Lecture Notes in Statistics 149

Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth, S. Zeger

Springer Science+Business Media, LLC

D. Bosq

Linear Processes in Function Spaces

Theory and Applications

Springer

D. Bosq
Université Pierre et Marie Curie
Institut de Statistique
4 Place Jussieu
75252 Paris Cedex 05
France

Library of Congress Cataloging-in-Publication Data

Bosq, Denis, 1939-
Linear processes in function spaces : theory and applications / D. Bosq.
p. cm. -- (Lecture notes in statistics ; 149)
Includes bibliographical references and index.
ISBN 978-0-387-95052-5 ISBN 978-1-4612-1154-9 (eBook)
DOI 10.1007/978-1-4612-1154-9
1. Function spaces. 2. Stochastic processes. I. Title. II. Lecture notes in statistics (Springer-Verlag) ; v. 149.

QA323 .B67 2000
515'.73--dc21

00-040041

Printed on acid-free paper.

Camera ready copy provided by the author.

9 8 7 6 5 4 3 2 1

ISBN 978-0-387-95052-5

to my father
to Agnès
to Françoise and Philippe Gaillot

Ainsi l'abbé Blanès n'avait pas communiqué sa science assez difficile à Fabrice; mais à son insu, il lui avait inoculé une confiance illimitée dans les signes qui peuvent prédire l'avenir.

If you can look into the seeds of time,
And say which grain will grow and which will not

Preface

Representation of continuous time stochastic processes as random variables in function spaces is an efficient tool in probability and statistics.

The main purpose of this book is to study linear processes in Hilbert and Banach spaces, keeping in mind the above representation and applications to statistical prediction over a whole time interval.

The reader interested in these applications may skim through Chapters 1 and 2, and focus on Chapters 3, 4, 8 (except Section 8.6), and 9. These parts have been used in a one-semester course for Ph.D. students.

People who are not allergic to theory may carry out a more detailed reading of the book.

In order to facilitate various approaches, this work contains some intentional repetitions. It also probably contains some (unintentional!) errors, and I thank in advance the readers who will point out them to me.

My thanks go to Alain Berlinet, Delphine Blanke, and Serge Guillas who read portions of this book and made helpful suggestions.

I am also indebted to Florence Merlevède for her kind participation in the writing of Chapter 7.

I have used recent or unpublished results by Philippe Besse, Hervé Cardot and David Stephenson; André Mas; Besnik Pumo; Pierre Bernard; and Andréa Cavallini, Gian Carlo Montanari, Mauro Loggini, Oliviero Lessi,† and Mario Cacciari. I also thank them.

I acknowledge Bernard Maurey and Joan Seignour for fruitful discussions and suggestions.

My greatest debt is to Dominique Guyard, who prepared the typescript with patience, care, and efficiency.

I thank the editors, especially John Kimmel, for accepting this work for their Lecture Notes in Statistics series. Finally, I thank Tony Orrantia for his very careful rereading.

Denis Bosq
Paris, January 2000

Contents

Preface vi

Notation xi

Synopsis 1

1. The object of study 1
2. Finite-dimensional linear processes 3
3. Random variables in function spaces 5
4. Limit theorems in function spaces 7
5. Autoregressive processes in Hilbert spaces 7
6. Estimation of covariance operators 8
7. Autoregressive processes in Banach spaces and representations of continuous-time processes 10
8. Linear processes in Hilbert spaces and Banach spaces 11
9. Estimation of autocorrelation operator and forecasting 12
10. Applications 13

1. Stochastic processes and random variables in function spaces 15

1.1. Stochastic processes 15

1.2. Random functions 21

1.3. Expectation and conditional expectation in Banach spaces 27

1.4. Covariance operators and characteristic functionals in Banach spaces 30

1.5. Random variables and operators in Hilbert spaces 33

1.6. Linear prediction in Hilbert spaces 38

Notes 42

2. Sequences of random variables in Banach spaces 43

2.1. Stochastic processes as sequences of B-valued random variables 43

2.2. Convergence of B-random variables 44

2.3. Limit theorems for i.i.d. sequences of B-random variables 47

2.4. Sequences of dependent random variables in Banach spaces 54

2.5. * Derivation of exponential bounds 66

Notes 70

3. Autoregressive Hilbertian processes of order one 71

3.1. Stationarity and innovation in Hilbert spaces 71

3.2. The ARH(1) model 73

3.3. Basic properties of ARH(1) processes 79

3.4. ARH(1) processes with symmetric compact autocorrelation operator 82

3.5. Limit theorems for ARH(1) processes 86

Notes 94

4. Estimation of autocovariance operators for ARH(1) processes 95

4.1. Estimation of the covariance operator 95

4.2. Estimation of the eigenelements of C 102

4.3. Estimation of the cross-covariance operators 112

4.4. Limits in distribution 118

Notes 125

5. Autoregressive Hilbertian processes of order p 127

5.1. The $ARH(p)$ model 127

5.2. Second order moments of $ARH(p)$ 133

5.3. Limit theorems for $ARH(p)$ processes 136

5.4. Estimation of autocovariance of an $ARH(p)$ 140

5.5. Estimation of the autoregression order 143

Notes 145

6. Autoregressive processes in Banach spaces 147

1. Strong autoregressive processes in Banach spaces 147

2. Autoregressive representation of some real continuous-time processes 150

3. Limit theorems 153

4. Weak Banach autoregressive processes 161

5. Estimation of autocovariance 164

6. The case of $C[0,1]$ 168

7. Some applications to real continuous-time processes 175

Notes 180

7. General linear processes in function spaces 181

7.1. Existence and first properties of linear processes 182

7.2. Invertibility of linear processes 184

7.3. Markovian representations of LPH: applications 188

7.4. Limit theorems for LPB and LPH 191

7.5. * Derivation of invertibility 195

Notes 202

8. Estimation of autocorrelation operator and prediction 203

8.1. Estimation of ρ if H is finite dimensional 204

8.2. Estimation of ρ in a special case 211

8.3. The general situation 218

8.4. Estimation of autocorrelation operator in $C[0,1]$ 222

8.5. Statistical prediction 226

8.6. * Derivation of strong consistency 229

Notes 236

9. Implementation of functional autoregressive predictors and numerical applications 237

9.1. Functional data 237

9.2. Choosing and estimating a model 240

9.3. Statistical methods of prediction 243

9.4. Some numerical applications 247

Notes 251

Figures 252

Appendix 263

1. Measure and probability 263

2. Random variables 264

3. Function spaces 265

4. Basic function spaces 266

5. Conditional expectation 267

6. Stochastic integral 267

References 269

Index 277

Notation

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ sets of natural numbers, integers, rational numbers, real numbers, complex numbers

$[a, b]$, $]a, b[$ closed interval, open interval

$\mathring{A}, \bar{A}$, A^c interior, closure and complement of A

$\bigcup_{i \in I} A_i$, $\bigcap_{i \in I} A_i$ union, intersection of $(A_i, i \in I)$

$A \times B$ cartesian product of A and B

$(\Omega, \mathcal{A}, P)$ probability space: Ω non empty set, $\mathcal{A}$ σ-algebra of subsets of Ω, P probability measure on $\mathcal{A}$

$\mathcal{B}_E$ σ-algebra of Borel sets on the metric space E

$\sigma(X_i,\ i \in I)$ σ-algebra generated by X_i, $i \in I$

i.i.d. independent and identically distributed random variables

EX, VX, C_X, P_X expectation, variance, covariance operator, distribution (of X)

$\mathrm{Cov}(X, Y)$, $\mathrm{Corr}(X, Y)$ covariance, correlation coefficient of X and Y

$C_{X,Y}$, $C_{Y,X}$ Cross-covariance operators of X and Y

$E^{\mathcal{A}_0}(X)$ or $E(X|\mathcal{A}_0)$ conditional expectation of X with respect to the sub-σ algebra $\mathcal{A}_0$

$\delta_{(a)}, \mathcal{N}(m, \sigma^2), \lambda$ Dirac measure at a, normal distribution, Lebesgue measure

$\mathcal{N}(m, C)$ Normal distribution with expectation m and covariance operator C

$(\xi_t,\ t \in I)$ or (ξ_t) real stochastic process

$(X_n,\ n \in \mathbb{Z})$ or (X_n) stochastic process in a function space

H, B, B^* (real) Hilbert space, Banach space, (topological) dual of B

$H_1 \oplus H_2$, $H_1 \otimes H_2$ direct sum, tensor product of Hilbert spaces

$\|\cdot\|$; $<\cdot,\cdot>$ norm on H, B or B^*; scalar product on H

$\mathrm{sp}\{x_i,\ i \in I\}$ linear space generated by $\{x_i,\ i \in I\}$

$\overline{\mathrm{sp}}\{x_i,\ i \in I\}$ closure of $\mathrm{sp}\{x_i,\ i \in I\}$ with respect to some topology

$\mathcal{L}(B, B')$ space of continuous linear operators from the Banach space B to the Banach space B'

$\mathcal{L} = \mathcal{L}(H, H)$ or $\mathcal{L}(B, B)$

ℓ^* adjoint of the linear operator ℓ

$e \otimes f$ element of $\mathcal{L}(H, H)$ defined by $(e \otimes f)(x) =< e, x > f$, $x \in H$, where e and f are fixed elements in H

$\mathcal{C}$ class of compact operators over H

$\mathcal{S}$ space of Hilbert-Schmidt operators over H

$\mathcal{N}$ space of nuclear operators over H

$\|\cdot\|_{\mathcal{S}}$ Hilbert-Schmidt norm

$\|\cdot\|_{\mathcal{N}}$ nuclear norm

$L^p_B(\Omega, \mathcal{A}, P)$, $1 \le p \le \infty$ space of (classes of) B-valued random variables X such that $\| X \|_{p,B} = \left(\int_B \| X \|^p \, dP \right)^{1/p} < \infty$, $1 \le p < \infty$, $\| X \|_{\infty,B} = \inf\{c : P(\| X \| > c) = 0\} < \infty$, $p = \infty$

$L^p := L^p_{\mathbb{R}}$, $1 \le p \le \infty$

$C_k[a, b]$, $k \ge 0$ Banach space of real function defined on $[a, b]$ with k continuous derivatives and equipped with the norm

$$\| x \| = \sum_{i=0}^{k} \sup_{a \le t \le b} |x^{(\ell)}(t)|$$

$\mathbf{1}_A$ indicator of $A : \mathbf{1}_A(x) = 1$, $x \in A$; $(1_A(x) = 0$, $x \notin A$

$Log\ x$ logarithm of x

$\mathrm{Log}_k x$ defined recursively by $\mathrm{Log}_k x = \mathrm{Log}(\mathrm{Log}_{k-1} x)$ if $x \ge e$; $\mathrm{Log}_k(x) = 1$ if $x < e$; $k \ge 2$

$[x]$ integer part of x

$x_+ = \max(0, x)$

$x_- = \max(0, -x)$

$u_n \sim v_n \quad \dfrac{u_n}{v_n} \to 1$

$u_n \simeq v_n$ there exist constants c_1 and c_2 such that $0 < c_1 v_n < u_n < c_2 v_n$ for large enough n

$u_n = o(v_n) \quad \dfrac{u_n}{v_n} \to 0$

$u_n = O(v_n)$ There is a $c > 0$ such that $u_n \le c v_n$, $n \ge 1$.

$\xrightarrow{\mathcal{D}}$, $\xrightarrow{w}$ convergence in distribution, weak convergence

$\xrightarrow[p]{}$ convergence in probability

$\xrightarrow[a.s.]{}$ or $\longrightarrow$ $a.s.$ almost sure convergence

■ End of proof

* section that may be omitted

$\perp$ weakly orthogonal

$\perp\!\!\!\!\perp$ orthogonal

$\perp\!\!\!\perp$ independent

Synopsis

S.1 The object of study

The final goal of this book is statistics of continuous-time processes.

Due to the fact that a stochastic process has the form $\xi(t, \omega)$, where t is time and ω is random, approach of this topic is twofold.

The local approach simply consists in considering the observed variables as a collection of data and then applying classical statistical techniques.

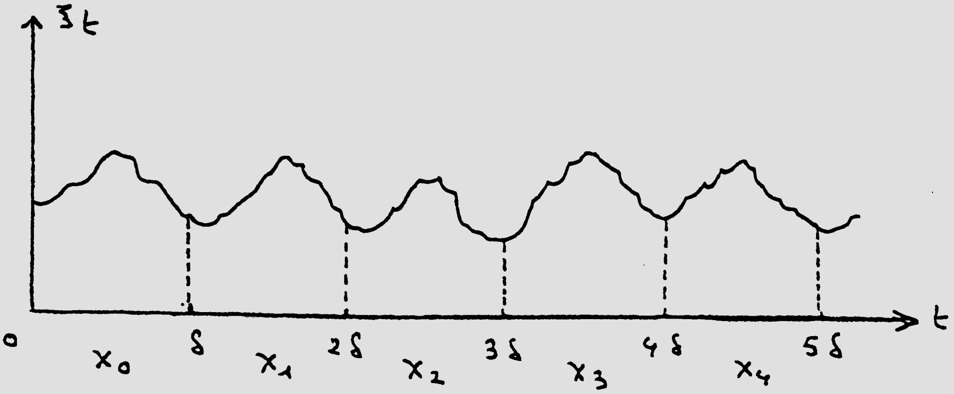
The global approach emphasizes interpretation of ξ as a random function, that is, a random variable that takes its values in some function space. This well-known interpretation is commonly used in probability theory.

However, it is not convenient for asymptotic inference even if one may express increase of time by defining a suitable "filtration." This purely theoretical method is not easy to handle.

In many cases a more convenient method consists in associating with $\xi = (\xi_t,\ t \in \mathbb{R})$ a sequence of random variables with values in an appropriate function space. For example, this may be obtained by setting

$$X_n(t) = \xi_{nh+t},\ 0 \le t \le h\ ,\ n \in \mathbb{Z},$$

so $X = (X_n,\ n \in \mathbb{Z})$ is an infinite-dimensional discrete-time process (see figure 1).



Representation of a continuous-time process as a sequence of F-random variables

Figure 1

This representation is especially fruitful if ξ possesses a seasonal component with period h and/or if one intends to forecast the future behavior of ξ over a time interval of length h (see figure 2 below). It can also be employed if data are curves defined on the same interval.

Starting from this idea, it is natural to study discrete time F-valued processes, where F is a function space. Curiously, in almost all literature on the subject, this theory is developed only in the special framework of independent equidistributed random variables. This is of course a very important and interesting case, but it is not suitable for many applications and makes no sense if one is interested in forecasting.

In this book we intend to develop theory of discrete time F-valued processes by studying the reasonable case where X is a linear process and F is a Hilbert or Banach space. In this context, the genuine **functional autoregressive process of order 1** plays a central role and will be carefully analyzed because it is used in practice (see Chapter 9) and provides efficient tools for studying more general models like autoregressive processes of order p and general linear processes.

Before describing the different parts of this book, let us recall some basic facts concerning linear processes in a finite-dimensional setting, i.e., in $\mathbb{R}^d$, $d \geq 1$.

S.2 Finite-dimensional linear processes

We focus on the **one-dimensional case**. Let $X = (X_n,\ n \in \mathbb{Z})$ be a sequence of square integrable real random variables defined on a probability space $(\Omega, \mathcal{A}, P)$. X is a (weakly) **stationary** process if EX_n does not depend on n and

$$\operatorname{Cov}(X_n,\ X_m) = \operatorname{Cov}(X_0,\ X_{m-n})\ ;\ n, m \in \mathbb{Z}.$$

A **white noise** $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ is a zero-mean stationary process such that

$$E(\varepsilon_n \varepsilon_m) = \sigma^2 \delta_{n,m}\ \ ;\ \ n, m \in \mathbb{Z},$$

where $\sigma^2 > 0$, and $\delta_{n,m} = 0$ if $n \neq m$ and 1 if $n = m$.

Given a zero-mean stationary process, one defines the spaces

$$\mathcal{M}_n = \overline{\operatorname{sp}}(X_i,\ i \leq n),\ \ n \in \mathbb{Z},$$

where closure is taken in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$. Let $\pi^{\mathcal{M}_n}$ be the orthogonal projector of $\mathcal{M}_n$. X is said to be **regular** if

$$\sigma^2 := E\left(X_n - \pi^{\mathcal{M}_{n-1}}(X_n)\right)^2 > 0,\ \ n \in \mathbb{Z}.$$

If X is regular it admits the so-called **Wold decomposition**

$$X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j} + Y_n,\ \ n \in \mathbb{Z},$$

where the series converges in $L^2(\Omega, \mathcal{A}, P)$ and

$$\varepsilon_n = X_n - \pi^{\mathcal{M}_{n-1}}(X_n),\ \ n \in \mathbb{Z}$$

is a white noise called **innovation** (process) of X; $(a_j,\ j \geq 1)$ is a real sequence with $a_0 = 1$, $\sum_{j=1}^{\infty} a_j^2 < \infty$. $Y = (Y_n,\ n \in \mathbb{Z})$ is the singular part of the decomposition in the sense that $E(Y_n \varepsilon_m) = 0$; $n, m \in \mathbb{Z}$ and $Y_m \in \bigcap_{n=-\infty}^{+\infty} \mathcal{M}_n$, $m \in \mathbb{Z}$. Thus Y_m is completely specified by observation of X until time n, for any n.

This remark leads to limit oneself to the study of **linear processes**, that is, processes of the form

$$X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j},\ n \in \mathbb{Z}, \tag{1}$$

where (ε_n) is a white noise and (a_j) is as above.

If in addition $\varepsilon_n \in \mathcal{M}_n$, $n \in \mathbb{Z}$ then (1) is the Wold decomposition of X. In that situation X is said to be **invertible** and we have

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j X_{n-j}, \quad n \in \mathbb{Z}, \tag{2}$$

with $\sum_{j=1}^{\infty} \rho_j^2 < \infty$.

This model is difficult to use in statistics, since it involves an infinite number of parameters. Thus it is preferable to employ a truncated version of (1) or to use (2).

Truncation of (1) gives the **moving average** process of order q , $MA(q)$, defined as

$$X_n = \sum_{j=0}^{q} a_j \varepsilon_{n-j}, \quad n \in \mathbb{Z},$$

where $a_0 = 1$ and $a_q \neq 0$. It is invertible if $1 + \sum_{j=1}^{q} a_j j^j \neq 0$ for $|z| \leq 1$.

By using (2) one obtains the **autoregressive** process of order p, $AR(p)$:

$$X_n = \varepsilon_n + \sum_{j=1}^{p} \rho_j X_{n-j}, \quad n \in \mathbb{Z},$$

where $\rho_p \neq 0$. It is a linear process if $1 - \sum_{j=1}^{p} \rho_j z^j \neq 0$ for $|z| \leq 1$.

If $p = 1$, one gets the simple $AR(1)$ model

$$X_n = \rho X_{n-1} + \varepsilon_n, \quad n \in \mathbb{Z},$$

and the above condition reduces to $|\rho| < 1$.

If $|\rho| \geq 1$ this equation has no stationary solution with innovation (ε_n), since existence of such a solution would imply $VX_n = \rho^2 VX_{n-1} + \sigma^2$ and hence $VX_0(1 - \rho^2) = \sigma^2 > 0$ which is a contradiction.

In practice one may use the $ARMA(p,q)$ model which combines the AR and MA models. More generally, one may employ the $SARIMA$ model, which includes trend and seasonal component (see Brockwell and Davis (1991)).

In the **multidimensional case** a linear process has the form

$$X_n = \sum_{j=1}^{\infty} a_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z},$$

where $(a_j,\ j \geq 1)$ is a sequence of matrices whose components are absolutely summable and (ε_n) is a multivariate white noise.

In particular, a zero-mean autoregressive process of order 1 in $\mathbb{R}^d$ satisfies

$$X_n = \rho(X_{n-1}) + \varepsilon_n,$$

where (ε_n) is a white noise in $\mathbb{R}^d$ and ρ is a $d \times d$ matrix such that $\det(I - z\rho) \neq 0$ for all complex numbers z such that $|z| \leq 1$.

We do not give more details, since almost all definitions and results in the infinite-dimensional case that appear in the sequel will directly apply to the multidimensional case.

S.3 Random variables in function spaces

Consider a continuous-time real stochastic process $\xi = (\xi_t,\ t \in T)$, where T is an interval in $\mathbb{R}$, defined on a probability space $(\Omega, \mathcal{A}, P)$ and such that $t \mapsto \xi_t(\omega)$ belongs to some function space F for all ω in Ω. Then, under simple conditions, ξ defines a ***F*-valued random variable**. Such conditions are studied in chapter 1 in the case where F is a separable Banach or Hilbert space.

Basic properties of random variables in function spaces are considered in Sections 1.3, 1.4, and 1.5. **Expectation** and **conditional expectation** do not pose major problems. On the contrary, new questions arise concerning **moments of order 2** because these moments are defined in terms of linear operators.

Classification of linear operators in a function space is much more intricate than that in a finite-dimensional context. Consider, for example, a separable Hilbert space H with norm $\|\cdot\|$ and scalar product $<\cdot,\cdot>$. Three classes of linear operators over H are particularly interesting for our purpose.

The first is the class $\mathcal{L}$ of bounded linear operators with norm

$$\| \ell \|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \| \ell(x) \|, \quad \ell \in \mathcal{L}.$$

The second is the class $\mathcal{S}$ of **Hilbert-Schmidt operators**. A linear operator s is Hilbert-Schmidt if

$$s(x) = \sum_{j=1}^{\infty} \lambda_j < x, e_j > f_j, \quad x \in H, \tag{3}$$

where (e_j) and (f_j) are orthonormal bases of H and (λ_j) is a real sequence such that $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$. $\mathcal{S}$, equipped with the norm $\| s \|_{\mathcal{S}} = \left(\sum_1^{\infty} \lambda_j^2 \right)^{1/2}$, is a Hilbert space.

Finally, an operator s is said to be **nuclear** if (3) holds with $\sum_{j=1}^{\infty} |\lambda_j| < \infty$. The class $\mathcal{N}$ of nuclear operators is a Banach space with respect to the norm $\|\cdot\|_{\mathcal{N}} = \sum_{j=1}^{\infty} |\lambda_j|$.

Some details concerning these classes of operators appear in Section 1.5.

