbf{B}^\intercal \tilde{\mathbf{P}}_{st}(\Lambda)\right)^\intercal + \left(\mathbf{A} - \mathbf{B}\mathbf{B}^\intercal \tilde{\mathbf{P}}_{st}(\Lambda)\right)\mathbf{X} = -\begin{pmatrix}\sigma^1\\ \sigma^2\end{pmatrix}\begin{pmatrix}\sigma^1\\ \sigma^2\end{pmatrix}^\intercal$$

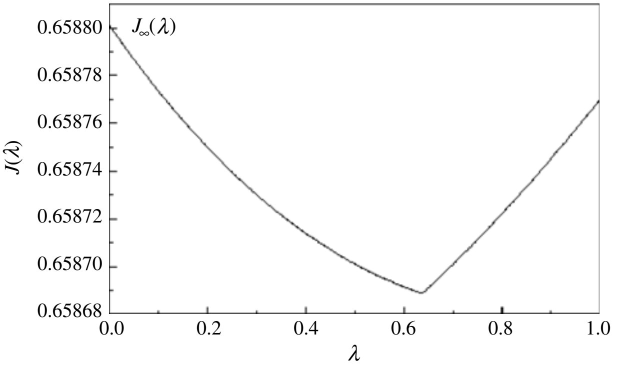

Fig. 16.4. The dependence of the performance index J_∞ on the weighting parameter λ.

So, for

$$\begin{array}{llll} a_1 = 2 & b_1 = 1 & \sigma_1 = 0.1 & q = 1 \\ a_2 = 2.15 & b_2 = 0.2 & \sigma_2 = 0.1 & r = 1 \end{array}$$

the function $J_\infty(\lambda)$ is shown at Fig. 16.4. One can see that the minimum value is achieved at

$$\lambda = \lambda^* \cong \begin{pmatrix} 0.63 \\ 0.37 \end{pmatrix}$$

16.8 Conclusion

- The min-max linear quadratic problems formulated for stochastic differential equations, containing in general a control-dependent diffusion term, are shown to be solved by the robust maximum principle formulated in this chapter.
- The corresponding Hamiltonian formalism is constructed based on the parametric families of the first and second order adjoint stochastic processes.
- The robust optimal control *maximizes* the Hamiltonian function (at every time t) which is equal to the sum (in general, an integral) of the standard stochastic Hamiltonians corresponding to each value of an uncertain parameter from a given compact (or, particular, finite) set.
- In the case of finite uncertainty set the construction of the min-max optimal controller is reduced to an optimization problem in a finite-dimensional simplex.

This page intentionally left blank

BIBLIOGRAPHY

Albert, A., Gardner, L., 1967. Stochastic Approximation and Nonlinear Regression. MIT Press, Cambridge, Massachusetts.

Bellman, R., 1960. Introduction to Matrix Analysis. McGraw-Hill Book Company, Inc., New York, USA.

Bensoussan, A., 1983. Lecture on Stochastic Control, Part 1. In: Lecture Notes in Math., vol. 972. Springer-Verlag, NY–Berlin, pp. 1–39.

Bensoussan, A., 1992. Stochastic Control of Partially Observable Systems. Cambridge University Press, Cambridge.

Bernhard, P. 1994. Minimax versus stochastic partial information control. In: Proc. 33rd Conf. Decision Control, Lake Buena Vista, Florida. pp. 2572–2577.

Bingham, N., 1986. Variants on the law of the iterated logarithm. Bull. London Math. Soc. 18, 433–467.

Bismut, J., 1977. On optimal control of linear stochastic equations with a linear quadratic criteria. SIAM J. Control 15(1), 1–4.

Bismut, J., 1978. An introductory approach to duality in stochastic control. SIAM Rev. 20, 62–78.

Blom, H., Everdij, M. 1993. Embedding adaptive JLQG into LQ martingale with a complete observable stochastic control matrix. In: Proc. 32nd Conf. Decision Control, San Antonio, Texas, pp. 849–854.

Blum, J., Hanson, D., Koopmans, L., 1963. On the strong law of large numbers for a class of stochastic processes. Z. Wahrsch. Verw. Gebiete 2, 1–11.

Boltyanskii, V., Poznyak, A., 1999. Robust maximum principle in minimax control. Int. J. Control 72(4), 305–314.

Boukas, E., Shi, P., Andijani, A., 1999. Robust inventory-production control problem with stochastic demand. Opt. Control Appl. Methods 20(1), 1–20.

Breiman, L., 1992. Probability. In: Classics in Applied Mathematics, vol. 7. SIAM, Philadelphia.