Now let X be a zero-mean H-valued random variable such that $E \| X \|^2 < \infty$. Its **covariance operator** is defined as

$$C_X(x) = E(< X, x > X), \quad x \in H.$$

If Y is a random variable of the same type, the **cross-covariance operators** of X and Y are defined as

$$C_{X,Y}(x) = E(< X, x > Y), \quad x \in H,$$

and

$$C_{Y,X}(x) = E(< Y, x > X), \quad x \in H.$$

These operators are nuclear and therefore Hilbert-Schmidt. In particular, C_X admits a decomposition of the form

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j < x, v_j > v_j, \quad x \in H, \tag{4}$$

where (v_j) is an orthonormal basis of H and (λ_j) is a positive sequence such that $\sum_{j=1}^{\infty} \lambda_j = E \| X \|^2 < \infty$. Clearly (4) is the diagonal form of C_X and (λ_j, v_j), $j \geq 1$ is a complete sequence of eigenelements of it. The above properties of C_X are extensively used in this book.

Another question that has a typical "infinite-dimensional solution" is **linear approximation**: consider the Hilbert space $L^2_H(\Omega, \mathcal{A}, P)$ of (classes of) H-valued random variables X such that $E \|X\|^2 < \infty$, equipped with the scalar product $E < X, Y >$; $X, Y \in L^2_H(\Omega, \mathcal{A}, P)$.

A natural notion of orthogonality in such a space is given by

$$X \perp Y \Leftrightarrow E < X, Y >= 0.$$

In fact this notion is too weak and must be replaced by the stronger

$$X \perp\!\!\!\perp Y \Leftrightarrow C_{X,Y} = 0.$$

Linked with this definition is the notion of **$\mathcal{L}$-closed subspace (LCS)** of $L^2_H(\Omega, \mathcal{A}, P)$. A subset $\mathcal{G}$ of $L^2_H(\Omega, \mathcal{A}, P)$ is an LCS if

- $\mathcal{G}$ is an Hilbertian subspace of $L^2_H(\Omega, \mathcal{A}, P)$ and
- $Z \in \mathcal{G},\ \ell \in \mathcal{L} \Rightarrow \ell(Z) \in \mathcal{L}$.

We are now in a position to define linear approximation. Let $\mathcal{X} = (X_i,\ i \in I)$ be a family of zero-mean random variables in $L^2_H(\Omega, \mathcal{A}, P)$. The best linear approximation of X given $(X_i,\ i \in I)$ is the orthogonal projection $\pi^{\mathcal{G}_\mathcal{X}}(X)$ of X over the LCS $\mathcal{G}_\mathcal{X}$ generated by $(X_i,\ i \in I)$.

We then have $X - \pi^{\mathcal{G}_\mathcal{X}}(X) \perp \mathcal{G}_\mathcal{X}$ but owing to special properties of LCS we also get $X - \pi^{\mathcal{G}_\mathcal{X}}(X) \underline{\perp} \mathcal{G}_\mathcal{X}$.

Note that this scheme also applies to the finite-dimensional situation (the two notions of orthogonality coincide only if $\dim H = 1$). Some properties of LCS's are presented in section 1.6.

S.4 Limit theorems in function spaces

Definitions and properties of stochastic convergences in a separable Banach space B appear in Section 2.2, while, classical limit theorems for i.i.d. sequences of B-valued random variables are presented in Section 2.3.

If B is a Hilbert space, rates of convergence may be obtained via large deviation inequalities and Berry-Esseen bound. In a general Banach space the situation is more awkward since "addition of variances" does not work. Then rates of convergence depend on some geometrical properties of B, which are described by types and cotypes (see Definition 2.3).

Extensions to dependent variables are considered in Section 2.4. In particular, various limit theorems concerning strong mixing processes and martingale differences are presented. These results are used in the following chapters to derive limit theorems for B-valued linear processes.

S.5 Autoregressive processes in Hilbert spaces

Let H be a separable Hilbert space with norm $\|\cdot\|$ and scalar product $< \cdot, \cdot >$.

Stationary process and white noise in such a space are easily defined by using cross-covariance operators instead of covariances. Now a stationary H-valued process $X = (X_n,\ n \in \mathbb{Z})$ is an autoregressive process of order 1 ($ARH(1)$) if it satisfies

$$X_n - \mu = \rho(X_{n-1} - \mu) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{5}$$

where $\mu \in H$, ρ is a bounded linear operator and (ε_n) is an H-white noise.

Existence of such a process is ensured by the mild condition

(c_0) There exists an integer $j_0 \geq 1$ such that $\| \rho^{j_0} \|_\mathcal{L} < 1$, which implies the apparently stronger

(c_1) There exist $a > 0$ and $0 < b < 1$ such that $\| \rho^j \|_{\mathcal{L}} \leq ab^j, \ j \geq 0$.

If (c_0) holds, then (5) has a unique stationary solution given by

$$X_n = \mu + \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z},$$

where the series converges in $L^2_H(\Omega, \mathcal{A}, P)$ and with probability 1. As a consequence, $\rho(X_{n-1} - \mu)$ is the orthogonal projection of $X_n - \mu$ over the LCS generated by $(X_i - \mu, \ i \leq n-1)$ therefore (ε_n) is the innovation process of $(X_n - \mu)$.

Properties of $ARH(1)$ are studied in Chapter 3. In particular, limit theorems with optimal rates are derived. Suppose for convenience that X is zero-mean and set $S_n = X_1 + \ldots + X_n$, $n \geq 1$. Then

$$\| n \, C_{S_n/n} - \sum_{h=-\infty}^{+\infty} C_{X_0, X_h} \|_{\mathcal{N}} \underset{n \to \infty}{\longrightarrow} 0 \, ,$$

hence

$$n \, E \, \| \frac{S_n}{n} \|^2 \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < X_0, X_h > .$$

Now if (ε_n) is i.i.d. and if $E\left(e^{\gamma \|X_0\|}\right) < \infty$ for some $\gamma > 0$, we have

$$P\left(\| \frac{S_n}{n} \| \geq \eta\right) \leq 4 \exp\left(-\frac{n\eta^2}{\alpha_0 + \beta_0 \eta}\right), \quad \eta > 0, \tag{6}$$

where α_0 and β_0 are positive and depend only on ρ and the distribution of ε_0.

Finally, $\dfrac{S_n}{\sqrt{n}}$ converges in distribution to a Gaussian random variable and a Berry-Esseen bound holds.

The above results easily extend to $ARH(p)$. This is because, if $X = (X_n, \ n \in \mathbb{Z})$ is $ARH(p)$, then $Y_n = (X_n, \ldots, X_{n-p+1})$, $n \in \mathbb{Z}$, is $ARH^p(1)$. Details appear in Chapter 5. Here the new question that arises is estimation of p. This problem is studied in a special case in Section 5.5 while a general empirical method is proposed in Chapter 9.

S.6 Estimation of covariance operators

Let $X = (X_n, \ n \in \mathbb{Z})$ be a zero-mean $ARH(1)$ and let C be the covariance operator of X_0.

The natural estimator of C is the **empirical covariance operator**, defined as

$$C_n(x) = \frac{1}{n}\sum_{i=1}^{n} < X_i, x > X_i, \quad x \in H.$$

The main tool for studying C_n is the following autoregressive representation. Set

$$Z_i(x) =< X_i, x > X_i - C(x), \ x \in H, \ i \in \mathbb{Z}.$$

Then, under mild conditions, $Z = (Z_i, \ i \in \mathbb{Z})$ is an autoregressive process of order 1 in the Hilbert space $\mathcal{S}$ of Hilbert-Schmidt operators, and the associated white noise is a martingale difference. Now noting that

$$C_n - C = \frac{Z_1 + \ldots + Z_n}{n},$$

one may use this $ARS(1)$ representation to derive sharp asymptotic results for $C_n - C$. This is performed in Chapter 4. Some proofs combine autoregressive representation with asymptotic, properties of martingale differences given in Chapter 2.

An important application is estimation of the eigenelements of C.

These are defined by

$$C(v_j) = \lambda_j v_j, \quad j \geq 1,$$

where (v_j) is an orthonormal basis of H and $\lambda_j \downarrow 0$. Similarly, empirical eigenelements are given by

$$C(v_{jn}) = \lambda_{jn} v_{jn}, \quad j \geq 1.$$

Convergence of empirical eigenelements are based on the following inequalities:

$$\sup_{j\geq 1} |\lambda_{jn} - \lambda_j| \leq \| C_n - C \|_{\mathcal{L}}$$

and

$$\| v_{jn} - v'_j \| \leq a_j \| C_n - C \|_{\mathcal{L}}, \quad j \geq 1,$$

where (a_j) is a sequence that depends on (λ_j) and

$$v'_j = \operatorname{sgn} < v_{jn}, v_j > v_j.$$

As a by-product one may estimate dimensions of eigensubspaces.

Now let D be the cross-covariance operator C_{X_0,X_1}. As above, a natural estimator of D is the **empirical cross-covariance operator** given by

$$D_n(x) = \frac{1}{n-1}\sum_{i=1}^{n-1} < X_i, x > X_{i+1}, \quad x \in H.$$

Here a "quasi-autoregressive representation" allows us to obtain the desired asymptotic results concerning $D_n - D$.

Finally, it is shown that $\sqrt{n}(C_n - C, \ D_n - D)$ converges in distribution toward an $\mathcal{S} \times \mathcal{S}$-valued Gaussian random operator. A consequence is asymptotic normality of $\sqrt{n}(\lambda_{jn} - \lambda_j)$ and $\sqrt{n}\ (v_{jn} - v'_j)$.

S.7 Autoregressive processes in Banach spaces and representations of continuous time processes

In practice, observation of processes with continuous or differentiable sample paths is common. In order to construct an associated random model, it is then natural to use Banach spaces whose elements are regular functions, instead of general Hilbert spaces. The counterpart of such a choice is complexity due to weakness of the geometrical properties of Banach spaces. An $ARB(1)$ process is defined similarly to an $ARH(1)$ (see Chapter 6).

Various continuous-time real processes have $ARB(1)$ representations in appropriate Banach spaces. A simple and illuminating example is the well-known Ornstein-Uhlenbeck process, defined as

$$\xi_t = \int_{-\infty}^{t} e^{-\theta(t-s)} dW(s), \quad t \in \mathbb{R},$$

where θ is a positive parameter and W is a bilateral standard Wiener process. It is easy to see that representation (5) holds in $C[0,h]$ (where $h > 0$ is fixed but arbitrary) under the form

$$X_n = \rho_\theta(X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z},$$

where ρ_θ is defined by

$$\rho_\theta(x)(t) = e^{-\theta t} x(h), \quad 0 \le t \le h \ , \ x \in C[0,h],$$

and (ε_n) is given by

$$\varepsilon_n(t) = \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW(s), \quad 0 \le t \le h, \ n \in \mathbb{Z}.$$

Here we have $\| \rho_\theta^j \|_{\mathcal{L}} = e^{-\theta(j-1)h}, \ j \ge 1$, thus (c_0) holds with $j_0 = 2$.

More generally, consider the Banach space $B = C_{q-1}[0,h]$ of real functions on $[0,h]$ having $q-1$ continuous derivatives $(q \ge 1)$, and define the process

$$\eta_t = \mu(t) + \xi_t, \quad t \in \mathbb{R},$$

where $\mu(\cdot)$ is a non constant real function with period h, whose restriction to $[0,h]$ belongs to B and where (ξ_t) is a stationary process solution of a stochastic differential equation of the form

$$\sum_{\ell=0}^{q} a_\ell d\xi^{(\ell)}(t) = dW_t,$$

where $a_0, \dots, a_q \ne 0$ are real coefficients.

Then (η_t) has an $ARB(1)$ representation, although it is not stationary (see Examples 6.3 and 6.4).

In the above examples the range of ρ is finite dimensional. This shows the **nonparametric** character of the $ARB(1)$ model, since it includes much more general situations than those considered above.

Limit theorems for $ARB(1)$ are similar to those for $ARH(1)$, provided (ε_n) is a strong white noise and B has type 2 (see Section 6.3).

Estimation of covariance operators is again based on a special autoregressive representation (Lemma 6.8). In particular, returning to Hilbert spaces, one may use autoregressive interpretation of $C_n - C$ in the Banach space of nuclear operators, to obtain $\| C_n - C \|_{\mathcal{N}} \underset{a.s.}{\longrightarrow} 0$ and consequently $\frac{1}{n}\sum_{i=1}^{n} \| X_i \|^2 \underset{a.s.}{\longrightarrow} E \| X_0 \|^2$.

The special case of $C_0 := C[0,1]$ is of course very important. In a general setting, results concerning $ARC_0(1)$ processes are poor, since C_0 does not possess good geometrical properties. However, if sample paths satisfy a Hölder condition, results become sharper. This condition also provides bounds that measure quality of approximation when one interpolates discrete data in order to reconstruct partially observed sample paths (Section 6.6).

S.8 Linear processes in Hilbert spaces and Banach spaces

A general linear process in a separable Banach space has the form

$$X_n = \mu + \varepsilon_n + \sum_{j=1}^{\infty} a_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{7}$$

where $\mu \in B$, (ε_n) is a white noise in B and (a_j) is a sequence of bounded linear operators over B. If $\sum_{j=1}^{\infty} \| a_j \|_{\mathcal{L}} < \infty$, the series in (7) converges with probability 1.

The first crucial problem is **invertibility**, since this property allows to consider (ε_n) as the innovation process of (X_n).

Under some mild conditions, a linear process in a Hilbert space (LPH) is invertible:

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j(X_{n-j}), \quad n \in \mathbb{Z},$$

with $\sum_{j=1}^{\infty} \| \rho_j \|_{\mathcal{L}} < \infty$ (Theorem 7.2).

The second question is **Markovian representation**: Let H^∞ be the cartesian product of a countable number of copies of H and let $w = (w_i,\ i \geq 1)$ be a sequence of strictly positive weights such that $\sum_{i=1}^{\infty} w_i < \infty$. Then the subspace H_w of H^∞ of those $x = (x_i,\ i \geq 1)$ such that $\sum_{i=1}^{\infty} w_i \| x_i \|^2 < \infty$ is a Hilbert space.

Now if (X_n) is an invertible LPH then, under some regularity conditions, $Y_n = (X_n,\ X_{n-1}, \ldots)$, $n \in \mathbb{Z}$, is an $ARH_w(1)$ (Theorem 7.3). One may employ this representation to obtain precise asymptotic and large deviation results for (X_n) by using the corresponding ones in Chapters 3 and 4. For example, inequality (6) holds for LPH (Theorem 7.5).

Finally, limit theorems can be obtained for linear processes in Banach spaces (see the end of Section 7.4).

S.9 Estimation of autocorrelation operator and forecasting

Identification of a zero-mean $ARH(1)$ requires estimation of the autocorrelation operator ρ.

In a **finite-dimensional** context one may use the relation $D = \rho C$, which yields $\rho = DC^{-1}$ provided invertibility of C. Then a natural estimator of ρ is $\overline{\rho}_n = D_n C_n^{-1}$. Asymptotic properties of $\overline{\rho}_n$ are studied in Section 8.1.

If H is infinite dimensional, unboundedness of C^{-1} induces an erratic behavior of C_n^{-1} and $\overline{\rho}_n$ is not satisfactory. One remedy is to project data over a suitable finite-dimensional subspace of H.

If (v_j) is known, a natural choice is $H_{k_n} = \mathrm{sp}\{v_1, \ldots, v_{k_n}\}$. If (v_j) is unknown, H_{k_n} is replaced by $\widetilde{H}_{k_n} = \mathrm{sp}\{v_{1n}, \ldots, v_{k_n n}\}$ to obtain a general estimator, say ρ_n.

Under mild conditions,

$$\| \rho_n - \rho \|_{\mathcal{L}} \to 0 \ \ a.s.$$

More stringent conditions yield the bound

$$P(\| \rho_n - \rho \|_{\mathcal{L}} \geq \eta) \leq c_1(\eta) \ \exp\left(-c_2(\eta)\frac{n}{(\mathrm{Log} n)^4}\right), \ \ \eta > 0,\ n \geq n_\eta,$$

where $c_1(\eta)$ and $c_2(\eta)$ are positive constants (Section 8.3).

Now consider an $ARC_0(1)$ process associated with ρ, defined by

$$\rho(x)(t) = \int_0^1 r(s,t)x(s)ds, \quad x \in C_0,$$

where r is continuous and such that $\sup_{0 \le s,t \le 1} |r(s,t)| < 1$. Then ρ induces a bounded linear operator over $L^2 := L^2\left([0,1],\ \mathcal{B}_{[0,1]}, \lambda\right)$, where λ is Lebesgue measure. Using this property, one may employ methods similar to those above to estimate ρ in this new context (see Section 8.4).

Finally, $\rho_n(X_n)$ is a natural statistical predictor of X_{n+1} that provides sharp approximation of the best predictor $\rho(X_n)$ (see Section 8.5).

S.10 Applications

Chapter 9 is devoted to implementation of the functional predictor $\rho_n(X_n)$.

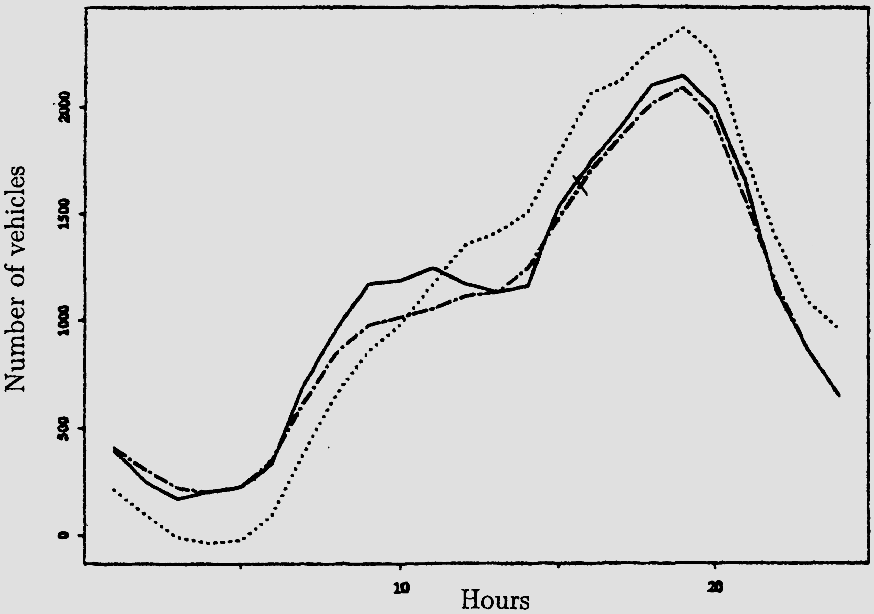
The successive (possible) steps can be summarized as follows.

1. Elimination of trend
2. Interpolating or smoothing of data (if they are discrete).
3. Test of "independence" against "$ARH(p)$ model".
4. Identification of p (if necessary).
5. Estimation of the mean, covariance operator and cross-covariance operator.
6. Determination of dimension k_n by validation, and computation of $\rho_n(X_n)$.

Some numerical examples appear in Section 9.4. They deal with the following topics:

- Simulations
- Prediction of electricity consumption
- Prediction of traffic (see Figure 2)
- Prediction of *el niño*
- Prediction of electrocardiograms.

The results show good accuracy and robustness of the functional predictor in various situations.



Prediction of traffic by functional autoregressive methods
Solid: data. Dotted: SARIMA predictor
Dashed: functional predictor.

Figure 2

1

Stochastic Processes and Random Variables in Function Spaces

This chapter deals with basic facts about probability theory over infinite-dimensional spaces. The underlying topic is the study of random processes considered as random functions.

Our main goal is to present the necessary background for the special theory developed in the following chapters. Most of the proofs are omitted. Some complement will appear in appendix.

Generalities about stochastic processes are presented in Section 1. Section 2 discusses the well-known interpretation of a stochastic process as a random variable with values in a function space. Sections 3 and 4 deal with moments in Banach spaces; for this study we need some facts concerning operators in function spaces. The important special case of Hilbert spaces appears in Section 5. Section 6 is devoted to linear prediction in Hilbert spaces.

1.1 Stochastic processes

Let $(\Omega, \mathcal{A}, P)$ be a probability space: Ω is the set where the random experiment takes place, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and P is a probability measure over $\mathcal{A}$. For simplicity we will assume that this space is complete (i.e., $\mathcal{A}$ contains the P-negligible sets).

Now let $\xi = (\xi_t,\ t \in T)$ be a family of random variables defined on $(\Omega, \mathcal{A}, P)$ and with values in a measurable space $(E, \mathcal{B})$. ξ is called a **stochastic process** (or **random process**) with **sample space** $(\Omega, \mathcal{A}, P)$, **state space** $(E, \mathcal{B})$, and **time set** T.

If T is countable (e.g. $T = \mathbb{N}$ or $\mathbb{Z}$), ξ is a **discrete-time process**, if T is an interval in $\mathbb{R}$ (e.g. $T = [a, b]$, $\mathbb{R}_+$ or $\mathbb{R}$), ξ is a **continuous-time process**. If $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, where $\mathcal{B}_{\mathbb{R}}$ is the Borel σ-algebra of $\mathbb{R}$, ξ is said to be a **real** stochastic process.

The random variable $\omega \mapsto \xi_t(\omega)$ is **the state of $\boldsymbol{\xi}$ at time $\boldsymbol{t}$** and the mapping $t \mapsto \xi_t(\omega)$ is the **sample path** of ξ associated with ω.

Random functions

In order to interpret ξ as a **random function** one may consider the space E^T of mappings from T to E equipped with the σ-algebra $\mathcal{S} = \sigma(\pi_t,\ t \in T)$, where $\pi_t : E^T \longrightarrow E$ is defined by $\pi_t(x) = x(t),\ x \in E^T$.

Noting that $\xi_t^{-1}(B) = \xi^{-1}\left[\pi_t^{-1}(B)\right] \in \mathcal{A}$ for all $B \in \mathcal{B}$ and all $t \in T$, we infer that ξ is $\mathcal{A} - \mathcal{S}$ measurable, and thus defines a random variable with values in $(E^T, \mathcal{S})$.

This allows us to define the **distribution of $\boldsymbol{\xi}$** by putting

$$P_\xi(S) = P(\xi \in S), \quad S \in \mathcal{S}.$$

The **finite-dimensional distributions** of ξ are the distributions of all the random vectors $(\xi_{t_1}, \ldots, \xi_{t_k})$, where $k \geq 1$ and $t_1, \ldots, t_k \in T$.

The following statement is useful.

THEOREM 1.1 *If E is a complete separable metric space and $\mathcal{B} = \mathcal{B}_E$ is its Borel σ-algebra, then the finite, dimensional distributions of ξ determine P_ξ.*

Equivalent processes

Recall that two random variables are said to be equivalent if they coincide almost surely (a.s.). In the context of stochastic processes the situation is somewhat more sophisticated: let $\xi = (\xi_t,\ t \in T)$ and $\eta = (\eta_t,\ t \in T)$ be two stochastic processes with the same sample space and state space. We say that η is a **modification** of ξ if

$$P(\eta_t = \xi_t) = 1, \quad t \in T. \tag{1.1}$$

We say ξ and η are **indistinguishable** if $\bigcap_{t \in T} \{\xi_t = \eta_t\} \in \mathcal{A}$ and

$$P\left(\bigcap_{t \in T} \{\xi_t = \eta_t\}\right) = 1. \tag{1.2}$$

Clearly (1.2) implies (1.1), but if T is uncountable, (1.1) does not imply (1.2), as shown by the following classical example.

Example 1.1
Let $(\Omega, \mathcal{A}, P)$ be $([0,1],\ \mathcal{B}_{[0,1]},\ \lambda)$ where, λ denotes Lebesgue measure, and consider the processes

$$\begin{aligned}\xi_t(\omega) &= 0,\ \omega \in [0,1],\ t \in [0,1],\\ \eta_t(\omega) &= 1_{\{t\}}(\omega),\ \omega \in [0,1],\ t \in [0,1].\end{aligned}$$

Since $\bigcap_{t\in[0,1]} \{\xi_t = \eta_t\} = \emptyset$, η is a modification of ξ when ξ and η are not indistinguishable.

Indistinguishability is a desirable property, since it means that ξ and η almost surely have same sample paths. A criterion is given by the following.

THEOREM 1.2 *Let ξ and η be real continuous-time processes with right-continuous sample paths (a.s.). Then, if η is a modification of ξ, η and ξ are indistinguishable.*

Let us now give some important examples of real processes.

Example 1.2
A **white noise** is a sequence $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ of real random variables such that $0 < \sigma^2 = E\varepsilon_n^2 < \infty$, $E\ \varepsilon_n = 0$, $E(\varepsilon_n \varepsilon_m) = \sigma^2 \delta_{n,m};\ n, m \in \mathbb{Z}$ ($\delta_{n,m} = 1$ if $m = n$, $= 0$ if $n \neq m$). If in addition (ε_n) is a sequence of independent random variables with the same distribution (**i.i.d.** random variables), then ε is called a **strong white noise**.

Example 1.3
A **real linear process** is defined by

$$\xi_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j},\quad n \in \mathbb{Z}, \tag{1.3}$$

where (ε_n) is a white noise and (a_j) a sequence of real numbers such that $a_0 = 1$ and $\sum_j a_j^2 < \infty$. The series in (1.3) converges in quadratic mean.

The random variable ξ_n may be interpreted as a linear combination of shocks which take place at times $n, n-1, \ldots$

If $a_j = \rho^j$, $j \geq 0$ where $-1 < \rho < 1$, (ξ_n) satisfies the relation

$$\xi_n = \rho \xi_{n-1} + \varepsilon_n\ ,\quad n \in \mathbb{Z}, \tag{1.4}$$

and is then called an **autoregressive process of order 1** $(\boldsymbol{AR}(\mathbf{1}))$. We will study a generalization of this model in an infinite-dimensional space in Chapters 3 and 6.

Example 1.4 A real process $\xi = (\xi_t,\ t \in T)$ is a **second order process** if $E\xi_t^2 < \infty$ for all t in T. For such a process one defines the **mean function**

$$m(t) = E\xi_t, \quad t \in T, \tag{1.5}$$

and the **covariance function**

$$c(s,t) = \mathrm{Cov}(\xi_s, \xi_t); \quad s,\ t \in T. \tag{1.6}$$

A covariance function is **symmetric**, i.e.,

$$c(s,t) = c(t,s)\ ; \quad s,\ t \in T,$$

and **positive** in the following special sense:

$$\sum_{1 \le i,j \le k} \alpha_i \alpha_j c(t_i, t_j) \ge 0, \ k \ge 1; \ \alpha_1, \dots, \alpha_k \in \mathbb{R}; \ t_1, \dots, t_k \in T.$$

Example 1.5
A second order process $\xi = (\xi_t,\ t \in T)$ is called **Gaussian** if all the random variables $\sum_{i=1}^{k} \alpha_i \xi_{t_i}$; $k \ge 1$; $\alpha_1, \dots, \alpha_k \in \mathbb{R}$; $t_1, \dots, t_k \in T$; are real Gaussian random variables (recall that constants are (degenerate) Gaussian random variables).

Theorem 1.1 and special properties of Gaussian vectors imply that m and c determine P_ξ.

Example 1.6
A **Wiener process** (or **Brownian motion process**) $W = (W_t,\ t \ge 0)$ is a Gaussian process such that $m \equiv 0$ and $c(s,t) = \sigma^2 \min(s,t)$, where σ^2 is a strictly positive constant. If $\sigma^2 = 1$, W is said to be **standard**.

W models the displacement of a particle in a homogeneous and motionless fluid: W_t is the absciss at time t of the orthogonal projection of the particle over a given axis.

A Wiener process has **independent stationary increments**. This means that

$$P_{(W_{t_1+h} - W_{t_0+h}, \dots, W_{t_k+h} - W_{t_{k-1}+h})} = P_{(W_{t_1} - W_{t_0})} \otimes \dots \otimes P_{(W_{t_k} - W_{t_{k-1}})};$$

$k \ge 2$, $0 \le t_0 < \dots < t_k$, $h \ge 0$.

Example 1.7
A **diffusion process** is a solution of a stochastic differential equation (see appendix)

$$d\xi_t = m(\xi_t, t)dt + \sigma(\xi_t, t)dW_t, \tag{1.7}$$

where W is a standard Wiener process and m and σ are measurable functions that satisfy some regularity conditions. Simple sufficient conditions for existence of a solution for (1.7) are

$$|m(x,t) - m(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq \kappa |x - y| \tag{1.8}$$

and

$$|m(x,t)|^2 + |\sigma(x,t)|^2 \leq \kappa(1 + x^2), \tag{1.9}$$

$t \geq 0$; $x, y \in \mathbb{R}$, where κ is a positive constant.

Equation (1.7) models the displacement of a particle in a fluid that is inhomogeneous and in motion. It is also used to describe the dynamics of many time-varying random phenomena in economics, finance, biology, chemistry, engineering, and other fields.

Example 1.8

A **Poisson process** $(N_t,\ t \geq 0)$ is defined by

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(\tau_n), \quad n \geq 0,$$

where $0 < \tau_1 < \ldots < \tau_n \uparrow +\infty$ (a.s.) is a discrete-time process with independent stationary increments.

Then it can be shown that there exists $\lambda > 0$ such that

$$P\left(N_{t_2} - N_{t_1} = k\right) = e^{-\lambda(t_2 - t_1)} \frac{[\lambda(t_2 - t_1)]^k}{k!},$$

$k = 0, 1, 2, \ldots$; $0 \leq t_1 < t_2$; and that $(N_t,\ t \geq 0)$ has independent stationary increments too.

The zero-mean process $(N_t - \lambda t,\ t \geq 0)$ is called a **compensated Poisson process.**

The variables $\tau_1, \tau_2, \ldots, \tau_n, \ldots$ are interpreted as random time arrivals of events such as telephone calls, particles impacts, etc.

Stationary processes

Suppose that T is an additive subgroup of $\mathbb{R}$ e.g., ($T = \mathbb{R}$ or $\mathbb{Z}$), and consider a real process $\xi = (\xi_t,\ t \in T)$ and set

$$\tau^h \xi = (\xi_{t+h},\ t \in T), \quad h \in T.$$

Then ξ is said to be **strictly stationary** if

$$P_{\tau^h \xi} = P_\xi, \quad h \in T. \tag{1.10}$$

Theorem 1.1 shows that ξ is strictly stationary if and only if

$$P_{(\xi_{t_1+h}, \ldots, \xi_{t_k+h})} = P_{(\xi_{t_1}, \ldots, \xi_{t_k})}, \quad k \geq 1, \quad t_1, \ldots, t_k, \ h \in T. \tag{1.11}$$

Definition (1.10) and property (1.11) directly extend to processes with values in a separable Banach space.

Now, if ξ is a second order process (see Example 1.4), it is **weakly stationary** if

$$m(t) = m, \quad t \in T,$$

and

$$c(s+h, t+h) = c(s,t); \quad s, t, h \in T.$$

For such a process one defines the **autocovariance** γ by setting

$$\gamma(h) = c(0,h), \quad h \in T, \tag{1.12}$$

and, if $\gamma(0) > 0$, the **autocorrelation** as

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in T. \tag{1.13}$$

A strictly stationary second order process is weakly stationary. The converse is not true. However, a weakly stationary Gaussian process is strictly stationary since its distribution is determined by m and γ.

A real linear process (see Example 1.3) is weakly stationary. It is strictly stationary if (ε_n) is a strong white noise.

Example 1.9
Consider the Langevin stochastic differential equation

$$d\xi_t = -\theta\xi_t dt + \sigma dW_t, \tag{1.14}$$

with $\theta > 0$, $\sigma > 0$, and $W_t = W_t^{(1)} 1_{\mathbb{R}_+}(t) + W_{-t}^{(2)} 1_{\mathbb{R}_-}(t)$, $t \in \mathbb{R}$, where $\left(W_t^{(1)}\right)$ and $\left(W_t^{(2)}\right)$ are two independent standard Wiener processes. $(W_t,\ t \in \mathbb{R})$ is then called a **bilateral Wiener process**.

It can be shown that (1.14) has a stationary solution given by the zero-mean Gaussian process

$$\xi_t = \sigma \int_{-\infty}^{t} e^{-\theta(t-s)} dW_s, \quad t \in \mathbb{R}, \tag{1.15}$$

where the stochastic integral is taken in Ito sense (see appendix).

This process is called an **Ornstein-Uhlenbeck (O.U.) process**. Its autocovariance is

$$\gamma(h) = \frac{\sigma^2}{2\theta} e^{-\theta|h|}, \quad h \in \mathbb{R}. \tag{1.16}$$

The O.U. process provides an alternative model for Brownian motion. Actually, the Wiener process has continuous but not differentiable sample paths (a.s.). Thus a Brownian motion particle has not a velocity! Defining the position of the particle by integrating the O.U. process, one obtains a particle with velocity.

In a more general context consider a **Markov stationary Gaussian** continuous-time process $\xi = (\xi_t,\ t \in \mathbb{R})$ that is zero-mean and nondegenerate and has continuous autocorrelation ρ such that $\rho(\delta) < 1$ for some $\delta > 0$. Then ξ has a modification which is an O.U. process.

We will see below that an O.U. process has autoregressive representations in some suitable Banach spaces.

1.2 Random functions

We have seen that a stochastic process ξ may be interpreted as a random variable with values in the function space $(E^T, \mathcal{S})$.

However, this space is not, in general, the suitable space to work with in order to obtain precise information about the probabilistic structure of ξ.

To see this, consider the processes in example 1.1. We have $P_\xi = P_\eta$, but

$$P\left(\sup_{0\leq t\leq 1} |\xi_t| > \frac{1}{2}\right) = 0\ ,\quad P\left(\sup_{0\leq t\leq 1} |\eta_t| > \frac{1}{2}\right) = 1,$$

and

$$P(t \mapsto \xi_t \text{ continuous}) = 1, \quad P(t \mapsto \eta_t \text{ continuous}) = 0.$$

This shows that the σ-algebra $\mathcal{S}$ is in some sense too small. Note that, on the contrary, E^T is in general too large, since it may contains very irregular functions.

Now, if the sample paths of ξ possess some regularity properties, it is possible to define ξ as a random variable in a nice subspace of E^T.

In this section we present some classical regularity conditions.

The weakest one is measurability. A real continuous-time process ξ is **measurable** if $(t,\omega) \mapsto \xi_t(\omega)$ is $\mathcal{B}_T \otimes \mathcal{A} - \mathcal{B}_{\mathbb{R}}$ measurable. Concerning measurability we have the following.

LEMMA 1.1 *If ξ is measurable its sample paths are $\mathcal{B}_T - \mathcal{B}_{\mathbb{R}}$ measurable. If in addition $E|\xi_t|^p < \infty$, $t \in T$, where p is an integer, then $t \mapsto E\xi_t^p$ is also measurable.*

From Lemma 1.1 it follows that a measurable process defines a random variable with values in $(\mathcal{M}(T,\mathbb{R}),\ \mathcal{S}')$ where $\mathcal{M}(T,\mathbb{R})$ is the space of $\mathcal{B}_T - \mathcal{B}_{\mathbb{R}}$ measurable functions and $\mathcal{S}'$ is the trace of $\mathcal{S}$ on $\mathcal{M}(T,\mathbb{R})$.

Measurability in Banach spaces

In order to work with more interesting function spaces we need a measurability criterion in a separable Banach space.

Let B be a separable Banach space with norm $\|\cdot\|$ and Borel σ-algebra $\mathcal{B}_B$. A ***B*-valued random variable** (or ***B*-random variable**) defined on

$(\Omega, \mathcal{A}, P)$ is an $\mathcal{A} - \mathcal{B}_B$ measurable mapping.

Now let B^* be the (topological) dual of B (i.e., the set of all continuous linear functionals defined on B). The usual norm over B^* is defined by

$$\| x^* \| = \sup_{\substack{x \in B \\ \| x \| \leq 1}} |x^*(x)|, \quad x^* \in B^*,$$

and $(B^*, \ \| \cdot \|)$ is a Banach space.

Let us consider the cylindrical σ-algebra on B, defined as

$$\mathcal{C}_B = \sigma(x^*, \ x^* \in B^*).$$

We see that $\mathcal{C}_B \subset \mathcal{B}_B$ and, in fact, separability of B entails $\mathcal{C}_B = \mathcal{B}_B$. A consequence of this equality is the following simple measurability criterion.

LEMMA 1.2 *If B is a separable Banach space, then $X : (\Omega, \mathcal{A}) \to (B, \mathcal{B}_B)$ is a B-valued random variable if and only if $x^*(X)$ is a real random variable for every x^* in B^*.*

We now give examples of applications.

Example 1.10

Let $C[0,1]$ be the space of continuous real functions defined on $[0,1]$, equipped with the sup-norm

$$\| x \| = \sup_{t \in [0,1]} |x(t)|, \quad x \in C[0,1].$$

Then $C[0,1]$ is a separable Banach space and a Riesz theorem states that its dual is the space $\mathcal{M}[0,1]$ of bounded signed measures over $([0,1], \ \mathcal{B}_{[0,1]})$.

Now let $\xi = (\xi_t, \ 0 \leq t \leq 1)$ be a real process with continuous sample paths. We claim that ξ defines a $C[0,1]$-random variable.

From Lemma 1.2 it suffices to show that $\int_0^1 \xi_t d\mu(t)$ is a real random variable for all μ in $\mathcal{M}[0,1]$. Since $\xi = \xi_+ - \xi_-$ and $\mu = \mu_+ - \mu_-$, we may assume that ξ and μ are positive.

Then consider the increasing sequence of processes

$$\xi_n(t) = \sum_{j=1}^{2^n} \mathbf{I}_{\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right[}(t) \inf_{\frac{j-1}{2^n} \leq t \leq \frac{j}{2^n}} \xi_t + \mathbf{I}_{\{1\}}(t)\xi(1),$$

$0 \leq t \leq 1, \ n \geq 1$. Note that $\inf_{t \in I_{jn}} \xi_t$ where $I_{jn} = \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right]$, is actually a random variable, since it is equal to $\inf_{t \in I_{jn} \cap \mathbb{Q}} \xi_t$.

Integrating with respect to μ we obtain the sequence of random variables

$$\int_0^1 \xi_n(t)d\mu(t) = \sum_{j=1}^{2^n} \mu\left(\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right[\right) \inf_{\frac{j-1}{2^n} \leq t \leq \frac{j}{2^n}} \xi_t + \mu(\{1\})\xi(1), \ n \geq 1.$$

Hence the integral

$$\int_0^1 \xi_t d\mu(t) = \lim_{n\uparrow\infty} \uparrow \int_0^1 \xi_n(t)d\mu(t)$$

is a random variable and the proof is complete.

Example 1.11

Let $B = L^p([0,1], \mathcal{B}_{[0,1]}, \lambda)$, where $1 < p < \infty$, with norm $\| x \| = \left(\int_0^1 |x(t)|^p\right)^{1/p}$, $x \in B$.

Then B is a separable Banach space and its dual space is $B^* = L^q([0,1], \mathcal{B}_{[0,1]}, \lambda)$ with $\frac{1}{p} + \frac{1}{q} = 1$.

If $\xi = (\xi_t,\ 0 \leq t \leq 1)$ is a measurable process with sample paths in $L^p([0,1],\ \mathcal{B}_{[0,1]}, \lambda)$, then $\int_0^1 \xi_t\ y(t)dt$ is a random variable for all y in B^* and from Lemma 1.2 it follows that ξ defines a B-valued random variable.

The above measurability results basically deal with continuous-time processes. For discrete time-processes the situation is simpler, as shown by the next example.

Example 1.12

Let ξ be a real linear process (example 1.3) and let μ be a measure on $\mathbb{Z}$ defined as

$$\mu = \sum_{n\in\mathbb{Z}} \alpha_n \delta_{(n)},$$

where $\alpha_n \geq 0$ and $\sum_n \alpha_n < \infty$. Then we have

$$\begin{aligned} E\left(\int \xi_n^2 d\mu\right) &= E\left(\sum_n \alpha_n \xi_n^2\right) = \sum_n \alpha_n E\xi_n^2 \\ &= \sigma^2 \left(\sum_n \alpha_n\right)\left(\sum_j a_j^2\right) < \infty, \end{aligned}$$

which implies $\int \xi_n^2 d\mu < \infty$ a.s., and ξ defines an $L^2(\mu)$-valued random variable since $\sum \alpha_n u_n \xi_n$ is a real random variable for all (u_n) in $L^2(\mu)$.

Regularity criteria

We finally indicate some classical regularity criteria for continuous-time processes. We begin with a measurability criterion.

THEOREM 1.3 *Let $\xi = (\xi_t,\ t \geq 0)$ be a real stochastic process which is continuous in probability. Then it admits a measurable modification.*

The next statement gives a sharp condition for continuity.

THEOREM 1.4 *Let $\xi = (\xi_t,\ 0 \leq t \leq 1)$ be a real stochastic process such that*

$$E|\xi_{t+h} - \xi_t|^p \leq k \frac{|h|}{|\mathrm{Log}|h||^{1+r}} \ ; \ t, t+h \in [0,1], \tag{1.17}$$

where $k > 0$, $0 < p < r$ are constants.

Then there is a modification of ξ with continuous sample paths.

Example 1.13

Let $\xi = (\xi_t,\ 0 \leq t \leq 1)$ be a zero-mean Gaussian process with a continuous covariance function c which satisfies

$$|c(t, t+h) - c(t,t)| \leq \kappa |h| \ ; \ \ t, t+h \in [0,1], \tag{1.18}$$

where κ is a constant.

Then we obtain

$$\begin{aligned} E(\xi_{t+h} - \xi_t)^2 &= c(t+h, t+h) - 2c(t, t+h) + c(t,t) \\ &= [c(t+h, t+h) - c(t, t+h)] - [c(t, t+h) - c(t,t)] \\ &\leq 2\kappa |h|. \end{aligned}$$

Now, since if $N \sim \mathcal{N}(0, \sigma^2)$ we have $E|N|^3 = \left(\frac{2}{\pi}\right)^{1/2} \sigma^3$, it follows that

$$E|\xi_{t+h} - \xi_t|^3 \leq k_1 |h|^{3/2},$$

where k_1 is a constant.

Hence (1.17) holds, with $p = 3$ and $r > 3$, so ξ has a modification with continuous sample paths.

This result applies to Wiener processes (example 1.6) and to Ornstein-Uhlenbeck processes (Example 1.9).

The last criterion provides an explicit form of the random variable associated with ξ. It is interesting in itself and will be used in the sequel of this book.

We first state a lemma

LEMMA 1.3 (Mercer lemma)

Let c be a covariance function continuous over $[0,1]^2$. Then there exists a *sequence* (φ_n) *of continuous functions and a decreasing sequence* (λ_n) *of positive numbers such that*

$$\int_0^1 c(s,t)\varphi_n(s)ds = \lambda_n\varphi_n(t),\ t \in [0,1],\ n \in \mathbb{N}, \tag{1.19}$$

and

$$\int_0^1 \varphi_n(s)\varphi_m(s)ds = \delta_{n,m}\ ;\ \ n,m \in \mathbb{N}. \tag{1.20}$$

Moreover,

$$c(s,t) = \sum_{n=0}^{\infty} \lambda_n\varphi_n(s)\varphi_n(t)\ ;\ \ s,t \in [0,1], \tag{1.21}$$

where the series converges uniformly on $[0,1]^2$*; hence*

$$\sum_{n=0}^{\infty} \lambda_n = \int_0^1 c(t,t)dt < \infty. \tag{1.22}$$

Now we may state the criterion.

THEOREM 1.5 (Karhunen-Loève (K.L.) expansion)
Let $\xi = (\xi_t,\ 0 \le t \le 1)$ *be a second order zero-mean measurable process with continuous covariance function* c*. Then*

$$\xi_t = \sum_{n=0}^{\infty} \eta_n\varphi_n(t),\ \ t \in [0,1], \tag{1.23}$$

where (η_n) *is a sequence of real zero-mean random variables such that*

$$E(\eta_n\eta_m) = \lambda_n\delta_{n,m}\ ;\ \ n,m \in \mathbb{N}; \tag{1.24}$$

and where the sequence (λ_n, φ_n) *is defined in the Mercer lemma. The series in (1.23) converges uniformly with respect to the* $L^2(\Omega, \mathcal{A}, P)$*-norm.*

Moreover, ξ *defines a* $L^2([0,1],\ \mathcal{B}_{[0,1]}, \lambda)$ *random variable via (1.23).*

Proof

Note first that measurability of ξ entails existence of the $\overline{\mathbb{R}}$-valued random variable $\int_0^1 |\xi_t\varphi_n(t)|dt$.