Caines, P., 1988. Linear Stochastic Systems. John Wiley and Sons, New York.

Chen, H., Guo, L., 1991. Identification and Stochastic Adaptive Control. Birkhaüser, Boston.

Cramér, H., 1957. Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, USA.

Cramer, H., Lidbetter, M., 1969. Stationary Random Processes. Princeton University Press, Princeton, NJ, USA.

Davis, M., 1977. Linear Estimation and Stochastic Control. Chapman and Hall, New York.

Devyaterikov, I., Poznyak, A., 1984. Optimization of mechanical systems under uncertainty. MPhTI Publisher, Moscow (in Russian).

Devyaterikov, I., Poznyak, A., 1985. Parameters Estimating in Linear Systems under Noise Perturbations. MPhTI Publisher, Moscow (in Russian).

Doob, J., 1953. Stochastic Processes. Wiley, New York.
Duncan, T., Guo, L., Pasik-Duncan, B., 1999. Adaptive continuous-time linear quadratic gaussian control. IEEE Trans. AC 44(9), 1653–1662.
Duncan, T., Varaiya, P., 1971. On the solution of a stochastic control system. SIAM J. Control 9, 354–337.
Etemadi, N., 1981. An elementary proof of the strong law of large numbers. Z. Wahrsch. Verw. Gebiete 55, 119–122.
Feller, W., 1935. Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung. Math. Z. 40, 521–559.
Fleming, W., Rishel, R., 1975. Optimal Deterministic and Stochastic Control. Springer-Verlag, Berlin.
Gihman, I., Skorohod, A., 1972. Stochastic Differential Equations. Springer-Verlag, New York.
Gut, A., 2005. Probability: A Graduate Course. Springer, New York.
Hall, P., Heyde, C., 1976. On a unified approach to the law of the iterated logarithm for martingales. Bull. Australian Math. Soc. 14, 435–447.
Hall, P., Heyde, C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York, USA.
Hardy, G., Littlewood, J., 1914. Some problems of diophantine approximation. Acta Math. 37, 155–239.
Hartman, P., Wintner, A., 1941. On the law of the iterated logarithm. Am. J. Math. 63, 169–176.
Hausdorff, F., 1913. Grundzuge der Mengenlehre. Veit, Leipzig.
Haussman, U., 1981. Some examples of optimal control, or: The stochastic maximum principle at work. SIAM Rev. 23, 292–307.
Haussman, U., 1982. On the existence of optimal controls for partially observed diffusions. SIAM J. Control 20, 385–407.
Hájek, J., Renyi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Math. Acad. Sci. Hungar. 6, 281–283.
Hoffman-Jörgensen, J., 1974. Sums of independent banach space valued random variables. Studia Math. LII, 159–186.
Hsu, P., Robbins, H., 1947. Complete convergence and the law of large numbers. Proc. Nat. Acad. Sci. USA 33, 25–31.
Huber, P., 1975. Robustness and designs. In: A Survey of Statistical Design and Linear Models. North-Holland Publishing Company.
Ibragimov, I., 1962. Some limit theorems for stationary sequences. Theory Probab. Appl. 7, 349–382.
Iosifecsu, M., Theodorescu, R., 1969. Random Processes and Learning. Springer-Verlag, New York.
Ito, K., 1951. On Stochastic Differential Equations. American Mathematical Society, New York, Memoirs No: 4.
Jain, N., 1975. Tail probabilities for sums of independed banach space valued random variables. Z. Wahrsch. Verw. Gebiete 33, 155–166.
Kahane, J.-P., 1985. Some Random Series of Functions, 2nd edn. Cambridge University Press, Cambridge.
Kalman, R., 1960. A new approach to linear filtering and prediction problem. ASME Trans. Part D: J. Basic Eng. 82, 35–45.