Now

$$\begin{aligned} E\left(\int_0^1 |\xi_t\varphi_n(t)|dt\right)^2 &\le E\left(\int_0^1 \xi_t^2 dt \cdot \int_0^1 \varphi_n^2(t)dt\right) \\ &\le \int_0^1 c(t,t)dt \int_0^1 \varphi_n^2(t)dt < \infty; \end{aligned}$$

hence $\int_0^1 |\xi_t \varphi_n(t)| dt < \infty$ with probability 1 and we may set

$$\eta_n = \int_0^1 \xi_t \varphi_n(t) dt, \quad n \in \mathbb{N} \quad (a.s.). \tag{1.25}$$

Then we clearly have $E\eta_n^2 < \infty$, $E\eta_n = 0$, and

$$E(\eta_n \eta_m) = \int_{[0,1]^2} \varphi_n(s) c(s,t) \varphi_m(t) ds dt = \lambda_n \delta_{n,m}. \tag{1.26}$$

On the other hand,

$$\begin{aligned} E(\xi_t \eta_n) &= E\left(\xi_t \int_0^1 \xi_s \varphi_n(s) ds\right) = \int_0^1 c(s,t) \varphi_n(s) ds \\ &= \lambda_n \varphi_n(t). \end{aligned} \tag{1.27}$$

Now we may write

$$E\left(\xi_t - \sum_{k=0}^{n} \eta_k \varphi_k(t)\right)^2 = E\xi_t^2 - 2\sum_{k=0}^{n} E(\xi_t \eta_k) \varphi_k(t) + E\left(\sum_{k=0}^{n} \eta_k \varphi_k(t)\right)^2,$$

then, from (1.26) and (1.27), it follows that

$$E\left(\xi_t - \sum_{k=0}^{n} \eta_k \varphi_k(t)\right)^2 = c(t,t) - \sum_{k=0}^{n} \lambda_k \varphi_k^2(t), \quad 0 \le t \le 1, \ n \in \mathbb{N},$$

and the Mercer lemma gives

$$\sup_{t \in [0,1]} E\left(\xi_t - \sum_{k=0}^{n} \eta_k \varphi_n(t)\right)^2 \underset{n \to \infty}{\longrightarrow} 0 \ .$$

Finally, eventually completing (φ_n) in order to obtain an orthonormal basis and using (1.23) and Lemma 1.2, it is easy to check that ξ defines a $L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$-random variable. ■

For example, it can be proved that a measurable version of the standard Wiener process has a K.L. expansion of the form

$$W_t = \sqrt{2} \sum_{n=0}^{\infty} \eta_n \sin\left[(2n+1)\frac{\pi}{2} t\right], \quad t \in [0,1], \tag{1.28}$$

where (η_n) is a sequence of independent zero-mean Gaussian random variables, with $E\eta_n^2 = \dfrac{4}{(2n+1)^2 \pi^2}$, $n \in \mathbb{N}$.

Concerning the Ornstein-Uhlenbeck process we have the following: for convenience suppose that $\theta = 1$ and $\sigma^2 = 2$ so that the covariance is $c(s,t) = e^{-|t-s|}$; then we have

$$\xi_t = \sqrt{2}\sum_{n=0}^{\infty}(1+\lambda_n)^{-1/2}\eta_n \sin\left[\omega_n\left(t-\frac{1}{2}\right)+(n+1)\frac{\pi}{2}\right],\ t\in[0,1], \tag{1.29}$$

where $\lambda_n = 2\left(1+\omega_n^2\right)^{-1}$, $n \geq 0$, and $\omega_0, \omega_1, \ldots$ are the positive roots of the equation

$$tg\omega = -2\omega(1-\omega^2)^{-1}$$

arranged in ascending order.

1.3 Expectation and conditional expectation in Banach spaces

Let B be a separable real Banach space, $\|\cdot\|$ its norm, and B^* its dual space. In the current section we study integration of B-random variables.

We will say that a B-random variable X is **weakly integrable** if $x^*(X)$ is integrable for all x^* in B^* and if there exists an element of B, denoted by EX, such that

$$E(x^*(X)) = x^*(EX), \quad x^* \in B^*. \tag{1.30}$$

EX is called the **weak expectation** or **weak integral** of X. A simple random variable is weakly integrable.

Now X is said to be **integrable** (or strongly integrable) if $E \parallel X \parallel < \infty$.

LEMMA 1.4 *Let X be an integrable B-random variable. Then there exists a sequence (X_n) of simple B-random variables such that*

$$\lim_{n\to\infty} E \parallel X_n - X \parallel = 0. \tag{1.31}$$

Moreover, X is weakly integrable.

Proof

Let $\varepsilon > 0$ and $(B_j,\ j \geq 1)$ be a partition of B in Borel sets of diameter less than $\frac{\varepsilon}{2}$. Let us set

$$Y_\varepsilon = \sum_{j=1}^{\infty} \mathbf{1}_{\{X\in B_j\}} x_j,$$

where $x_j \in B_j$, $j \geq 1$. Then $\parallel X - Y_\varepsilon \parallel \leq \frac{\varepsilon}{2}$, and Y_ε is integrable since $\parallel Y_\varepsilon \parallel \leq \parallel X \parallel + \frac{\varepsilon}{2}$.

Hence we can find an integer N_ε such that

$$\sum_{j>N_\varepsilon} \| x_j \| P(X \in B_j) < \frac{\varepsilon}{2}.$$

Putting $X_\varepsilon = \sum_{j=1}^{N_\varepsilon} x_j \mathbf{I}_{\{X \in B_j\}}$ we get

$$E \| X - X_\varepsilon \| \leq E \| X - Y_\varepsilon \| + E \| Y_\varepsilon - X_\varepsilon \| < \varepsilon. \tag{1.32}$$

If (ε_n) is a sequence converging to zero, the above result shows that it is possible to construct (X_n) satisfying (1.31).

Noting that for all x^* in B^* we have

$$|x^*(X_n - X)| \leq \| x^* \| \, \| X_n - X \|,$$

we obtain existence of $Ex^*(X)$ and

$$\lim_{n\to\infty} E[x^*(X_n)] = E[x^*(X)]. \tag{1.33}$$

On the other hand, (1.31) entails that (EX_n) is a Cauchy sequence. Setting $\lim_{n\to\infty} EX_n = m$, we get

$$E[x^*(X)] = x^*(m);$$

Thus X is weakly integrable and $EX = m$. ∎

If X is integrable, EX is called the **integral** or **expectation** of X. It is also denoted by $\int XdP$. Finally, one sets

$$\int_A XdP = \int \mathbf{I}_A XdP, \quad A \in \mathcal{A}.$$

Example 1.14

Let $\xi = (\xi_t,\ 0 \leq t \leq 1)$ be a real random process with continuous sample paths. Then it defines a $C[0,1]$-random variable (Example 1.10), and if $E\left(\sup_{0\leq t\leq 1} |\xi_t|\right) < \infty$, it is integrable and

$$E\left(\int_0^1 \xi_t d\mu(t)\right) = \int_0^1 (E\xi)(t)d\mu(t), \quad \mu \in \mathcal{M}([0,1]),$$

which implies $(E\xi)(t) = E[\xi(t)]$, $t \in [0,1]$.

We now list some important properties of expectation.

(1) The space $L_B^1(P)$ of equivalence classes of integrable B-random variables (with respect to the equivalence relation $X = Y$ a.s.) is a Banach space with respect to the norm $\| X \|_{1,B} = E \| X \|$.

(2) E defines a continuous linear operator from $L_B^1(P)$ to B, which satisfies the **contractive property** $\| EX \| \leq E \| X \|$.

(3) Let B_1 and B_2 be two separable Banach spaces and let $\ell \in \mathcal{L}(B_1, B_2)$ (the space of continuous linear operators from B_1 to B_2). If $X \in L_{B_1}^1(P)$, then $\ell(X) \in L_{B_2}^1(P)$ and

$$E\ell(X) = \ell E(X).$$

(4) **Dominated convergence :** If $X_n \to X$ a.s. in B and $\| X_n \| \leq Y$ a.s., where $n \geq 1$ and Y is an integrable real random variable, then $X_n \in L_B^1(P)$, $n \geq 1$, $X \in L_B^1(P)$, and $E \| X_n - X \| \underset{n\to\infty}{\longrightarrow} 0$.

(5) The space $L_B^p(P)$, where $1 < p < \infty$, of classes of B-random variables X such that $E \|X\|^p < \infty$, is a Banach space with respect to the norm $\| X \|_{p,B} = (E \| X \|^p)^{1/p}$.

(6) The space $L_B^\infty(P)$ of essentially bounded B-random variables (i.e., where there exists b such that $P(\| X \| < b) = 1$) is a Banach space with respect to the norm $\| X \|_{\infty,B} = \inf\{b > 0 : P(\| X \| < b) = 1\}$.

Conditional expectation

Conditional expectation can be defined similarly as in the case of real random variables. Let $\mathcal{A}_0$ be a sub-σ-algebra of $\mathcal{A}$. The **conditional expectation** relative to $\mathcal{A}_0$ is the mapping $E^{\mathcal{A}_0}$ from $L_B^1(\Omega, \mathcal{A}, P)$ to $L_B^1(\Omega, \mathcal{A}_0, P)$ such that

$$\int_A E^{\mathcal{A}_0} X dP = \int_A X dP, \quad A \in \mathcal{A}_0, \ X \in L_B^1(\Omega, \mathcal{A}, P). \tag{1.34}$$

Classical properties of conditional expectation are valid in the context of separable Banach spaces. In particular, the following properties hold.

(1) $E^{\mathcal{A}_0}$ is linear, and contractive in the sense that

$$\| E^{\mathcal{A}_0} X \| \leq E^{\mathcal{A}_0} \| X \| \quad (a.s.).$$

(2) Property (3) of expectation holds in the form

$$E^{\mathcal{A}_0} \ell(X) = \ell\left(E^{\mathcal{A}_0} X\right).$$

(3) If X and $\mathcal{A}_0$ are independent ($X \perp\!\!\!\perp \mathcal{A}_0$), then

$$E^{\mathcal{A}_0} X = EX;$$

in particular,

$$E^{\{\emptyset,\Omega\}} X = EX.$$

(4) If $X \in L^1_B(\Omega, \mathcal{A}_0, P)$, then $E^{\mathcal{A}_0}X = X$.

Finally, note that if

$$X = \sum_{j=1}^{k} \mathbf{1}_{A_j} x_j \;\; ; \;\; A_j \in \mathcal{A}, \;\; x_j \in B, \; j = 1, \ldots, k,$$

then

$$E^{\mathcal{A}_0}X = \sum_{j=1}^{k} P^{\mathcal{A}_0}(A_j)x_j,$$

where $P^{\mathcal{A}_0}$ denotes conditional probability relative to $\mathcal{A}_0$.

1.4 Covariance operators and characteristic functionals in Banach spaces

Covariance operator

Let $X \in L^2_B(\Omega, \mathcal{A}, P)$ such that $EX = 0$. The covariance operator C_X of X is the bounded linear operator from B^* to B, defined by

$$C_X(x^*) = E[x^*(X)X], \;\; x^* \in B^*. \tag{1.35}$$

If $EX \neq 0$, one may set $C_X = C_{X-EX}$, so we will deal only with zero-mean random variables in the remainder of this section.

C_X is completely determined by the **covariance function** of X, defined as

$$c_X(x^*, y^*) = y^*[C_X(x^*)] = E[x^*(X)y^*(X)] = \mathrm{Cov}(x^*(X), y^*(X));$$
$$x^*, y^* \in B^*. \tag{1.36}$$

It is often more convenient to use c_X instead of C_X, as shown in the examples given below.

Example 1.15
If $B = \mathbb{R}^d$, C_X is defined as

$$C_X(x_1, \ldots, x_d) = E\left[\left(\sum_{i=1}^{d} x_i X_i\right) X\right],$$

where $x^* = (x_1, \ldots, x_d)$ and $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$. Concerning c_X we have

$$c_X[(x_1, \ldots, x_d), (y_1, \ldots, y_d)] = \sum_{1 \le i,j \le d} x_i E(X_i X_j) y_j,$$

$(x_1, \ldots, x_d) \in \mathbb{R}^d$, $(y_1, \ldots, y_d) \in \mathbb{R}^d$. Thus c_X is the bilinear form associated with the covariance matrix of X.

Example 1.16

In example 1.11, if $E\left(\int_0^1 |\xi_t|^p dt\right)^{2/p} < \infty$ then the L^p-random variable X associated with ξ admits a covariance operator given by

$$C_X(x^*) = E\left[\int_0^1 X_t x^*(t)dt \cdot X\right], \quad x^* \in L^q(\lambda),$$

while

$$\begin{aligned} c_X(x^*, y^*) &= E\left(\int_0^1 x^*(t)X_t dt \int_0^1 y^*(t)X_t dt\right) \\ &= \int_{[0,1]^2} x^*(s)E(X_s X_t)y^*(t)dsdt, \quad x^*, y^* \in L^q(\lambda). \end{aligned} \tag{1.37}$$

Hence c_X is determined by the covariance function of any version of X.

Example 1.17

If $B = C[0,1]$ and $E\left(\sup_{0\le t\le 1} |X_t|^2\right) < \infty$, then

$$c_X(\mu, \nu) = \int_{[0,1]^2} E(X_s X_t)d\mu(s)d\nu(t)\ ; \quad \mu,\ \nu \in \mathcal{M}([0,1]); \tag{1.38}$$

and c_X is determined by the covariance function of X.

We now indicate a condition for a mapping to be a covariance operator.

THEOREM 1.6 *Let C be a mapping from B^* to B. Then there exists a B-random variable X such that $C = C_X$ if and only if*

$$C(x^*) = \sum_{j=1}^{\infty} x^*(x_j)x_j, \quad x^* \in B^*, \tag{1.39}$$

where (x_j) is a fixed sequence in B such that $\sum_{j=1}^{\infty} \| x_j \|^2 < \infty$.

If B is a Hilbert space we will see that (x_j) has a specific interpretation.

Cross-covariance operators

Let $X \in L^2_{B_1}(P)$ and $Y \in L^2_{B_2}(P)$ where B_1 and B_2 are separable Banach spaces and suppose that $EX = 0$ and $EY = 0$. Then the cross-covariance operators of X and Y are defined by

$$\begin{aligned} C_{X,Y} : B_1^* &\mapsto B_2 \\ x^* &\mapsto E(x^*(X)Y) \end{aligned}$$

and

$$\begin{aligned} C_{Y,X} : B_2^* &\mapsto B_1 \\ y^* &\mapsto E(y^*(Y)X). \end{aligned}$$

Observe that $C_{X,Y}$ and $C_{Y,X}$ are bounded linear operators and that

$$y^*[C_{X,Y}(x^*)] = x^*[C_{Y,X}(y^*)] = E(x^*(X)y^*(Y)), \ \ x^* \in B_1^*, \ y^* \in B_2^*.$$

So the **cross-covariance function**

$$c_{X,Y}(x^*, y^*) = E(x^*(X)y^*(Y)), \ \ x^* \in B_1^*, \ y^* \in B_2^*,$$

completely determines $C_{X,Y}$ and $C_{Y,X}$.

Examples similar to Examples 1.15 and 1.16 show that $(s,t) \mapsto E(X_s Y_t)$ determines $c_{X,Y}$ as soon as X and Y may be interpreted as random functions associated with suitable function spaces.

Characteristic functional

The **characteristic functional** of a B-random variable X is the mapping from B^* to $\mathbb{C}$ defined by

$$\varphi_X(x^*) = E(e^{ix^*(X)}), \ \ x^* \in B^*.$$

Associated to φ_X is the **Fourier transform of $P_X = \mu$**, defined by

$$\varphi_\mu(x^*) = \int e^{ix^*(x)} d\mu(x) = \varphi_X(x^*).$$

Elementary properties of $\varphi = \varphi_X$ are the same as in the finite dimensional case:

(1) $\varphi(0) = 1$, and φ is continuous with respect to the norm in B^* and **positive**, i.e.,

$$\sum_{1 \le i,j \le k} z_i \overline{z_j} \varphi(x_i^* - x_j^*) \ge 0$$

for all $k \ge 1$, $z_1, \ldots, z_k \in \mathbb{C}$, $x_1^*, \ldots, x_k^* \in B^*$.

(2) φ determines P_X (or μ).

(3) If X and Y are independent B-valued random variables, then

$$\varphi_{X+Y}(x^*) = \varphi_X(x^*)\varphi_Y(x^*), \ \ x^* \in B^*.$$

The characteristic functional of a Gaussian random variable has a special form. First, a B-random variable X is said to be **Gaussian** if $x^*(X)$ is a Gaussian real random variable for all x^* in B^*. So X is Gaussian if and only if $E \| X \|^2 < \infty$ and, for all x^* in B^*,

$$\varphi_X(x^*) = \exp(ix^*(EX) - \frac{1}{2}x^*[C_X(x^*)]). \tag{1.40}$$

It is important to note that, given m in B and a covariance operator C, it is not always possible to find a Gaussian B-random variable X such that $EX = m$ and $C_X = C$. An example of such a phenomenon will appear in Chapter 2.

1.5 Random variables and operators in Hilbert spaces

In this section we consider the special case where $B = H$ is a separable real Hilbert space with scalar product $< \cdot, \cdot >$ and norm $\| \cdot \|$. We will see that the geometric structure of H induces some significant regularity properties of H-random variables and their associated parameters.

Expectation and conditional expectation

First, $L^2_H(\Omega, \mathcal{A}, P)$ is also a Hilbert space, with scalar product given by

$$(X, Y) = E < X, Y > \ , \quad X, Y \in L^2_H(P).$$

This fact leads us to interpret the conditional expectation $E^{\mathcal{A}_0}$ as the orthogonal projector of $L^2_H(\Omega, \mathcal{A}, P)$ onto $L^2_H(\Omega, \mathcal{A}_0, P)$. Thus, if $X \in L^2_H(\Omega, \mathcal{A}, P)$, $E^{\mathcal{A}_0}X$ is characterized by

$$E^{\mathcal{A}_0}X \in L^2_H(\Omega, \mathcal{A}_0, P), \tag{1.41}$$

$$E < X, Z >= E < E^{\mathcal{A}_0}X, Z > \ , \quad Z \in L^2_H(\Omega, \mathcal{A}_0, P), \tag{1.42}$$

and (1.42) is equivalent to

$$\int_A X dP = \int_A E^{\mathcal{A}_0} X dP \ , \quad A \in \mathcal{A}_0. \tag{1.43}$$

In particular, expectation is the orthogonal projector onto $L^2_H(\Omega, \{\emptyset, \Omega\}, P)$, a space that is isomorphic to H itself.

Operators in Hilbert spaces

At this stage we need some classical properties of operators acting on H.

Let $\mathcal{L} = \mathcal{L}(H, H)$ be the space of **bounded** linear operators from H to H. The **uniform norm**

$$\| \ell \|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \| \ell(x) \|, \quad \ell \in \mathcal{L}, \tag{1.44}$$

supplies a Banach space structure on $\mathcal{L}$.

To each $\ell \in \mathcal{L}$ one may associate its **adjoint** ℓ^*, defined by

$$< \ell^*(x), y > = < x, \ell(y) > ; \; x, y \in H. \tag{1.45}$$

Now ℓ is said to be **compact** if there exist two orthonormal bases (e_j) and (f_j) of H and a sequence (λ_j) of real numbers tending to zero, such that

$$\ell(x) = \sum_{j=1}^{\infty} \lambda_j < x, e_j > f_j, \quad x \in H, \tag{1.46}$$

or in a more concise form,

$$\ell = \sum_{j=1}^{\infty} \lambda_j e_j \otimes f_j. \tag{1.47}$$

Equation (1.46) (or (1.47)) is called the **spectral decomposition** of ℓ. Note that eventually substituting $-f_j$ for f_j, one may choose positive λ_j's.

An equivalent definition is the following: ℓ is called **compact** if it transforms each weakly convergent sequence into a strongly convergent one; that is: if $< x_n, y > \underset{n\to\infty}{\longrightarrow} < x, y >$ for all y in H then $\| \ell(x_n) - \ell(x) \| \underset{n\to\infty}{\longrightarrow} 0$ as $n \to \infty$. The set of compact operators over H will be denoted by $\mathcal{C}$. Note that compactness is a rather restricting property since, for example, the identity operator is not compact whenever H is infinite dimensional.

A compact operator ℓ is said to be a **Hilbert-Schmidt operator** if $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$.

The space $\mathcal{S}$ of Hilbert-Schmidt operators is a separable Hilbert space with respect to the scalar product

$$< s_1, s_2 >_{\mathcal{S}} = \sum_{1 \leq i,j \leq \infty} < s_1(g_i), h_j >< s_2(g_i), h_j >, \tag{1.48}$$

where (g_i) and (h_j) are two arbitrary orthonormal bases of H. This scalar product does not depend on the choice of these bases.

The associated **Hilbert-Schmidt norm** satisfies

$$\| s \|_{\mathcal{S}} = \left(\sum_j \lambda_j^2 \right)^{1/2} = \left(\sum_j \| s(g_j) \|^2 \right)^{1/2}, \; s \in \mathcal{S}, \tag{1.49}$$

where the λ_j's are those that appear in the decomposition of s as a compact operator and where (g_j) is any orthonormal basis in H. Conversely, if an operator ℓ satisfies $\sum_j \| \ell(g_j) \|^2 < \infty$ for some orthonormal basis (g_j) in H, then it is a Hilbert-Schmidt operator.

An operator ℓ is **symmetric** if

$$< \ell(x), y > = < x, \ell(y) > , \quad x, y \in H, \tag{1.50}$$

and **positive** if

$$< \ell(x), x > \geq 0 , \quad x \in H. \tag{1.51}$$

If s is a symmetric positive Hilbert-Schmidt operator decomposition, (1.46) becomes

$$s(x) = \sum_{j=1}^{\infty} \lambda_j < x, e_j > e_j \quad , \quad x \in H, \tag{1.52}$$

and we have

$$s(e_j) = \lambda_j e_j \quad , \quad j \geq 1;$$

thus (λ_j, e_j), $j \geq 1$ is a complete sequence of eigenelements of s and λ_j is positive for each j. Moreover one may assume that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_j \geq \ldots \geq 0$. Since $\lambda_j \downarrow 0$ each eigensubspace is finite dimensional.

We now turn to the concept of **nuclear operator**: a compact operator is called nuclear if $\sum_j |\lambda_j| < \infty$. The **nuclear norm** is defined by

$$\| \ell \|_{\mathcal{N}} = \sum_{j=1}^{\infty} |\lambda_j|. \tag{1.53}$$

The space $(\mathcal{N}, \| \cdot \|_{\mathcal{N}})$ of nuclear operators over H is a separable Banach space.

Finally, we have the inclusions

$$\mathcal{N} \subset \mathcal{S} \subset \mathcal{C} \subset \mathcal{L} \tag{1.54}$$

and the inequalities

$$\| \cdot \|_{\mathcal{N}} \ \geq \ \| \cdot \|_{\mathcal{S}} \ \geq \ \| \cdot \|_{\mathcal{L}} \, . \tag{1.55}$$

Example 1.18

Take $H = L^2([0,1], \ \mathcal{B}_{[0,1]}, \lambda)$ and consider the **kernel operator** ℓ_K defined by

$$\ell_K(x)(t) = \int_0^1 K(s,t) x(s) ds, \quad t \in [0,1], \ x \in H,$$

where $K \in L^2([0,1]^2, \mathcal{B}_{[0,1]^2}, \lambda \otimes \lambda) = H \otimes H$. Then ℓ_K is a bounded linear operator on H and

$$\| \ell_K \|_{\mathcal{L}} \leq \left(\int_{[0,1]^2} K^2(s,t) ds dt \right)^{1/2} .$$

If (g_j) denotes an orthonormal basis in H, then $(g_i \otimes g_j)$ is an orthonormal basis in $H \otimes H$; therefore

$$\int_{[0,1]^2} K^2(s,t) ds dt = \sum_{i,j} \left(\int K(s,t) g_i(s) g_j(t) ds dt \right)^2$$

$$= \sum_j \sum_i \left(\int \ell_K(g_j)(s) g_i(s) ds \right)^2 = \sum_j \| \ell_K(g_j) \|^2 .$$

Thus ℓ_K is a Hilbert-Schmidt operator.

In the special case where K is a continuous symmetric, positive kernel, one may use the Mercer lemma for obtaining expansion (1.21):

$$K(s,t) = \sum_{j=0}^{\infty} \lambda_j \varphi_j(s) \varphi_j(t) \quad ; \quad s,t \in [0,1].$$

Since $\sum_{j=0}^{\infty} \lambda_n = \int_0^1 K(t,t) dt < \infty$, ℓ_K becomes a nuclear operator.

To complete this section we indicate some useful properties of $\mathcal{S}$ and $\mathcal{N}$.

(1) If $\ell \in \mathcal{L}$ and $s \in \mathcal{S}$ and then $\ell s \in \mathcal{S}$, $s\ell \in \mathcal{S}$, with

$$\| \ell s \|_{\mathcal{S}} \leq \| \ell \|_{\mathcal{L}} \| s \|_{\mathcal{S}} \text{ and } \| s\ell \|_{\mathcal{S}} \leq \| \ell \|_{\mathcal{L}} \| s \|_{\mathcal{S}} .$$

(2) If $s \in \mathcal{S}$, then $s^* \in \mathcal{S}$ and $\| s^* \|_{\mathcal{S}} = \| s \|_{\mathcal{S}}$.

(3) If $s_1, s_2 \in \mathcal{S}$ then $s_1 s_2 \in \mathcal{N}$ and

$$\| s_1 s_2 \|_{\mathcal{N}} = \| s_1 \|_{\mathcal{S}} \| s_2 \|_{\mathcal{S}} .$$

Conversely, if $\ell \in \mathcal{N}$, there exist Hilbert-Schmidt operators s_1 and s_2 such that $\ell = s_1 s_2$.

Covariance operators in Hilbert spaces

Consider an H-random variable X such that $E \| X \|^2 < \infty$ and $EX = 0$. Since H^* may be identified with H, the covariance operator of X is simply given by

$$C_X(x) = E[< X, x > X] , \quad x \in H. \tag{1.56}$$

It is symmetric positive and its norm satisfies

$$\begin{aligned} \| C_X \|_{\mathcal{L}} &= \sup_{\|x\|\leq 1} \| C_X(x) \| = \sup_{\substack{\|x\|\leq 1 \\ \|y\|\leq 1}} | < C_X(x), y > | \\ &= \sup_{\|x\|\leq 1} < C_X(x), x > . \end{aligned} \tag{1.57}$$

Now it is easy to show that C_X is compact, so it satisfies (1.46). Actually using symmetry and positivity, one obtains the decomposition

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j < x, v_j > v_j, \quad x \in H. \tag{1.58}$$

Here $\lambda_j =< C_X(v_j), v_j >= E(< X, v_j >^2)$, so

$$\sum_1^{\infty} \lambda_j = E \| X \|^2 < \infty.$$

Thus C_X is a nuclear operator and

$$\| C_X \|_{\mathcal{N}} = E \| X \|^2 . \tag{1.59}$$

On the other hand we have

$$\| C_X \|_{\mathcal{S}} = \left[\sum_{j=1}^{\infty} \left(E < X, v_j >^2 \right)^2 \right]^{1/2} . \tag{1.60}$$

These properties will play a significant part in the statistics of autoregressive processes on H.

Finally, we indicate a characterization of covariance operators.

THEOREM 1.7 *An operator $C : H \to H$ is a covariance operator if and only if it is symmetric positive and nuclear.*

Concerning **cross-covariance operators** we have

$$C_{X,Y}(x) = C^*_{Y,X}(x) = E(< X, x > Y), \quad x \in H, \tag{1.61}$$

$X, Y \in L^2_H(P), \quad EX = EY = 0.$

These operators are nuclear and

$$\| C_{X,Y} \|_{\mathcal{N}} = \| C_{Y,X} \|_{\mathcal{N}} \leq E(\| X \| \, \| Y \|). \tag{1.62}$$

The characteristic functional may be written as

$$\varphi_X(x) = E(\exp i < x, X >) , \quad x \in H, \tag{1.63}$$

and if X is Gaussian as

$$\varphi_X(x) = \exp(i < x, EX > -\frac{1}{2} < C_X(x), x >), \quad x \in H. \tag{1.64}$$

Contrary to the general case, every covariance operator C can be associated to a Gaussian random variable X such that $C_X = C$.

1.6 Linear prediction in Hilbert spaces

The problem of linear approximation of a nonobserved random variable X by a linear function of observed random variables $(X_i,\ i \in I)$ has a simple and well known statement in a finite-dimensional setting:

If $X \in L^2(\Omega, \mathcal{A}, P)$ and $X_i \in L^2(\Omega, \mathcal{A}, P)$, $i \in I$, are zero-mean, the best linear approximation of X is its orthogonal projection over the smallest closed subspace of $L^2(\Omega, \mathcal{A}, P)$ containing $(X_i,\ i \in I)$. This subspace is the closure of

$$\mathrm{sp}\ (X_i,\ i \in I) = \left\{ \sum_{i \in J} a_i X_i\,,\ J \text{ finite } \subset I\,,\ a_i \in \mathbb{R},\ i \in J \right\}.$$

If the variables are in $L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$, the usual procedure is to consider the closed subspace generated by the components of the observed random vectors and then to project each component of the nonobserved random vector.

More generally, in an infinite-dimensional Hilbert space it is convenient to project over a rich enough subspace of $L^2_H(\Omega, \mathcal{A}, P)$. In this context, we introduce the notion of **$\mathcal{L}$-closed subspace** (LCS) (or **hermetically closed subspace**) in $L^2_H(\Omega, \mathcal{A}, P)$.

DEFINITION 1.1 *$\mathcal{G}$ is said to be an* **$\mathcal{L}$-closed subspace** *(or* **hermetically closed subspace***) of $L^2_H(\Omega, \mathcal{A}, P)$ if*

1) $\mathcal{G}$ is a Hilbertian subspace of $L^2_H(\Omega, \mathcal{A}, P)$.

2) If $X \in \mathcal{G}$ and $\ell \in \mathcal{L}$, then $\ell(X) \in \mathcal{G}$.

$\mathcal{G}$ is said to be a zero-mean LCS if it contains only zero-mean H-random variables.

We now study some properties of LCS's

THEOREM 1.8 *Let F be a subset of $L^2_H(P)$. Then the LCS $\mathcal{G}_F$ generated by F is the closure of*

$$\mathcal{G}'_F = \left\{ \sum_{i=1}^{k} \ell_i(X_i),\ \ \ell_i \in \mathcal{L},\ \ X_i \in F,\ \ i = 1, \ldots, k,\ \ k \geq 1 \right\}. \qquad (1.65)$$

Proof

Observe that an intersection of LCS's is again an LCS; thus $\mathcal{G}_F$ does exist.

Now we clearly have

$$F \subset \mathcal{G}'_F \subset \mathcal{G}_F,$$

and since $\mathcal{G}_F$ is closed we get

$$\overline{\mathcal{G}'_F} \subset \mathcal{G}_F.$$

Now, if $Y \in \overline{\mathcal{G}'_F}$ and $\varepsilon > 0$ is given, there exists $Z(\varepsilon) \in \mathcal{G}'_F$ such that $\| Y - Z(\varepsilon) \|_{L^2_H(P)} < \varepsilon$.

Then, if $\ell \in \mathcal{L}$, it follows that

$$\| \ell(Y) - \ell(Z(\varepsilon)) \|_{L^2_H(P)} \leq \| \ell \|_{\mathcal{L}} \varepsilon.$$

Noting that $\ell(Z(\varepsilon)) \in \mathcal{G}'_F$, we may assert that for every $\ell \in \mathcal{L}$, $\ell \neq 0$, $\varepsilon > 0$, and $Y \in \overline{\mathcal{G}'_F}$, there exists $Z' = \ell\left(Z\left(\varepsilon/_{\|\ell\|_{\mathcal{L}}}\right)\right)$ such that

$$\| \ell(Y) - Z' \|_{L^2_H(P)} < \varepsilon.$$

Hence $\ell(Y) \in \overline{\mathcal{G}'_F}$ and $\overline{\mathcal{G}'_F}$ is an LCS. Finally, $\overline{\mathcal{G}'_F} = \mathcal{G}_F$. ■

In order to show usefulness of hermetically closed subspaces we define two notions of orthogonality in $L^2_H(P)$.

DEFINITION 1.2 *Let $X, Y \in L^2_H(P)$ be such that $EX = EY = 0$.*

1) X and Y are said to be **weakly orthogonal** *($X \perp Y$) if*

$$E < X, Y >= 0.$$

2) X and Y are said to be **orthogonal** *($X \perp\!\!\!\perp Y$) if, for all $x, y \in H$,*

$$E(< x, X >< y, Y >) = 0$$

or, equivalently, if

$$C_{X,Y} = 0.$$

It is easy to verify that orthogonality is strictly stronger than weak orthogonality and that independence is strictly stronger than orthogonality: $\perp\!\!\!\perp \Rightarrow \perp\!\!\!\perp \Rightarrow \perp$.

Consider the special case of a zero-mean Gaussian $H \times H$-random vector, say, $Z = (X, Y)$. Then we have

$$X \perp\!\!\!\perp Y \Rightarrow X \perp\!\!\!\perp Y. \tag{1.66}$$

This can be checked directly by noting that $\varphi_{(X,Y)}(x, y) = \varphi_X(x)\varphi_Y(y)$; $x, y \in H$.

However, $X \perp Y$ does not imply $X \perp\!\!\!\perp Y$, as shown by the following elementary counterexample.

Let ξ and η be two standard Gaussian real random variables, and set

$$X = \xi e_1 + \eta e_2$$
$$Y = -\eta e_1 + \xi e_2,$$

where $\{e_1, e_2\}$ is an orthonormal system in H. Then $X \perp Y$ when $X \not\perp\!\!\!\perp Y$, since $E(< X, e_1 >< Y, e_2 >) = 1$.

The following result shows the importance of LCS's with respect to orthogonality.

THEOREM 1.9 *Let X be a zero-mean element of $L_H^2(P)$ and $\mathcal{G}$ an LCS. Then $X \perp \mathcal{G}$ implies $X \boldsymbol{\perp} \mathcal{G}$.*

Proof

$X \perp \mathcal{G}$ if and only if

$$E(< X, Z >) = 0\,, \quad Z \in \mathcal{G}. \tag{1.67}$$

Since $\mathcal{G}$ is an LCS, $Z \in \mathcal{G}$ implies $\ell(Z) \in \mathcal{G}$ for all $\ell \in \mathcal{L}$. Then, by (1.67),

$$E < X, \ell(Z) >= 0 \;\;, \quad Z \in \mathcal{G},\ \ell \in \mathcal{L}. \tag{1.68}$$

In particular, one may choose $\ell =< \cdot, z > x$, where x and z are fixed elements in H. Then (1.68) gives

$$E(< X, x >< Z, z >) = 0 \;\;, \quad x, z \in H,$$

that is, $X \boldsymbol{\perp} Z$. ∎

The next statement provides a useful property of LCS's.

THEOREM 1.10 *The orthogonal complement of an LCS is also an LCS.*

Proof

Let $\mathcal{G}$ be an LCS and $\mathcal{G}'$ its orthogonal complement. If $X \in \mathcal{G}'$, $Z \in \mathcal{G}$, and $\ell \in \mathcal{L}$, we have $X \perp Z$ and, since $\mathcal{G}$ is LCS, $X \perp \ell^*(Z)$. Then for all Z in $\mathcal{G}$

$$E(< \ell(X),\ Z >) = E(< X, \ell^*(Z) >) = 0\,, \quad \ell \in \mathcal{L},\ X \in \mathcal{G}'. \tag{1.69}$$

Thus $Z \perp \ell(X)$; hence $\ell(X) \in \mathcal{G}'$ and $\mathcal{G}'$ is an LCS. ∎

Finally, note that the LCS $\mathcal{G}_{\mathcal{X}}$ generated by a family $\mathcal{X} = (X_i,\ i \in I)$ of H-random variables is, in general, strictly greater than the closed subspace $\mathcal{V}_{\mathcal{X}}$ generated by $\mathcal{X}$.

However if $\mathcal{X} = \mathcal{X}_c$ is the family of constant H-random variables, then $\mathcal{X}_c$, $\mathcal{V}_{\mathcal{X}_c}$ and $\mathcal{G}_{\mathcal{X}_c}$ coincide and are isomorphic to H.

Application to linear prediction

The problem of linear prediction in a Hilbert space is the main motivation for introducing LCS's. Now we may give a natural definition.

DEFINITION 1.3 *Let X be a zero-mean element in $L_H^2(P)$ and $\mathcal{G}$ a zero-mean LCS. The orthogonal projection of X on $\mathcal{G}$ is called* **Prediction of X given $\mathcal{G}$** *or* **$\mathcal{G}$-Prediction of X**.

From Theorem 1.9 we infer that weak orthogonality and orthogonality coincide for $\mathcal{G}$. Thus, if $\pi^{\mathcal{G}}$ denotes the orthogonal projector of $\mathcal{G}$, we have the following characterization:

$$\begin{cases} \pi^{\mathcal{G}}(X) \in \mathcal{G} \\ C_{X-\pi^{\mathcal{G}}(X),\, Z} = 0, \ Z \in \mathcal{G} \end{cases} . \tag{1.70}$$

In the general case where X and $\mathcal{G}$ are not necessarily zero-mean, one may observe that

$$\mathcal{G} = \mathcal{G}_0 \oplus \mathcal{G}_{\mathcal{X}_c},$$

where $\mathcal{G}_0$ is the zero-mean LCS generated by $\sum_{i=1}^{k} \ell_i(Z_i - EZ_i)$; $\ell_1, \ldots, \ell_k \in \mathcal{L}$; $Z_1, \ldots Z_k \in \mathcal{G}$; $k \geq 1$. (see Theorem 1.8).

Then we have

$$\pi^{\mathcal{G}} = \pi^{\mathcal{G}_0} + \pi^{\mathcal{G}_{\mathcal{X}_c}}$$

and since $\pi^{\mathcal{G}_{\mathcal{X}_c}} = E$ (expectation!) we get

$$\begin{aligned} \pi^{\mathcal{G}}(X) &= \pi^{\mathcal{G}_0}(X) + EX \\ &= \pi^{\mathcal{G}_0}(X - EX) + EX. \end{aligned}$$

Finally, we consider the **Gaussian case**. For the sake of simplicity we only consider (X, Y), a zero-mean Gaussian $H \times H$-random variable, and $\mathcal{G} = \mathcal{G}_Y$, the LCS generated by Y. On the other hand, let $\mathcal{A}_Y$ be the σ-algebra generated by Y. Then we have

THEOREM 1.11

$$E^{\mathcal{A}_Y}(X) = \pi^{\mathcal{G}_Y}(X). \tag{1.71}$$

Proof

Theorem 1.9 entails that $C_{X-\pi^{\mathcal{G}}(X),Y} = 0$. Now $(X - \pi^{\mathcal{G}_Y}(X), Y)$ is clearly Gaussian and, using (1.66), we conclude that $X - \pi^{\mathcal{G}_Y}(X) \perp\!\!\!\perp Y$. Therefore $X - \pi^{\mathcal{G}_Y}(X) \perp\!\!\!\perp \mathcal{A}_Y$ and $E^{\mathcal{A}_Y}(X - \pi^{\mathcal{G}_Y}(X)) = 0$; hence (1.71) holds. ■

The above result remains valid if Y is an H^p-valued random variable.

NOTES

1.1 A general reference for stochastic processes theory is still Doob (1953). More recent are Revuz and Yor (1991), Karatzas and Shreve (1991), and many others. Sobczyk (1991) gives a comprehensive treatment of processes and stochastic differential equations with applications.

Existence and construction of Wiener processes and Ornstein-Uhlenbeck processes, and more generally of diffusion processes are discussed in all the books cited above.

Proof of Theorem 1.1 appears (for example) in Parthasarathy (1967) and proof of Theorem 1.2 in Ash and Gardner (AG) (1975).

1.2 Interpretation of stochastic processes as random variables in function spaces is standard and has been considered by many authors. Perhaps the text most interested in statistical applications of this representation is Grenander (1981).

Lemma 1.2 comes from Mourier (1953). A proof appears in Vakhania, Tarieladze, and Chobanyan (VTC) (1987). Theorem 1.3 is established in AG, Theorem 1.4 in Cramér and Leadbetter (1967), Lemma 1.3 in Mercer (1909) or Riesz and Nagy (1955). Theorem 1.5 comes from Karhunen (1946) and Loève (1948). Equation (1.28) appears in AG and (1.29) in Pugachev (1959).

1.3 Integration theory in Banach spaces arises in many works, in particular VTC or Ledoux and Talagrand (1991).

1.4 Covariance operators and characteristic functional in Banach spaces are considered in VTC. Proof of Theorem 1.6 is in that book.

1.5 The theory of operators in Hilbert space has been extensively studied. The results presented in this section are established in Guelfand and Vilenkin (1967) or Akhiezer and Glazman (1961). Theorem 1.7 is proved in VTC.

1.6 Details about linear prediction in $\mathbb{R}^d$ can be found in Brockwell and Davis (1991). The theory of $\mathcal{L}$-closed subspace's ("sous-espaces clos") is developed in Fortet (1995).

2

Sequences of Random Variables in Banach Spaces

This chapter discusses asymptotics for random variables in function spaces. The main statistical application will be asymptotic inference when data are curves.

In Section 1 we develop representation of processes as sequences of random variables in Banach spaces. This is different from the interpretation presented in Chapter 1, where a given stochastic process was associated with one B-random variable. That representation of **one** process as a **sequence** of random variables will be used extensively in the remainder of this book.

Section 2 deals with classical properties of stochastic convergences. Limit theorems for i.i.d. B-random variables appear in Section 3, while the case of dependent random variables is considered in Section 4. Some proofs are postponed to Section 5.

As in Chapter 1, proofs of some classical results are omitted.

2.1 Stochastic processes as sequences of B-valued random variables

In Chapter 1 we saw that a real stochastic process may be interpreted as a random variable in a suitable function space.

That representation must be adapted in order to deal with asymptotics. Actually if a process $\xi = (\xi_t,\ t \in \mathbb{R}_+)$ is observed over a time interval $[0, \tau]$ with τ tending to infinity one has to continually change the function space he/she works with.

In order to avoid this difficulty it is often interesting to use the following method: set $\tau = nh$, where n is an integer and h is a positive real coefficient, and suppose, for convenience, that the sample paths of ξ are continuous. Then, defining

$$X_k(t) = \xi_{kh+t} \ , \ \ 0 \leq t \leq h \ ; \ \ k = 0, 1, \ldots \tag{2.1}$$

one obtains a sequence $X_0, X_1, \ldots$ of random variables in $C[0, h]$. Note that $(X_0, \ldots, X_{n-1})$ contains the same information as $(\xi_t, \ 0 \leq t \leq \tau)$.

Clearly other function spaces adapted to regularity of sample paths may be employed. The choice of h is compelling in many concrete situations. If, for example, consumption of electricity in a town is observed continuously during one year, the time interval "one day" should be a convenient h. (See Example 9.3 in Chapter 9.)

Figure 1 in the Synopsis provides an example of representation (2.1). Observe that $X_0, X_1, \ldots$ are correlated since $X_0(h) = X_1(0), \ldots$

Other forms of this kind of representation may be considered, since it is possible to choose overlapping or disjoint intervals. Finally, the same description holds if data are independent copies of the same process $\xi = (\xi_t, \ t \in T)$, where T is a fixed time interval.

This sequential interpretation of stochastic processes is one of the leading strands of this book. It is used in the numerical applications that appear in Chapter 9.

2.2 Convergence of B-random variables

Let B be a separable Banach space with norm $\| \cdot \|$, Borel σ-algebra $\mathcal{B}_B$ and dual space B^*.

The classical stochastic convergences in B are defined as follows.

DEFINITION 2.1 *Let $(Y, Y_n, \ n \geq 1)$ be a family of B-valued random variables defined on a probability space $(\Omega, \mathcal{A}, P)$.*

1) (Y_n) converges to Y **almost surely**, *written $Y_n \underset{a.s.}{\longrightarrow} Y$, if*

$$P\left\{\omega : \| Y_n(\omega) - Y(\omega) \| \underset{n\to\infty}{\longrightarrow} 0 \right\} = 1.$$

2) (Y_n) converges to Y **in probability** $\left(Y_n \underset{p}{\longrightarrow} Y\right)$ *if for each $\varepsilon > 0$,*

$$P(\| Y_n - Y \| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0 \ .$$

3) (Y_n) converges to Y in $L^2_B(\Omega, \mathcal{A}, P)$ $\left(Y_n \underset{L^2_B(P)}{\longrightarrow} Y\right)$ if

$$E(\| Y_n - Y \|^2) \underset{n\to\infty}{\longrightarrow} 0 \, .$$

Clearly

$$Y_n \underset{L^2_B(P)}{\longrightarrow} Y \Rightarrow Y_n \underset{p}{\longrightarrow} Y \tag{2.2}$$

and

$$Y_n \underset{a.s.}{\longrightarrow} Y \Rightarrow Y_n \underset{p}{\longrightarrow} Y. \tag{2.3}$$

On the other hand, **weak convergence** deals with convergence of sequences of probability measures.

DEFINITION 2.2 and Portmanteau Theorem

Let $(\mu, \mu_n,\ n \geq 1)$ be a family of probability measures over $(B, \mathcal{B}_B)$. (μ_n) converges **weakly** *to μ $\left(\mu_n \underset{w}{\longrightarrow} \mu\right)$ if any of the five following equivalent conditions hold:*

1) $\int f d\mu_n \to \int f d\mu$ for all f in the class of bounded continuous real functions on B.

2) $\int f d\mu_n \to \int f d\mu$ for all bounded uniformly continuous real f.