Kaz'min, S., Poznyak, A., 1992. Recurrent estimates of ARX models with noises described by arma processes. Autom. Remote Control 53(10), 1549–1556.
Khan, R., 1974. A probabilistic proof of Stirling's formula. Am. Math. Monthly 81, 366–369.
Khintchine, A., 1924. Uber einen satz der wahrscheinlichkeitsrechnung. Fund. Math. 6, 9–20.
Kiefer, J., Wolfowitz, J., 1952. Stochastic estimation of the maximum of a regression function. Ann. Math. Stat. 23, 462–466.
Kolmogorov, A., 1929. Das gesetz des iterierten logarithmus. Math. Ann. 101, 126–135.
Kolmogorov, A., 1933. Grundbegriffe der wahscheinlichkeitsrechnung. Ergeb. Math. Grenzgeb. 2(3); English translation: Foundations of the Theory of Probability, Chelsea Publishing Co., New York, 1950.
Kushner, H., 1972. Necessary conditions for continuous parameter stochastic optimization problems. SIAM J. Control 10, 550–565.
Kushner, H., Yin, G., 1997. Stochastic Approximation Algorithms and Applications. Springer-Verlag, New York.
Lindeberg, J., 1922. Eine neue herleitung des exponentialgezetzes in der wahrscheinlichkeitsrechnung. Math. Z. 15, 211–225.
Liptser, R.S., Shiryayev, A., 1989. Theory of Martingales. Kluwer Academic Publishers, Dordrecht and Boston.
Ljung, L., 1987. System Identification: Theory for the User. Prentice Hall PTR, New Jersey.
Ljung, L., 1999. System Identification: Theory for the User, 2nd edn. Prentice Hall PTR, New Jersey.
Ljung, L., Soderstrom, T., 1983. Theory and Practice of Recursive Identification. MIT Press, London.
Loéve, M., 1977. Probability Theory, vol. I, 4th edn. Springer-Verlag, Berlin and New York.
Lévy, P., 1937. Th'eorie de l'Addition des Variables Aléatoires. Gauthier-Villars, Paris, (2nd edn, 1954).
Lyapunov, A., 1900. Sur une proposition de la théorie des probabilités. Acad. Sci. St.-Petersbourg 13(5), 211–225.
Marcinkiewicz, J., Zygmund, A., 1937. A sur les fonctions indépendantes. Fund. Math. 29, 60–90.
McLeish, D., 1975. A maximal inequality and dependent strong law. Ann. Probab. 3, 829–839.
Nazin, A., 2001. Stochastic approximation method with averaging. Moscow Aviation Institute, Moscow (in Russian).
Nazin, A., Poznyak, A., 1986. Adaptive choice of variants. In: Theoretical Backgrounds of Technical Cybernetics. Nauka, Moscow (in Russian).
Nevel'son, M., Khas'minski, R., 1972. Stochastic Approximation and Recurrent Estimation. Nauka, Moscow (in Russian).
Nevue, J., 1975. Discrete-parameter Martingales. North-Holland, Amsterdam.
Petrov, V., 1975. Sums of Independent Random Variables. Springer-Verlag, Berlin–NY.
Polyak, B., 1990. New stochastic approximation type procedure. Automat. i Telemekh. (7), 98–107 (in Russian).

Pontryagin, L.S., Boltyansky, V.G., Gamkrelidze, R.V., Mishenko, E.F., 1969. Mathematical Theory of Optimal Processes. Interscience, New York (translated from Russian).

Poznyak, A., 1992. New form or strong large numbers law for dependent sequences and its applications to the analysis of recurrent identification algorithms. Reports of Russian Academy of Sciences, ser. Kibernetica i Teoriya Regulirovaniya 323(4), 645–648 (in Russian).

Poznyak, A., 2001. Strong law of large numbers for dependent vector processes with decreasing correlation: Double averaging concept. Math. Probl. Eng. 7, 87–95.

Poznyak, A., 2002a. Robust stochastic maximum principle: a measured space as uncertainty set. In: Passik-Duncan, B. (Ed.), Stochastic Theory and Control. In: Lecture Notes in Control and Information Sciences, vol. 280. Springer, Berlin–Heidelberg–New York, pp. 385–394.

Poznyak, A., 2002b. Robust stochastic maximum principle: Complete prood and discussions. Math. Probl. Eng. 8(4–5), 389–412.

Poznyak, A., 2008. Advanced Mathematical Tools for Automatic Control Engineering. In: Deterministic Technique, vol. 1. Elsevier.

Poznyak, A., Najim, K., Gomez, E., 2000. Self-Learning Control of Finite Markov Chains. Marcel and Decker, NY.

Poznyak, A.S., 1980. Estimating the parameters of autoregression processes by the least square method. Int. J. System Sci. 5(11), 577–588.

Poznyak, A., Taksar, M., 1996. Robust control of linear stochastic systems with fully observable state. Appl. Math. 24(1), 35–46.

Poznyak, A., Tchikin, D., 1985. Gradient procedures of stochastic approximation with dependent noise and their asymptotic behavior. Int. J. System Sci. 16(8), 917–949.

Poznyak, A., Duncan, T., Pasik-Duncan, B., Boltyansky, V., 2002a. Robust optimal control for minimax stochastic linear quadratic problem. Int. J. Control 75(14), 1054–1065.