3) $\mu_n(A) \to \mu(A)$ for all A in $\mathcal{B}_B$ such that $\mu(\partial A) = 0$ where, ∂A denotes the boundary of A.

4) $\limsup \mu_n(F) \leq \mu(F)$ for all closed sets F.

5) $\liminf \mu_n(G) \geq \mu(G)$ for all open sets G.

Now let $(Y, Y_n,\ n \geq 1)$ be a family of B-random variables. We say that (Y_n) converges to Y **in distribution** *or* **weakly** *$(Y_n \xrightarrow{\mathcal{D}} Y)$ if $P_{Y_n} \underset{w}{\longrightarrow} P_Y$, where P_{Y_n} (resp. P_Y) is the distribution of Y_n (resp. Y).*

The implication below holds.

$$Y_n \underset{p}{\longrightarrow} Y \Rightarrow Y_n \xrightarrow{\mathcal{D}} Y. \tag{2.4}$$

The converses of (2.2), (2.3), and (2.4) fail. However, we have some special converses. First we have a Levy theorem:

THEOREM 2.1 (Geffroy-Ito-Nisio (GIN) Theorem)

Let $(X_i,\ i \geq 1)$ be a sequence of B-valued independent random variables. Set $S_n = X_1 + \ldots + X_n$, $n \geq 1$. The following conditions are equivalent:

(1) (S_n) *converges weakly.*

(2) (S_n) *converges in probability.*

(3) (S_n) *converges almost surely.*

Another interesting result is the following representation theorem.

THEOREM 2.2 (Dudley-Skorokhod Theorem)

If $Y_n \xrightarrow{\mathcal{D}} Y$, *there exists a family* $(Z, Z_n,\ n \geq 1)$ *of* B*-random variables, defined on some probability space* $(\Omega', \mathcal{A}', P')$ *and such that*

1) $P_Z = P_Y,\ P_{Z_n} = P_{Y_n}$, *for all* $n \geq 1$.

2) $Z_n(\omega') \longrightarrow Z(\omega'),\ \omega' \in \Omega'$.

Let us now indicate some useful properties and characterizations of weak convergence over the set $\mathcal{P}$ of probability measures on $(B, \mathcal{B}_B)$.

1. Consider **Prokhorov's metric**, defined as

$$\delta(\mu,\nu) = \inf\left\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon,\ A \in \mathcal{B}_B\right\},\ \mu,\nu \in \mathcal{P}, \tag{2.5}$$

where $A^\varepsilon = \left\{x \in B,\ \inf_{y\in A} \| x - y \| < \varepsilon\right\}$. Then

$$\mu_n \xrightarrow[w]{} \mu \Leftrightarrow \delta(\mu_n, \mu) \to 0, \tag{2.6}$$

and $\mathcal{P}$ equipped with Prokhorov's metric is **separable** and **complete**.

2. **Prokhorov's Theorem**

Let $\Gamma \subset \mathcal{P}$. *Then* $\overline{\Gamma}$ *is compact with respect to the topology induced by* δ *if and only if* Γ *is* **tight**, *that is, for each positive* ε *there exists a compact set* K_ε *in* B *such that* $\mu(K_\varepsilon) > 1 - \varepsilon$ *for all* μ *in* Γ.

This classical result allows us to obtain criteria for weak convergence.

THEOREM 2.3 *Let* $(\mu, \mu_n,\ n \geq 1)$ *be a family of probability measures over* $(B, \mathcal{B}_B)$. *Then*

1) If $B = H$, *a separable Hilbert space,* $\mu_n \xrightarrow[w]{} \mu$ *if and only if* (μ_n) *is tight and*

$$\varphi_n(x) \to \varphi(x),\ x \in H, \tag{2.7}$$

where φ_n *and* φ *are the respective Fourier transforms of* μ_n *and* μ.

2) If $B = C[0,1]$, *then* $\mu_n \xrightarrow[w]{} \mu$ *if and only if* (μ_n) *is tight and*

$$\mu_n \pi^{-1}_{t_1,\ldots,t_k} \xrightarrow[w]{} \mu \pi^{-1}_{t_1,\ldots,t_k}, \tag{2.8}$$

$k \geq 1$, $t_1, \ldots, t_k \in [0,1]$, with

$$\pi_{t_1,\ldots,t_k}(x) = (x(t_1), \ldots, x(t_k)), \quad x \in C[0,1].$$

3. The following properties of convergence in distribution will be used in the remainder of the text:

$$Y_n \xrightarrow{\mathcal{D}} Y \text{ and } \| Y_n - Z_n \| \xrightarrow[p]{} 0 \text{ imply } Z_n \xrightarrow{\mathcal{D}} Y. \tag{2.9}$$

$$Y_n \xrightarrow{\mathcal{D}} Y \text{ and } Z_n \xrightarrow[p]{} z \text{ (a constant)} \tag{2.10}$$

imply $(Y_n, Z_n) \xrightarrow{\mathcal{D}} (Y, z)$;
where convergence takes place in $B \times B$.
Let B' be a separable Banach space, then

$$Y_n \xrightarrow{\mathcal{D}} Y, \; h : B \mapsto B' \text{ continuous imply } h(Y_n) \xrightarrow{\mathcal{D}} h(Y). \tag{2.11}$$

More generally, let $(h, h_n,\ n \geq 1)$ be a family of measurable mappings from B to B' and let N be the set of x in B such that $h_n(x_n) \to h(x)$ fails to hold for some sequence (x_n) converging to x. Then we have

$$Y_n \xrightarrow{\mathcal{D}} Y \text{ and } P(Y \in N) = 0 \text{ imply } h_n(Y_n) \xrightarrow{\mathcal{D}} h(Y). \tag{2.12}$$

2.3 Limit theorems for i.i.d. sequences of B-random variables

Let $(X_n,\ n \geq 1)$ be a sequence of i.i.d. B-random variables. In this section we study the asymptotic behavior of $S_n = X_1 + \ldots + X_n$ (or $\overline{X}_n = \frac{S_n}{n}$), $n \geq 1$, as n tends to infinity.

Let us begin with the **law of large numbers**.

THEOREM 2.4 (Strong law of large numbers)

Let $(X_n,\ n \geq 1)$ be a sequence of i.i.d. integrable B-random variables with expectation m. Then

$$\frac{S_n}{n} \xrightarrow[a.s.]{} m. \tag{2.13}$$

Proof

We may suppose that $m = 0$. Now, using Lemma 1.4 (Section 1.3), we may associate to each X_i and each $\varepsilon > 0$ a simple random variable

$$X_{i,\varepsilon} = f_\varepsilon(X_i) = \sum_{j=1}^{N_\varepsilon} x_j \mathbf{I}_{\{X_i \in B_j\}}$$

such that

$$E \parallel X_i - X_{i,\varepsilon} \parallel < \varepsilon.$$

By the scalar law of large numbers

$$\limsup_{n\to\infty} \parallel \frac{S_n}{n} - \frac{\sum_{i=1}^n X_{i,\varepsilon}}{n} \parallel \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \parallel X_i - X_{i,\varepsilon} \parallel < \varepsilon \quad a.s. \tag{2.14}$$

Now we have

$$\frac{1}{n}\sum_{i=1}^n X_{i,\varepsilon} = \sum_{j=1}^{N_\varepsilon} x_j \left(\frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{X_i \in B_j\}} \right), \; n \geq 1,$$

and, since for each j the i.i.d. Bernoulli random variables $\mathbf{I}_{\{X_i \in B_j\}}$, $i \geq 1$ satisfy the scalar law of large numbers, we obtain

$$\frac{1}{n}\sum_{i=1}^n X_{i,\varepsilon} \underset{a.s.}{\longrightarrow} \sum_{j=1}^{N_\varepsilon} x_j P(X_1 \in B_j) = EX_{1,\varepsilon}.$$

Noting that $\parallel EX_{1,\varepsilon} \parallel < \varepsilon$ and using (2.14) we get

$$\limsup \parallel \frac{S_n}{n} \parallel < 2\varepsilon \;\; a.s., \;\; \varepsilon > 0,$$

which entails (2.13). ■

We now present **large deviation inequalities**, which allow us to obtain a convergence rate in the law of large numbers. We first give a lemma that provides some useful bounds.

LEMMA 2.1 *Let $X_1, \ldots, X_n$ be independent B-valued random variables such that $Ee^{\lambda\|X_i\|} < \infty$, $1 \leq i \leq n$, for some $\lambda > 0$. Then for all $x_1, \ldots, x_n$ in B we have*

$$V \parallel S_n \parallel \leq \sum_{i=1}^n E \parallel X_i - x_i \parallel^2, \tag{2.15}$$

$$Ee^{\lambda\|S_n\|} \leq \exp\left[\lambda E \parallel S_n \parallel + \sum_{i=1}^n E\left[e^{\lambda\|X_i - x_i\|} - 1 - \lambda \parallel X_i - x_i \parallel\right]\right] \tag{2.16}$$

and

$$Ee^{\lambda\|S_n\|} \leq e^{\lambda E\|S_n\|} \prod_{i=1}^n E\left[e^{\lambda\|X_i - EX_i\|} - \lambda \parallel X_i - EX_i \parallel\right]. \tag{2.17}$$

If $B = H$, a separable Hilbert space and $EX_i = 0$, $1 \leq i \leq n$, then

$$E[ch(\lambda \parallel S_n \parallel)] \leq \prod_{i=1}^n E\left[e^{\lambda\|X_i\|} - \lambda \parallel X_i \parallel\right]. \tag{2.18}$$

Proof: see Section 5. ∎

The above bounds lead to some exponential-type inequalities in Banach spaces. Let us begin with the general case.

THEOREM 2.5 *Let $X_1, \ldots, X_n$ be independent B-random variables.*

1) If, for some positive constants $b = b(n)$ and $\ell = \ell(n)$ and some $x_1, \ldots, x_n$ in B, we have

$$\sum_{i=1}^{n} E \parallel X_i - x_i \parallel^k \leq \frac{k!}{2} \ell^2 b^{k-2} \; ; \; k = 2, 3, \ldots, \tag{2.19}$$

then for all $t > 0$

$$P(| \parallel S_n \parallel - E \parallel S_n \parallel | \geq t) \leq 2 \exp\left(-\frac{t^2}{2\ell^2 + 2bt}\right). \tag{2.20}$$

2) If $B = H$, a separable Hilbert space, and if (2.19) holds, with $x_i = EX_i = 0$, $1 \leq i \leq n$, then

$$P(\parallel S_n \parallel \geq t) \leq 2 \exp\left(-\frac{t^2}{2\ell^2 + 2bt}\right), \quad t > 0. \tag{2.21}$$

Proof: see Section 5. ∎

Note that if $X_1, \ldots, X_n$ are i.i.d., if $x_1 = \ldots = x_n = 0$, and if $E \parallel X_i \parallel^k \leq \frac{k!}{2} \ell_0^2 \, b_0^{k-2}$, $k \geq 2$, $1 \leq i \leq n$, then (2.19) is valid with $b = b_0$ and $\ell = \ell_0 \sqrt{n}$.

We now turn to the case of bounded variables.

THEOREM 2.6 *Let $X_1, \ldots, X_n$ be independent B-random variables.*

1) If $\parallel X_i - x_i \parallel \leq b$ a.s., $1 \leq i \leq n$, then

$$P(| \parallel S_n \parallel - E \parallel S_n \parallel | \geq t) \leq 2 \exp\left(-\frac{t^2}{2c_n^2 + (2/3)bt}\right), \tag{2.22}$$

$t > 0$, where $c_n^2 = \sum_{i=1}^{n} E \parallel X_i - x_i \parallel^2$.

2) If $\parallel X_i \parallel \leq b$ a.s. and $EX_i = 0$, $1 \leq i \leq n$, and if $B = H$, a separable Hilbert space, then

$$P(\parallel S_n \parallel \geq t) \leq 2 \exp\left(-\frac{t^2}{2c_n^2 + (2/3)bt}\right), \; t > 0, \tag{2.23}$$

where $c_n^2 = \sum_{i=1}^{n} E \parallel X_i \parallel^2$.

Proof: see Section 5. ∎

In a Hilbert space, (2.21) and (2.23) yield rates of convergence in the law of large numbers. As an example we have the following.

COROLLARY 2.1 *Let $(X_i,\ i \geq 1)$ be a sequence of H-valued i.i.d. random variables such that $\| X_i \| \leq b$, $E \| X_i \|^2 = \sigma^2$ and $EX_i = 0$. Then we have*

$$\limsup_{n\to\infty} \left(\frac{n}{\operatorname{Log} n}\right)^{1/2} \| \overline{X}_n \| \leq \sigma\sqrt{2} \ a.s. \tag{2.24}$$

Proof

In (2.23), choose $t = a\sigma(n \operatorname{Log} n)^{1/2}$, where $a > \sqrt{2}$. Apply Borel-Cantelli Lemma and then make $a \downarrow \sqrt{2}$. ∎

Note that the law of the iterated logarithm (see Theorem 2.10 below) gives a slightly more precise speed than (2.24).

In a general Banach space, the situation is more intricate: the rate will depend on some geometrical properties of B, which are introduced below.

Type and cotype of Banach space

First let us consider a finite sequence $X_1, \ldots, X_n$ of independent real random variables that are square integrable and zero-mean. We then have

$$E\left|\sum_{i=1}^{n} X_i\right|^2 = \sum_{i=1}^{n} E|X_i|^2. \tag{2.25}$$

This relation remains valid in a Hilbert space, using norm instead of absolute value, but does not extend to general Banach spaces. However, in some regular enough Banach spaces, properties weaker than (2.25) hold. We summarize them in the following.

DEFINITION 2.3 *Let B be a Banach space and $p \in]1,2]$ (resp. $q \in [2,\infty[$) a real number.*

B is of **type p** *(resp. of* **cotype q***) if there exists a strictly positive constant c (resp. c') such that for any finite sequence $X_1, \ldots, X_n$ of independent B-random variables such that $E \| X_i \|^p < \infty$ (resp. $E \| X_i \|^q < \infty$) and $EX_i = 0$, $1 \leq i \leq n$, the following inequality (2.26) (resp. (2.27)) holds.*

$$E \left\| \sum_{i=1}^{n} X_i \right\|^p \leq c \sum_{i=1}^{n} E \| X_i \|^p . \tag{2.26}$$

$$E \left\| \sum_{i=1}^{n} X_i \right\|^q \geq c' \sum_{i=1}^{n} E \| X_i \|^q . \tag{2.27}$$

Example 2.1
$L^p([0,1], \mathcal{B}_{[0,1]}, \lambda)$ has type 2 if $p \geq 2$ and type p if $1 < p \leq 2$. $C[0,1]$ has neither type nor cotype (except the trivial type $p = 1$ and, by convention, the trivial cotype $q = \infty$).

Clearly a Hilbert space has type and cotype 2. Conversely, a Banach space with type and cotype 2 is isomorphic to a Hilbert space.

Now if B has a type we have the following rate.

COROLLARY 2.2 *Let $(X_i,\ i \geq 1)$ be a sequence of B-valued i.i.d. random variables such that $\| X_i \| \leq b$, $E \| X_i \|^2 = \sigma^2$, and $EX_i = 0$.*

Suppose that B is a separable Banach space of type p. Then, if $p = 2$ (2.23) remains valid with cc_n^2 instead of c_n^2, if $1 < p < 2$ we have

$$\| \overline{X}_n \| = O\left(\frac{1}{n^{1-1/p}}\right) \quad a.s. \tag{2.28}$$

Proof

Since B has type p we may write

$$E \| \overline{X}_n \| \leq \frac{1}{n}(E \| S_n \|^p)^{1/p} \leq \frac{c^{1/p}}{n^{1-1/p}}(E \| X_1 \|^p)^{1/p}$$

and the result follows from (2.22) and Borel-Cantelli Lemma. ■

Central limit theorem (CLT)

Concerning the CLT we first deal with the case of a Hilbert space.

THEOREM 2.7 *Let $(X_i,\ i \geq 1)$ be a sequence of i.i.d. H-random variables, where H is a separable Hilbert space. If $E \|X_1\|^2 < \infty$, $EX_1 = m$, and $C_{X_1} = C$ then*

$$n^{-1/2} \sum_{i=1}^{n} (X_i - m) \xrightarrow{\mathcal{D}} N, \tag{2.29}$$

where $N \sim \mathcal{N}(0, C)$.

Proof

Let us set $Y_n = n^{-1/2} \sum_{i=1}^{n} (X_i - m)$, $n \geq 1$. For all x in H we have

$$< x, Y_n > = n^{-1/2} \sum_{i=1}^{n} [< x, X_i > - < x, m >],\ n \geq 1,$$

and the classical scalar CLT yields

$$< x, Y_n > \xrightarrow{\mathcal{D}} < x, N > .$$

It remains to show that (P_{Y_n}) is tight. To this aim we consider an orthonormal basis (v_j) of H composed of eigenvectors of C. We have

$$C(v_j) = \lambda_j v_j, \ \ j = 1, 2, \ldots,$$

where the eigenvalues λ_j satisfy $\sum_{j=1}^{\infty} \lambda_j < \infty$ (see Section 1.5).

Then, since $C_{Y_n} = C$ it follows that

$$r_N^2 := E\left(\sum_{j=N}^{\infty} < Y_n, v_j >^2\right) = \sum_{j=N}^{\infty} \lambda_j.$$

Now, given $\varepsilon > 0$, we consider the compact set $K_\varepsilon = \bigcap_{k=1}^{\infty} B_k$, where

$$B_k = \left\{x \in H : \sum_{j=N_k}^{\infty} < x, v_j >^2 \leq \ell_k^{-1}\right\},$$

with $1 = N_1 < N_2 < \ldots < N_k \ldots; \ell_k \to \infty$, and $\sum_{k=1}^{\infty} \ell_k r_{N_k} < \varepsilon$.

Then we may write

$$1 - P_{Y_n}(K_\varepsilon) \leq \sum_{k=1}^{\infty} P_{Y_n}(B_k^c) \leq \sum_{k=1}^{\infty} \ell_k r_{N_k}^2 < \varepsilon, \ \ n \geq 1,$$

which proves the tightness of (P_{Y_n}). Applying Theorem 2.2, we obtain asymptotic normality for (Y_n). ■

The problem of CLT in a general Banach space is much more difficult. First we give an example which shows that existence of $E \parallel X_1 \parallel^2$ is not sufficient for the CLT to hold.

Example 2.2
Let c_0 be the space of real sequences tending to 0. It is a Banach space with respect to the sup-norm.

Consider a family $(\xi_{ji}, \ i \geq 1, \ j \geq 1)$ of real independent random variables such that

$$P(\xi_{ji} = 1) = P(\xi_{ji} = 0) = P(\xi_{ji} = -1) = \frac{1}{3}, \ i \geq 1, \ j \geq 1.$$

Let us set

$$X_i = (a_j \xi_{ji}, \ j \geq 1), \ \ i \geq 1,$$

where $a_j = \left(\frac{2}{3}\mathrm{Log}(j+1)\right)^{-1/2}, \ \ j \geq 1.$

Then (X_i) is a sequence of c_0-valued random variables. Let us show that this sequence does not satisfy the CLT. Suppose that

$$n^{-1/2}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N = (N_j,\ j \geq 1).$$

Then (2.11) gives

$$a_j n^{-1/2}\sum_{i=1}^{n} \xi_{ji} \xrightarrow{\mathcal{D}} N_j \sim \mathcal{N}(0, (\operatorname{Log}(j+1))^{-1}),\ j \geq 1,$$

but we have

$$P(|N_j| \geq \varepsilon) \leq \frac{2(\operatorname{Log}(j+1))^{1/2}}{\sqrt{2\pi}} \int_{\varepsilon}^{+\infty} e^{-\frac{x^2}{2}\operatorname{Log}(j+1)} dx,\ \varepsilon > 0,\ j \geq 1.$$

As j tends to infinity, the bound is equivalent to $\dfrac{1}{\varepsilon\sqrt{2\pi}} j^{-\varepsilon^2/2}(\operatorname{Log} j)^{-1/2}$.

Then, choosing $\varepsilon \leq \sqrt{2}$ and applying Borel-Cantelli Lemma, we obtain $P\left((N_j) \underset{j\to\infty}{\longrightarrow} 0\right) = 0$, which proves that $P(N \notin c_0) = 1$. ■

The following statement provides central limit theorems in some Banach spaces.

THEOREM 2.8 *Let $X = (X_i,\ i \geq 1)$ be a sequence of B-valued i.i.d. random variables.*

1) *If B is of type 2 and $E \| X_i \|^2 < \infty$, then X satisfies the CLT.*

2) *If B is of cotype 2 and if X_1 is pregaussian (i.e. there exists a B-valued Gaussian random variable Y such that $C_Y = C_{X_1}$), then X satisfies the CLT.*

3) *If $B = C[0,1]$ and*

$$|X_1(t) - X_1(s)| \leq M_1|t-s|,\ 0 \leq s,\ t \leq 1, \tag{2.30}$$

where M_1 is a positive random variable such that $EM_1^2 < \infty$, then X satisfies the CLT.

Example 2.3

Let us set $X_i(t) = \xi_i \cos 2\pi t + \eta_i \sin 2\pi t$, $0 \leq t \leq 1$, $i \geq 1$, where ξ_1, η_1, ξ_2, $\eta_2, \ldots$ is a family of i.i.d. real random variables such that $E\xi_i^2 = E\eta_I^2 = \sigma^2 < \infty$, $\quad i \geq 1$.

Then (X_i) is an i.i.d. sequence of $C[0,1]$-random variables and

$$|X_1(t) - X_1(s)| \leq 2\pi(|\xi_1| + |\eta_1|)|t-s|,\ 0 \leq s,\ t \leq 1;$$

thus (2.30) is satisfied and the CLT holds. ■

In a Hilbert space, accuracy of normal approximation is quantified by a **Berry-Esseen type bound**:

THEOREM 2.9 *If $X_1, \ldots, X_n$ are H-valued i.i.d. random variables such that $E \parallel X_1 \parallel^3 < \infty$ and $EX_1 = 0$, then*

$$\sup_{r \geq 0} |P(\parallel \sqrt{n}\ \overline{X}_n \parallel \leq r) - P(\parallel N \parallel \leq r)| \leq \frac{\gamma}{\sqrt{n}}, \tag{2.31}$$

where $N \sim \mathcal{N}(0, C_{X_1})$ and γ is a real coefficient that depends only on $E \parallel X_1 \parallel^3$ and C_{X_1}.

Again the situation is more complicated in a general Banach space. It can be shown that the rate $n^{-1/6}$ is attainable in some spaces (such as $C[0,1]$) and that it cannot be improved.

Law of the iterated logarithm (LIL)

We finally state the LIL in a Banach space. Here Log u is the maximum of 1 and the usual logarithm of u. On the other hand

$$d(x, A) = \inf(\parallel x - y \parallel,\ y \in A),\ x \in B,\ A \subset B,$$

and $c(x_n)$ is the set of limit points of the sequence (x_n).

THEOREM 2.10 (Law of the iterated logarithm)

Let $(X_i,\ i \geq 1)$ be a sequence of i.i.d. zero-mean B-random variables, where B is a separable Banach space of type 2.

Suppose that $E\left(\frac{\parallel X_1 \parallel^2}{\text{Log Log} \parallel X_1 \parallel}\right) < \infty$ and that the random variables $x^*(X_1)$, $x^* \in B^*$, $\parallel x^* \parallel \leq 1$ are uniformly integrable. Then there exists a compact convex symmetric set K such that

$$\lim_{n \to \infty} d\left(\frac{S_n}{\sqrt{2n\text{Log Log}n}}, K\right) = 0 \ \ a.s. \tag{2.32}$$

and

$$c\left(\frac{S_n}{\sqrt{2n\text{Log Log}n}}\right) = K \ \ a.s. \tag{2.33}$$

Moreover,

$$K = \{x \in B:\ \textit{where}\ x = E(\xi X_1),\ \xi \in L^2(\Omega, \mathcal{A}, P),\ E\xi = 0,\ E\xi^2 \leq 1\}.$$

2.4 Sequences of dependent random variables in Banach spaces

The independence assumption considered above is often not suitable in practice. Actually situation as in Figure 1 appears as soon as one observes a continuous-time process over a long period of time.

In the remaining chapters we will study linear models adapted to such a situation. In this section we consider various kinds of correlation between B-random variables that will appear in the remainder of this book. Let us begin with a general law of large numbers.

THEOREM 2.11 *Let $(X_i, i \geq 1)$ be a sequence of zero-mean B-random variables where B is a separable Banach space.*

Suppose that there exist two real sequences $(f(p) : p \geq 0)$ and $(\varphi(m) : m \geq 0)$ such that

(a) $E \parallel X_n + \ldots + X_{n+p-1} \parallel^2 \leq f(p),\ p \geq 1,\ n \geq 1,$

(b) $f(p) = o(p^2)$ *as* $p \to \infty,$

(c) $(\varphi(m))$ *is an increasing sequence of integers and* $\varphi(0) = 0,$

(d) $\sum_{m=1}^{\infty} \left[\frac{\varphi(m+1)}{\varphi(m)} - 1 \right]^2 < \infty,$ *and*

(e) $\sum_{m=1}^{\infty} \frac{f[\varphi(m)]}{\varphi^2(m)} < \infty.$

Then

$$\frac{S_n}{n} \longrightarrow 0$$

almost surely and in $L^2_B(P)$.

Proof

Note first that convergence in $L^2_B(P)$ is clear since (a) and (b) imply

$$\limsup_{n\to\infty} E \parallel \frac{S_n}{n} \parallel^2 \leq \limsup_{n\to\infty} \frac{f(n)}{n^2} = 0.$$

To prove almost sure convergence, we may apply Tchebychev inequality and (a) to obtain

$$P\left(\parallel \frac{S_{\varphi(m)}}{\varphi(m)} \parallel \geq \eta \right) \leq \frac{E \parallel S_{\varphi(m)} \parallel^2}{\eta^2 \varphi^2(m)} \leq \frac{1}{\eta^2} \frac{f[\varphi(m)]}{\varphi^2(m)},\ \eta > 0.$$

Thus (e) and the Borel-Cantelli Lemma yield

$$\parallel \frac{S_{\varphi(m)}}{\varphi(m)} \parallel \underset{a.s.}{\longrightarrow} 0 \quad \text{as } m \to \infty. \tag{2.34}$$

Now let us set

$$Y_m = \max_{\varphi(m) < n \leq \varphi(m+1)} \parallel \frac{S_n - S_{\varphi(m)}}{n} \parallel^2,\ \ m \geq 1.$$

Noting that

$$Y_m \leq \frac{1}{\varphi^2(m)}(\| X_{\varphi(m)+1} \| + \ldots + \| X_{\varphi(m+1)} \|)^2,$$

we have

$$EY_m \leq \frac{1}{\varphi^2(m)} \left[\sum_{i=1}^{\varphi(m+1)-\varphi(m)} \left(E \| X_{\varphi(m)+i} \|^2 \right)^{1/2} \right]^2 .$$

Then, applying (a) for $p = 1$, it follows that

$$EY_m \leq \frac{1}{\varphi^2(m)}[\varphi(m+1) - \varphi(m)]^2 f(1).$$

Now the Markov inequality, assumption (d), and, Borel-Cantelli Lemma give

$$Y_m \underset{a.s.}{\longrightarrow} 0 \text{ as } m \to \infty.$$

Using (c) we may associate to each integer $n \geq 1$ an integer m_n such that

$$\varphi(m_n) < n \leq \varphi(m_n + 1).$$

Since $n \to \infty$ entails $m_n \to \infty$, we obtain

$$Y_{m_n} \underset{a.s.}{\longrightarrow} 0 \text{ as } n \to \infty. \tag{2.35}$$

Finally, consider the relation

$$\frac{S_n}{n} = \frac{S_n - S_{\varphi(m_n)}}{n} + \frac{S_{\varphi(m_n)}}{\varphi(m_n)} \frac{\varphi(m_n)}{n}.$$

It implies

$$\| \frac{S_n}{n} \| \leq Y_{m_n}^{1/2} + \| \frac{S_{\varphi(m_n)}}{\varphi(m_n)} \|,$$

and from (2.34) and (2.35) it follows that

$$\frac{S_n}{n} \underset{a.s.}{\longrightarrow} 0,$$

which completes the proof. ■

Example 2.4
If (X_i) satisfies (a) with f defined as $f(0) = 0$, $f(1) = c$, $f(p) = cp^2(\text{Log } p)^{-\delta}$, $p \geq 2$, where $\delta > 2$, then it suffices to choose $\varphi(0) = 0$ and $\varphi(m) = [\exp(m^\alpha)]$, $m \geq 1$, with $\frac{1}{\delta} < \alpha < \frac{1}{2}$, to obtain (c), (d) and (e).

Note that the sequence f defined above leads to a nearly minimal condition since, if $X_n = X_0$ for all n with $0 < E \parallel X_0 \parallel^2 < \infty$ and $EX_0 = 0$, we have $E \parallel X_n + \ldots + X_{n+p-1} \parallel^2 = p^2 E \parallel X_0 \parallel^2$ and the law of large numbers does not hold.

A rate of convergence appears in the following corollary.

COROLLARY 2.3 *Let $(X_i,\ i \geq 1)$ be a sequence of zero-mean B-random variables such that*

$$E \parallel X_n + \ldots + X_{n+p-1} \parallel^2 \leq cp^\gamma,\ p \geq 1,\ n \geq 1, \tag{2.36}$$

where $c > 0$ and $\gamma \in]0, 2[$ are constants.

Then, for all $\beta > \dfrac{1}{2}$,

$$\frac{n^{(2-\gamma)/4}}{(\text{Log } n)^\beta} \parallel \frac{S_n}{n} \parallel \underset{a.s.}{\longrightarrow} 0. \tag{2.37}$$

Proof

Consider the sequence of integers defined as $\nu_m = [m^{2/(2-\gamma)}]$, $m \geq 1$. From Tchebychev inequality and (2.36) it follows that, for all $\eta > 0$,

$$P\left(\frac{\nu_m^{(2-\gamma)/4}}{(\text{Log } \nu_m)^\beta} \parallel \frac{S_{\nu_m}}{\nu_m} \parallel \geq \eta\right) = O\left(\frac{1}{m(\text{Log } m)^{2\beta}}\right).$$

Thus, if $\beta > \dfrac{1}{2}$, the Borel-Cantelli Lemma entails

$$\frac{\nu_m^{(2-\gamma)/4}}{(\text{Log}\nu_m)^\beta} \parallel \frac{S_{\nu_m}}{\nu_m} \parallel \underset{a.s.}{\longrightarrow} 0 \text{ as } m \to \infty. \tag{2.38}$$

Now for large enough m we have $\nu_{m+1} > \nu_m$ and we may set

$$Y_m = \max_{\nu_m < n \leq \nu_{m+1}} \frac{\parallel S_n - S_{\nu_m} \parallel^2}{n^{(2+\gamma)/2}(\text{Log } n)^{2\beta}}.$$

Using arguments similar to those in the proof of Theorem 2.11, we obtain the bound

$$EY_m \leq \frac{c(\nu_{m+1} - \nu_m)^2}{\nu_m^{(2+\gamma)/2}(\text{Log}\nu_m)^{2\beta}},$$

which implies

$$EY_m \leq \frac{c'm}{(\text{Log}\nu_m)^{2\beta}} \left(\frac{\nu_{m+1}}{\nu_m} - 1\right)^2,$$

where c' is constant.

Now for large enough m we have

$$\frac{\nu_{m+1}}{\nu_m} - 1 \leq \frac{4}{m(2-\gamma)};$$

therefore

$$EY_m = O\left(\frac{1}{m(\text{Log}m)^{2\beta}}\right),$$

so

$$Y_m \underset{a.s.}{\longrightarrow} 0 \text{ as } m \to \infty.$$

Finally, if $n \to \infty$, we have $Y_{m(n)} \underset{a.s.}{\longrightarrow} 0$, where $m(n)$ is defined by $\nu_{m(n)} < n \leq \nu_{m(n)+1}$. Hence

$$\frac{S_n - S_{\nu_{m(n)}}}{n^{(2+\gamma)/4}(\text{Log}n)^{\beta}} \underset{a.s.}{\longrightarrow} 0 \tag{2.39}$$

and (2.37) follows from (2.38) and (2.39). ■

More precise convergence rates can be obtained in some special cases. Among them we will consider strongly mixing processes and martingale differences.

Strong mixing

Let $X = (X_n,\ n \in \mathbb{Z})$ be a B-valued random process. Consider the σ-algebras $\mathcal{P}_n = \sigma(X_i,\ i \leq n)$ and $\mathcal{F}_m = \sigma(X_i,\ i \geq m)$, $n, m \in \mathbb{Z}$, respectively associated with the past relative to n and the future relative to m. Correlation between past and future may be quantified by the **strong mixing coefficients** defined as

$$\alpha_k = \sup_{n \in \mathbb{Z}} \sup_{\substack{B \in \mathcal{P}_n \\ C \in \mathcal{F}_{n+k}}} |P(B \cap C) - P(B)P(C)|,\ k \geq 1. \tag{2.40}$$

If $\lim\limits_{k \to \infty} \alpha_k = 0$, X is said to be **strongly mixing**. If $\alpha_k = O(r^k)$, where $0 < r < 1$, X is said to be **geometrically strongly mixing** (GSM).

Example 2.5

If $B = \mathbb{R}$ and if $X_n = \sum\limits_{j=0}^{\infty} a_j \varepsilon_{n-j}$, $n \in \mathbb{Z}$, is a linear process (see Example 1.3) such that $|a_j| = O(s^j)$, where $0 < s < 1$, and (ε_n) is a strong white noise such that ε_n has a density, then (X_n) is GSM. In particular, if (X_n) is a strictly stationary ARMA process associated with a strong white noise (ε_n) where ε_n has a density, then (X_n) is GSM.

We now give large deviation inequalities that depend on strong mixing coefficients. Let us begin with an inequality for real stochastic processes.

LEMMA 2.2 *Let* $(\xi_i,\ i \in \mathbb{Z})$ *be a zero-mean real stochastic process such that* $\sup_{i\in\mathbb{Z}} \| \xi_i \|_\infty \leq b < \infty$ *and with mixing coefficients* $(\alpha_k,\ k \geq 1)$. *Then*

$$P\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n}\right| > \varepsilon\right) \leq 4\exp\left(-\frac{\varepsilon^2}{8b^2}q\right) + 22\left(1 + \frac{4b}{\varepsilon}\right)^{1/2} q\alpha_{[n/2q]}, \tag{2.41}$$

$n \geq 2,\ q \in \left\{1, \ldots, \left[\frac{n}{2}\right]\right\},\ \varepsilon > 0.$

Now as a consequence we obtain an inequality for H-valued processes.

THEOREM 2.12 *Let* $X = (X_i,\ i \in \mathbb{Z})$ *be a zero-mean* H*-valued stochastic process, where* H *is a separable Hilbert space. Suppose that* $\sup_{i\in\mathbb{Z}} \| X_i \|_\infty \leq b < \infty$ *and that the covariance operator* C_{X_i} *of* X_i *does not depend on* i. *Denote by* $(\alpha_k,\ k \geq 1)$ *the strong mixing coefficients of* X *and by* $(\lambda_j,\ j \geq 1)$ *the eigenvalues of* C_{X_0}. *Then*

$$\begin{aligned} P\left(\| \frac{S_n}{n} \| > \varepsilon\right) \quad &\leq 4\nu \exp\left(-\frac{(1-\delta)\varepsilon^2}{8\nu b^2}q\right) \\ &+22\nu\left(1 + \frac{4b\sqrt{\nu}}{\varepsilon(1-\delta)^{1/2}}\right)^{1/2} q\alpha_{[n/2q]} + \frac{1}{\delta\varepsilon^2}\sum_{j>\nu}\lambda_j, \end{aligned} \tag{2.42}$$

$\nu = 1, 2, \ldots;\ n = 2, 3, \ldots;\ q = 1, 2, \ldots, \left[\frac{n}{2}\right];\ \varepsilon > 0;\ \delta \in]0, 1[.$

Proof

Let $(v_j,\ j \geq 1)$ be an orthonormal basis of H formed with eigenvectors of C_{X_0}. For every $\delta \in]0, 1[$ we may write

$$\begin{aligned} P\left(\| \frac{S_n}{n} \| > \varepsilon\right) \quad &\leq P\left(\sum_{j=1}^{\nu} < \frac{S_n}{n},\ v_j >^2 \geq (1-\delta)\varepsilon^2\right) \\ &+P\left(\sum_{j>\nu} < \frac{S_n}{n},\ v_j >^2 \geq \delta\varepsilon^2\right). \end{aligned} \tag{2.43}$$

Now we have the bound

$$P\left(\sum_{j=1}^{\nu} < \frac{S_n}{n},\ v_j >^2 \geq (1-\delta)\varepsilon^2\right) \leq \sum_{j=1}^{\nu} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} < X_i, v_j >\right| > \frac{\varepsilon\sqrt{1-\delta}}{\sqrt{\nu}}\right).$$

Applying Lemma 2.2 to $(< X_i, v_j >,\ i \in \mathbb{Z})$ we obtain

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} < X_i, v_j >\right| > \frac{\varepsilon\sqrt{1-\delta}}{\sqrt{\nu}}\right) &\leq 4\exp\left(-\frac{(1-\delta)\varepsilon^2}{8\nu b^2}q\right) \\ +22\left(1+\frac{4b\sqrt{\nu}}{\varepsilon\sqrt{1-\delta}}\right)^{1/2} q\alpha_{[n/2q]}\ , &\quad 1 \leq j \leq \nu. \end{aligned} \tag{2.44}$$

On the other hand, the Markov inequality yields

$$P\left(\sum_{j>\nu} < \frac{S_n}{n},\ v_j >^2 \geq \delta\varepsilon^2\right) \leq \frac{1}{\delta\varepsilon^2}\sum_{j>\nu}\left[E < \frac{S_n}{n},\ v_j >^2\right].$$

Noting that $\left(\frac{1}{n}\sum_{i=1}^{n} < X_i, v_j >\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} < X_i, v_j >^2$, we get

$$\begin{aligned} P\left(\sum_{j>\nu} < \frac{S_n}{n},\ v_j >^2 \geq \delta\varepsilon^2\right) &\leq \frac{1}{\delta\varepsilon^2}\sum_{j>\nu}\left[E < X_0, v_j >^2\right] \\ &\leq \frac{1}{\delta\varepsilon^2}\sum_{j>\nu}\lambda_j. \end{aligned} \tag{2.45}$$

Finally, (2.42) follows from (2.43), (2.44), and (2.45). ■

COROLLARY 2.4 *If in addition there exists $a > 0$ and $r \in]0,1[$ such that*

$$\alpha_k \leq ar^k\ ,\quad k \geq 1, \tag{2.46}$$

and

$$\lambda_j \leq ar^j\ ,\quad j \geq 1, \tag{2.47}$$

then

$$P\left(\| \frac{S_n}{n} \| > \varepsilon\right) \leq k_1 \exp\left(-k_2 n^{1/3}\right),\ \varepsilon > 0,\ n \geq 1, \tag{2.48}$$

where k_1 and k_2 depend only on ε and P_X.

Moreover, as $n \to \infty$,

$$\| \frac{S_n}{n} \| = O\left(\frac{(\text{Log}\ n)^{3/2}}{n^{1/2}}\right)\quad a.s. \tag{2.49}$$

Proof

In (2.42) choose $q = [n^{2/3}]$, $\nu = [n^{1/3}]$, and $\delta \in]0,1[$; then (2.48) follows.

For (2.49), take $\varepsilon = A\frac{(\text{Log}\ n)^{3/2}}{n^{1/2}}$, $q_n = B\left(\left[\frac{n}{\text{Log} n}\right] + 1\right)$, and $\nu =$

$C[\text{Log } n]$, $n \geq 3$, where A, B, and C are positive constants. One then obtains

$$P\left(\frac{n^{1/2}}{(\text{Log } n)^{3/2}} \parallel \frac{S_n}{n} \parallel > A\right) \leq 4 \text{ Log } n \exp\left(-\frac{1-\delta}{8b^2}\frac{A^2B}{C}\text{Log } n\right)$$

$$+22\nu\left(1+\frac{4b\sqrt{C}}{\sqrt{1-\delta}}\frac{n^{1/2}}{\text{Log } n}\right)^{1/2} B \cdot 2\frac{n}{\text{Log } n} \cdot a\, e^{-\frac{1}{2B}\left(\text{Log } \frac{1}{r}\right) \text{Log } n}$$

$$+\frac{1}{\delta A^2}\frac{n}{(\text{Log } n)^3} a\, e^{-\text{Log } \frac{1}{r} C[\text{Log } n]} \frac{r}{1-r}.$$

By choosing $C = \dfrac{3}{\text{Log } \frac{1}{r}}, B = \dfrac{1}{5} \text{ Log } \dfrac{1}{r}$, and $A = \dfrac{11b}{\sqrt{1-\delta}\, \text{Log } \frac{1}{r}}$

we get $\displaystyle\sum_{n=1}^{\infty} P\left(\frac{n^{1/2}}{(\text{Log } n)^{3/2}} \parallel \frac{S_n}{n} \parallel > A\right) < \infty$ and hence (2.49) by the Borel-Cantelli Lemma. ■

We now turn to the general case, keeping the same notation.

THEOREM 2.13 *Let $X = (X_i,\ i \in \mathbb{Z})$ be a zero-mean H-valued process with equidistributed X_i's and such that $E\left(e^{\gamma\|X_0\|}\right) < \infty$ for some $\gamma > 0$. Then*

$$P\left(\parallel \frac{S_n}{n} \parallel > \varepsilon\right) \leq 4\nu \exp\left(-\frac{(1-\delta)^3\varepsilon^2}{128\nu m^2} q\right)$$

$$+22\nu\left(1+\frac{16m\sqrt{\nu}}{\varepsilon(1-\delta)^{3/2}}\right)^{1/2} q\alpha_{[n/2q]} + \frac{4}{\delta(1-\delta)^2\varepsilon^2}\sum_{j>\nu}\lambda_j + ne^{-\gamma m} E\left(e^{\gamma\|X_0\|}\right); \tag{2.50}$$

$\nu = 1, 2, \ldots$; $n \geq 2$; $q = 1, \ldots, \left[\frac{n}{2}\right]$; $m > \dfrac{2}{\gamma}$ Log $\dfrac{2c}{(1-\delta)\varepsilon}$; *where* $c = \left(E\left(e^{\gamma\|X_0\|}\right) E \parallel X_0 \parallel^2\right)^{1/2}$; $0 < \delta < 1$; $\varepsilon > 0$.

Proof

Let us set $Y_i = X_i\ \mathbf{1}_{\|X_i\| \leq m}$ and $Z_i = X_i\ \mathbf{1}_{\|X_i\| > m}$, $i \in \mathbb{Z}$. Then if $m >$

$\frac{2}{\gamma}\operatorname{Log}\frac{2c}{(1-\delta)\varepsilon}$, where $0<\delta<1$,

$$\| EY_i \| = \| EZ_i \| \leq E\left(\mathbf{I}_{\|X_i\|>m} \| X_i \|\right)$$

$$\leq (P(\| X_i \| > m))^{1/2}(E \| X_i \|^2)^{1/2}$$

$$\leq ce^{-\frac{\gamma}{2}m} < \frac{(1-\delta)\varepsilon}{2}, \quad \varepsilon > 0.$$

Therefore

$$P\left(\| \frac{1}{n}\sum_{i=1}^{n} Y_i \| > (1-\delta)\varepsilon\right) \leq P\left(\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i) \| > \frac{(1-\delta)\varepsilon}{2}\right).$$

Now we may use (2.42) under the special form

$$P\left(\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i) \| > \frac{(1-\delta)\varepsilon}{2}\right) \leq 4\nu \exp\left(-\frac{(1-\delta)^3\varepsilon^2}{32\nu(4m^2)}q\right)$$
$$+22\nu\left(1+\frac{16m\sqrt{\nu}}{\varepsilon(1-\delta)\sqrt{1-\delta}}\right)^{1/2} q\alpha'_{[n/2q]} + \frac{4}{\delta(1-\delta)^2\varepsilon^2}\sum_{j>\nu} E < Y_0 - EY_0, v_j >^2,$$

where $\alpha'_{[n/2q]}$ denotes strong mixing coefficient for $(Y_i - EY_i)$ and where (v_j) **is an orthonormal basis formed with eigenvectors of** C_{X_0}.