Poznyak, A., Duncan, T., Pasik-Duncan, B., Boltyansky, V., 2002b. Robust stochastic maximum principle for multi-model worst case optimization. Int. J. Control 75(13), 1032–1048.

Poznyak, A., Tsypkin, Y., 1989. Generalized instrumental variables method in linear plants identification problems. Docladi of Academy of Sciences of USSR, ser. Kibernetica i Teoriya Regulirovaniya RAN 306(5), 1068–1072.

Protter, P., 1977. On the existence, uniquiness, convergence, and explosions of solutions of stochastic integral equations. Ann. Probab. 5, 243–261.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Ann. Math. Stat. 22, 400–407.

Rosenblatt, M., 1956. A central limit theorem and strong mixing condition. Proc. Natl. Acad. Sci. USA 42, 43–47.

Ruppert, D. 1988. Efficient estimators from a slowly convergent Robbins-Monro process. Tech. report no: 781. School of Operation Reserach and Industrial Engineering, Cornell University, Ithaca, NY.

Sacks, J., 1958. Asymptotic distribution of stochastic approximation procedure. Ann. Math. Stat. 29, 373–405.

Saridis, G., 1977. Self-Organizing Control of Stochastic Systems. Dekker, New York.

Shiryayev, A., 1984. Probability. In: Graduate Texts in Mathematics, vol. 95. Springer-Verlag, New York.

Siegmund, D., Robbins, H.A., 1971. Convergence Theorem for Nonnegative Almost Supermartingales and Some Applications. In: Rustaji, J.S. (Ed.), Optimizing Methods in Statistics, Academic Press, New York, pp. 233–257.
Skorohod, A., 1956. Limit theorems for stochastic processes. Theory Probab. Appl. 1, 261–290.
Soderstrom, T., Stoica, P., 1980. The Instrumental Variables Method for System Identification. Springer-Verlag, Berlin.
Stout, W., 1970a. The Hartman–Wintner law of the iterated logarithm for martingales. Ann. Math. Stat. 41, 2158–2160.
Stout, W., 1970b. A martingale analogue of Kolmogorov's law of the iterated logarithm. Z. Wahrsch. Verw. Gebiete 15, 279–290.
Strassen, V., 1964. An invariant principle for the law of the iteratived logarithm. Z. Wahrsch. Verw. Gebiete 3, 211–226.
Stratonovich, R., 1966. A new representation for stochastic integrals and equations. SIAM J. Control Optim. 4, 362–371.
Taksar, M., Poznyak, A., Iparraguirre, A., 1998. Robust output feedback control for linear stochastic systems in continuous time with time-varying parameter. IEEE Trans. AC 43(8), 1133–1136.
Taksar, M., Zhou, X., 1998. Optimal risk and dividend control for a company with a debt liability. Insurance: Math. Econom. 22, 105–122.
Tsypkin, Y., 1971. Adaptation and Learning and Automatic Systems. Academic Press, New York.
Tsypkin, Y., 1984. Information Theory of Identification. Nauka, Moscow (in Russian).
Tsypkin, Y., Polyak, B. 1977. Robust likelihood method. In: Dynamic Systems: Mathematical Methods of Oscillation Theory, vol. 12, Gor'kii State University, pp. 22–46.
Ugrinovskii, V., Petersen, I., 1997. Finite horizon minimax optimal control of nonlinear continuous time systems with stochastic uncertainties. Dynamics Control 10, 63–86.
Ville, J., 1939. Étude Critique de la Notion de Collectif (Monograhie des Probabilités). Gauthier-Villars, Paris.
Wald, A., 1947. Sequential Analysis. Wiley, New York.
Wasan, M., 1969. Stochastic Approximation. Cambridge University Press, London and New York.
Whittle, G., 1984. Probability Theory. Chapman and Hall, London.
Williams, D., 1991. Probability with Martingales. Cambridge University Press, Cambridge.
Yaz, G. 1991. Minimax control of discrete nonlinear stochastic systems with white noise uncertainty. In: Proc. 30th Conf. Decision Control, Brighton, England, pp. 1815–1816.
Yong, J., Zhou, X., 1999. Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer-Verlag.
Yoshida, K., 1979. Functional Analysis. Narosa Publishing House, New Delhi.
Zhou, X.Y., 1991. A unified treatment of maximum principle and dynamic programming in stochastic controls. Statist. Stoch. Rep. 36, 137–161.

This page intentionally left blank

Index