Now we clearly have $\alpha'_{[n/2q]} \leq \alpha_{[n/2q]}$ and

$$E\left(< Y_0 - EY_0, v_j >^2\right) = \operatorname{Var} < Y_0, v_j > \leq E\left(< Y_0, v_j >^2\right)$$

$$= E\left(\mathbf{I}_{\|X_0\|\leq m} < X_0, v_j >^2\right) \leq E < X_0, v_j >^2 = \lambda_j,\ j > \nu.$$

Hence

$$P\left(\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i) \| > \frac{(1-\delta)\varepsilon}{2}\right) \leq 4\nu \exp\left(-\frac{(1-\delta)^3\varepsilon^2}{128\nu m^2}q\right)$$
$$+22\nu\left(1+\frac{16m\sqrt{\nu}}{\varepsilon(1-\delta)\sqrt{1-\delta}}\right)^{1/2} q\alpha_{[n/2q]} + \frac{4}{\delta(1-\delta)^2\varepsilon^2}\sum_{j>\nu}\lambda_j. \tag{2.51}$$

On the other hand,

$$\| \frac{1}{n}\sum_{i=1}^{n} Z_i \| > \delta\varepsilon \Rightarrow \exists i \in \{1,\dots,n\} \ : \ \| X_i \| > m;$$

thus

$$P\left(\| \frac{1}{n}\sum_{i=1}^{n} Z_i \| > \delta\varepsilon\right) \leq n\, P(\| X_0 \| > m),$$

and hence

$$P\left(\| \frac{1}{n}\sum_{i=1}^{n} Z_i \| > \delta\varepsilon\right) \leq n\, e^{-\gamma m} E\left(e^{\gamma\|X_0\|}\right). \tag{2.52}$$

Finally, using (2.51) and (2.52), we arrive at (2.50). ■

COROLLARY 2.5 *If in addition (2.46) and (2.47) hold, then*

$$P\left(\| \frac{S_n}{n} \| > \varepsilon\right) \leq k_1 \exp\left(-k_2 n^{1/5}\right),\ \varepsilon > 0,\ n \geq 1, \tag{2.53}$$

where k_1 and k_2 depend only on ε and P_X.

Moreover, as $n \to \infty$,

$$\| \frac{S_n}{n} \| = O\left(\frac{(\mathrm{Log} n)^{5/2}}{n^{1/2}}\right) \quad a.s. \tag{2.54}$$

Proof

In (2.50) choose $q \sim n^{4/5}$, $\nu \sim n^{1/5}$, and $m \sim n^{1/5}$. Hence (2.53) holds.

Concerning (2.54) it suffices to take, in (2.50), $\varepsilon = A\dfrac{(\mathrm{Log}\ n)^{5/2}}{n^{1/2}}$, $q = B\left(\left[\dfrac{n}{\mathrm{Log} n}\right] + 1\right)$, $\nu = C[\mathrm{Log}\ n]$, and $m = D[\mathrm{Log}\ n]$, with suitable values of A, B, C, and D and then to apply the Borel-Cantelli Lemma. Details are left to the reader. ■

Martingale differences

Consider a **filtration** $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \ldots \subset \mathcal{A}_n \subset \ldots$ of sub-σ-algebras of $\mathcal{A}$ and a sequence $(X_i,\ i \geq 1)$ of integrable B-random variables.

(X_i) is a **martingale difference** with respect to $(\mathcal{A}_i)$ if it is **adapted** to $(\mathcal{A}_i)$ (i.e., X_i is $\mathcal{A}_i$-measurable for each i) and

$$E^{\mathcal{A}_{i-1}} X_i = 0,\ \ i \geq 1. \tag{2.55}$$

This definition extends to sequences indexed by $\mathbb{Z}$. Concerning martingale differences we have the following rate.

THEOREM 2.14 *Let $(X_i,\ i \geq 1)$ be a martingale difference in a separable Hilbert space H such that $\| X_i \| \leq b$ a.s., $i \geq 1$, and*

$$E^{\mathcal{A}_{i-1}} \| X_i \|^2 \leq d_i,\ \ i \geq 1, \tag{2.56}$$

where $d_1, d_2, \ldots$ are constants.

Then, for all $t > 0$ and all $n \geq 1$,

$$P(\| S_n \| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} d_i + (2/3)bt}\right) \tag{2.57}$$

and

$$\| \frac{S_n}{n} \| = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \tag{2.58}$$

Proof

First observe that (2.56) is satisfied at least if $d_i = b^2$, $i \geq 1$.

Now the proof of (2.57) is similar to that of (2.23), where here $c_n^2 = \sum_{i=1}^{n} d_i$, and (2.58) follows easily from (2.57) and the Borel-Cantelli Lemma.

■

We finally indicate without proof a pointwise result that is valid in a more general setting.

THEOREM 2.15 *Let $X = (X_n,\ n \in \mathbb{Z})$ be a strictly stationary B-valued process such that $EX_0 = 0$, $E \| X_0 \|^2 < \infty$ and*

$$E(< x^*, X_i >< x^*, X_j >) = 0, \quad x^* \in B^*, \quad i \neq j.$$

Then

$$\frac{S_n}{n} \to 0 \quad a.s.$$

Central limit theorems for dependent B-random variables

We first state a central limit theorem for martingale differences in a Hilbert space.

THEOREM 2.16 *Let $(X_i,\ i \geq 1)$ be an H-valued martingale difference and $(e_j,\ j \geq 1)$ be an orthonormal basis of H.*

Suppose that

$$n^{-1/2} E\left(\max_{1 \leq i \leq n} \| X_i \|\right) \to 0, \quad n \to \infty, \tag{2.59}$$

$$\frac{1}{n} \sum_{1 \leq i \leq n} < X_i, e_\ell >< X_i, e_k > \underset{a.s.}{\longrightarrow} \psi_{\ell,k},\ n \to \infty,\ \ell,\ k \geq 1, \tag{2.60}$$

where $(\psi_{\ell,k})$ is a family of real numbers, and

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(\sum_{i=1}^{n} r_N^2\left(n^{-1/2} X_i\right) > \varepsilon\right) = 0,\ \varepsilon > 0, \tag{2.61}$$

where $r_N^2(x) = \sum_{i=N}^{\infty} < x, e_i >^2, \quad x \in H.$

Then $n^{-1/2}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0,C)$, *where the covariance operator* C *satisfies*

$$< C(e_\ell),\ e_k >= \psi_{\ell,k}\ ,\ \ \ell,\ k \geq 1.$$

Concerning strong mixing we have the CLT under a sharp condition.

THEOREM 2.17 *Let* $(X_n,\ n \in \mathbb{Z})$ *be an* H*-valued strictly stationary zero-mean process with strong mixing coefficients* $(\alpha_k,\ k \geq 1)$. *Let* Q *be the quantile function of* $\| X_0 \|$ *(i.e., the generalized inverse function of* $x \mapsto P(\| X_0 \| > x)$*). Assume that*

$$\sum_{k=1}^{\infty} \int_0^{\alpha_k} Q^2(u)du < \infty. \tag{2.62}$$

Then for every $i, j \geq 1$,

$$\lim_{n\to\infty} \frac{1}{n} E\left(\sum_{k=1}^{n} < X_k,\ e_i > \sum_{k=1}^{n} < X_k,\ e_j >\right) = \sigma_{ij}$$

exists, and

$$n^{-1/2}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0,C),$$

where C *is such that*

$$< C(e_i),\ e_j >= \sigma_{ij}\ \ ;\ \ \ i, j \geq 1.$$

It can be proved that (2.62) is, in some sense, a minimal condition for CLT in the strongly mixing case. A sufficient condition for (2.62) is existence of a $\delta > 0$ such that $E \| X_0 \|^{2+\delta} < \infty$ and $\sum_{k=1}^{\infty} \alpha_k^{\delta/(2+\delta)} < \infty$.

In order to state a CLT in Banach spaces we need a definition.

DEFINITION 2.4 *Let* $X = (X_n,\ n \in \mathbb{Z})$ *be a* B*-valued random process.* X *is said to be* **(weakly) stationary** *if* $E \| X_n \|^2 < \infty$, $n \in \mathbb{Z}$, $EX_n = \mu$ *does not depend on* n *and the covariance functions of* X *satisfy*

$$E(x^*(X_{n+h} - \mu)y^*(X_{m+h} - \mu)) = E(x^*(X_n - \mu)y^*(X_m - \mu)); \tag{2.63}$$

$n, m, h \in \mathbb{Z}$; $x^*, y^* \in B^*$.

Now we have:

THEOREM 2.18 *Let* $X = (X_n,\ n \in \mathbb{Z})$ *be a weakly stationary strongly mixing* B*-valued stochastic process such that:*

$$EX_i = 0,\ i \in \mathbb{Z},\ \text{and} \sup_{i\in\mathbb{Z}} E \| X_i \|^{2+\gamma} < \infty \text{ for some } \gamma > 0, \tag{2.64}$$

$$\alpha_k = O\left(k^{-(1+\varepsilon)(1+2\gamma^{-1})}\right) \textit{ for some } \varepsilon > 0, \tag{2.65}$$

there exists a sequence (π_N) *of linear operators over* B *such that* $\dim \pi_N(B) = N$, $\| \pi_N \|_{\mathcal{L}} = O(N^r)$ *for some* $r > 0$, *and*

$$\sup_{a>0, n\geq 1} E \| n^{-1/2} \sum_{j=a+1}^{a+n} (X_j - \pi_N(X_j)) \|^2 = O(N^{-s}) \tag{2.66}$$

for some $s > 0$.

Then

$$T_n = n^{-1/2} \sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0, C),$$

where C *is associated with the covariance function*

$$\begin{aligned} c(x^*, y^*) &= E(x^*(X_1)y^*(X_1)) + \sum_{k\geq 2} E(x^*(X_1)y^*(X_k) \\ &+ x^*(X_k)y^*(X_1)), \ x^*, y^* \in B^*. \end{aligned}$$

Moreover,

$\delta(P_{T_N}, P_N) = O(n^{-\eta})$ *for some* $\eta > 0$, *where* δ *is Prokhorov's metric.*

2.5 * Derivation of exponential bounds

This section contains the proofs of Lemma 2.1 and Theorems 2.5 and 2.6.

Proof of Lemma 2.1

Consider the real martingale difference;

$$\xi_i = E^{\mathcal{A}_i} \| S_n \| - E^{\mathcal{A}_{i-1}} \| S_n \| \ ; \ \ i = 1, \ldots, n$$

where $\mathcal{A}_i = \sigma(X_1, \ldots, X_i)$, $i \geq 1$, and $\mathcal{A}_0 = \{\phi, \Omega\}$, and set

$$Y_i = X_i - x_i, \ \delta_i = \| S_n \| - \| S_n - Y_i \|, \ \eta_i = E^{\mathcal{A}_i}\delta_i, \ 1 \leq i \leq n.$$

Noting that $\delta_i \leq \| Y_i \|$, we get for each i

$$\eta_i \leq \| Y_i \| \qquad (a.s.). \tag{2.67}$$

On the other hand, independence of $S_n - Y_i$ and X_i entails

$$\xi_i = \eta_i - E^{\mathcal{A}_{i-1}}\eta_i. \tag{2.68}$$

From (2.67) and (2.68) it follows that

$$E^{\mathcal{A}_{i-1}}\left(\xi_i^2\right) \leq E \| Y_i \|^2 .$$

Thus

$$\sum_{i=1}^{n} E^{\mathcal{A}_{i-1}}\left(\xi_i^2\right) \leq \sum_{i=1}^{n} E \parallel X_i - x_i \parallel^2,$$

which implies (2.15). ■

Now, using (2.67), (2.68), and the elementary inequality

$$e^u - u \leq e^v - v \ , \quad |u| \leq v,$$

we successively obtain

$$E^{\mathcal{A}_{i-1}} e^{\xi_i} = e^{-E^{\mathcal{A}_{i-1}}\eta_i}\left[E^{\mathcal{A}_{i-1}}\left(e^{\eta_i} - \eta_i\right) + E^{\mathcal{A}_{i-1}}(\eta_i)\right]$$

and

$$E^{\mathcal{A}_{i-1}} e^{\xi_i} \leq e^{-b_i} \ (A_i + b_i), \tag{2.69}$$

where $b_i = E^{\mathcal{A}_{i-1}}\eta_i$ and $A_i = E\left(e^{\|Y_i\|} - \parallel Y_i \parallel\right)$.

Therefore, since $e^{-b_i}(A_i + b_i) \leq e^{A_i - 1}$, we get

$$E^{\mathcal{A}_{i-1}} e^{\xi_i} \leq e^{A_i - 1}. \tag{2.70}$$

Now we have

$$\sum_{i=1}^{n} \xi_i = \parallel S_n \parallel - E \parallel S_n \parallel;$$

thus

$$E e^{\|S_n\|} = e^{E\|S_n\|} E\left(e^{\sum_{i=1}^{n}\xi_i}\right),$$

and, using (2.70) repeatedly, we obtain

$$E e^{\|S_n\|} \leq \exp\left(E \parallel S_n \parallel + \sum_{i=1}^{n}(A_i - 1)\right).$$

Substituting λX_i for X_i, we get (2.16). ■

Let us now prove (2.17). First we consider the σ-algebras $\mathcal{T}_i = \sigma(X_k, \ k \in \{1, \ldots, n\} - \{i\})$. Then, by convexity of the norm and Jensen inequality,

$$E^{\mathcal{A}_{i-1}} \parallel S_n \parallel = E^{\mathcal{A}_{i-1}} E^{\mathcal{T}_i} \parallel S_n \parallel$$

$$\geq E^{\mathcal{A}_{i-1}} \parallel E^{\mathcal{T}_i} S_n \parallel = E^{\mathcal{A}_{i-1}} \parallel S_n - Y_i' \parallel,$$

with $Y_i' = X_i - EX_i$.

Hence

$$b_i = E^{\mathcal{A}_{i-1}}\eta_i = E^{\mathcal{A}_{i-1}} \parallel S_n \parallel - E^{\mathcal{A}_{i-1}} \parallel S_n - Y_i' \parallel \geq 0.$$

Using this last relation, together with (2.69) and the elementary inequality

$$e^{-b}(A+b) \leq A \quad , \quad b \geq 0,$$

we finally obtain

$$E^{\mathcal{A}_{i-1}} e^{\xi_i} \leq A_i,$$

and in the same way as above we conclude that

$$Ee^{\|S_n\|} \leq e^{E\|S_n\|} \prod_{i=1}^{n} A_i,$$

which is (2.17) with X_i instead of λX_i. ■

It remains to show (2.18). For this purpose we write

$$\begin{aligned} E[\mathrm{ch} \parallel S_n \parallel] &= \sum_{k=0}^{\infty} \frac{E \parallel S_n \parallel^{2k}}{(2k)!} \\ &= \sum_{k=0}^{\infty} E\left(\sum_{1\leq i,j\leq n} < X_i, X_j > \right)^k \frac{1}{(2k)!} \\ &= \sum_{k=0}^{\infty} E\,(\Sigma' < X_{i_1}, X_{j_1} > \ldots < X_{i_k}, X_{j_k} >) \frac{1}{(2k)!} \\ &\leq \sum_{k=0}^{\infty} E\,(\Sigma' \parallel X_{i_1} \parallel \parallel X_{j_1} \parallel \ldots \parallel X_{i_k} \parallel \parallel X_{j_k} \parallel) \frac{1}{(2k)!}, \end{aligned}$$

where Σ' denotes summation over all $2k$-tuples $(i_1, \ldots, i_k, j_1, \ldots, j_k)$ with elements in $\{1, \ldots, n\}$ satisfying $h_i \neq 1$ for any $i \in \{1, \ldots, n\}$, where h_i is the number of terms of the $2k$-tuple equal to i.

Therefore

$$E[\mathrm{ch} \parallel S_n \parallel] \leq \sum_{k=0}^{\infty} E\Sigma'' C_{2k}^{h_1,\ldots,h_n} \parallel X_1 \parallel^{h_1} \ldots \parallel X_n \parallel^{h_n} \frac{1}{(2k)!}, \tag{2.71}$$

where Σ'' denotes summation over all n-uples $(h_1, \ldots, h_n)$ of nonnegative integers such that $h_1 + \ldots + h_n = 2k$; $h_i \neq 1$; $i = 1, \ldots, n$; and

$$C_{2k}^{h_1,\ldots,h_n} = \frac{(2k)!}{h_1! \ldots h_n!}.$$

Since the right-hand side of (2.71) does not exceed $\prod_{i=1}^{n} E\left(e^{\|X_i\|} - \parallel X_i \parallel\right)$, we obtain (2.18) and the proof of Lemma 2.1 is now complete. ■

Proof of Theorem 2.5

Using (2.19) it easy to show that, if $\lambda < b^{-1}$,

$$\sum_{i=1}^{n} E\left(e^{\lambda \| X_i - x_i \|} - 1 - \lambda \parallel X_i - x_i \parallel\right) \leq \frac{\lambda^2 \ell^2}{2(1-\lambda b)}. \tag{2.72}$$

On the other hand we have the classical Bernstein bound

$$P(\parallel S_n \parallel \geq a) \leq e^{-\lambda a} E\left(e^{\lambda \| S_n \|}\right), \ a > 0. \tag{2.73}$$

Taking $a = t + E \parallel S_n \parallel$, $\lambda = \dfrac{t}{\ell^2 + bt}$ and using (2.16) we obtain

$$P(\parallel S_n \parallel - E \parallel S_n \parallel \geq t) \leq \exp\left(-\frac{t^2}{2\ell^2 + 2bt}\right).$$

Applying this result to $-X_i$, $1 \leq i \leq n$, we get the same bound for $P(E \parallel S_n \parallel - \parallel S_n \parallel \geq t)$; hence (2.20) holds.

Concerning (2.21), it suffices to use (2.18) and

$$P(\parallel S_n \parallel \geq t) \leq 2e^{-\lambda t} E[\mathrm{ch}\lambda \parallel S_n \parallel]. \tag{2.74}$$

Thus the proof of Theorem 2.5 is complete. ∎

Proof of Theorem 2.6

Using the growth of the function $\dfrac{e^u - 1 - u}{u^2}$ and $\parallel X_i - x_i \parallel \leq b$, we obtain the bound

$$\sum_{i=1}^{n} E\left(e^{\lambda \| X_i - x_i \|} - 1 - \lambda \parallel X_i - x_i \parallel\right) \leq \frac{e^{\lambda b} - 1 - \lambda b}{b^2} c_n^2.$$

Using (2.72) and (2.73) again, together with (2.15) and (2.16) and choosing $\lambda = \dfrac{1}{b} \mathrm{Log}\left(1 + \dfrac{tb}{c_n^2}\right)$, we get (2.22) and (2.23). ∎

NOTES

2.1 Representation (2.1) is used in Bosq (1990, 1991-a).

2.2 The classical Portmanteau Theorem is established in Billingsley (1968). Skorokhod (1956) proved Theorem 2.2. Prokhorov's metric and results are in Prokhorov (1956). Properties (2.9), (2.10), (2.11), and (2.12) come from Billingsley (1968).

2.3 Theorem 2.4 is in an early paper by Mourier (1953). The large deviation inequalities (Lemma 2.1, Theorems 2.5 and 2.6) are taken from Pinelis and Sakhanenko (1985). Similar inequalities appear in Yurinskii (1976). Type and cotype of Banach spaces were introduced by Maurey (1973) and independently by Hoffmann-Jorgensen (1974). For a general discussion about type and cotype we refer to Ledoux and Talagrand (LT) (1991).

Varadhan (1962) has obtained the CLT in Hilbert spaces. The counter-example 2.2 is given by Paulauskas and Rackauskas (1989). Theorem 2.8 appears in LT and Yurinskii (1982) has proved Theorem 2.9. Concerning optimality of the approximation rate $n^{-1/6}$ we refer again to Paulauskas and Rackauskas (1989). Theorem 2.10 is established in LT.

2.4 Theorem 2.11 and Corollary 2.3 seem to be new; they are extensions of a strong law of large numbers for dependent real random variables, which appears in Doob (1953).

Strong mixing was introduced by Rosenblatt (1956) and has been used by many authors since then. A general study has been performed by Doukhan (1994).

Lemma 2.2 is established in Bosq (1998). For other Bernstein type inequalities using mixing coefficients see, for example, Doukhan (1994). Theorem 2.14 is essentially due to Pinelis and Sakhanenko (1985). Theorem 2.15 has been established by Beck and Warren (1974).

The central limit theorem 2.16 comes from Jakubowski (1988). Theorem 2.17 is obtained by Merlevède and Peligrad (1997). Optimality of (2.62) has been noticed by Bradley (1997). Finally, Theorem 2.18 appears in Dehling (1983).

3

Autoregressive Hilbertian Processes of Order 1

In this chapter we particularize the representation of a continuous-time process as a sequence of B-random variables (recall Figure 1). Actually, we consider the case where B is a Hilbert space and the induced discrete time process is a linear Markov sequence. This leads to define the autoregressive Hilbertian process of order 1, denoted ARH(1), a flexible model that is used in practice to model and predict continuous-time random experiments (see Chapter 9).

Some more general linear models are considered in the remaining chapters. However, we shall see that they can be reduced to ARH(1) by using a suitable Markov representation. So, due to its generality and genuineness, the ARH(1) model plays a central role in the theory of functional linear processes.

Section 1 gives some ideas about stationarity in Hilbert spaces. We discuss existence and uniqueness of ARH(1) in Section 2 and establish some basic properties of this process in Sections 3 and 4. Limit theorems appear in Section 5.

3.1 Stationarity and innovation in Hilbert spaces

Throughout this chapter, H denotes a separable Hilbert space equipped with scalar product $< \cdot, \cdot >$, norm $\| \cdot \|$ and Borel σ-algebra $\mathcal{B}_H$. The H-random variables considered below are defined over the same probability space $(\Omega, \mathcal{A}, P)$ supposed to be rich enough and complete.

We first define an $\boldsymbol{H}$**-white noise**.

DEFINITION 3.1 *A sequence $\varepsilon = (\varepsilon_n, \ , n \in \mathbb{Z})$ of H-random variables is said to be an H-***white noise** *(WN) if*

1) $0 < E \parallel \varepsilon_n \parallel^2 = \sigma^2 < \infty$, $E\varepsilon_n = 0$, $C_\varepsilon := C_{\varepsilon_n}$ *do not depend on n, and*

2) ε_n *is orthogonal to* ε_m*;* $n, m \in \mathbb{Z}$*;* $n \neq m$*; i.e.,*

$$E(< \varepsilon_n, x >< \varepsilon_m, y >) = 0, \quad x, y \in H. \tag{3.1}$$

ε is said to be an ***H*** **strong white noise** *(SWN) if it satisfies 1) and*

2') (ε_n) *is a sequence of i.i.d. H-random variables.*

An SWN is a WN and the converse fails. It holds if ε is Gaussian.
Let us now give examples of Hilbertian white noises.

Example 3.1
Set $H = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$, where λ denotes Lebesgue measure, consider a measurable bilateral Wiener process W (cf. Example 1.9), and put

$$\varepsilon_n(t) = W_{n+t} - W_n, \quad 0 \leq t \leq 1, \quad n \in \mathbb{Z}. \tag{3.2}$$

Then Theorem 1.5 and (1.28) show that (ε_n) defines a sequence of H-random variables. Since increments of W are independent stationary, ε is a strong white noise.

More generally we have the following

Example 3.2
Take H and W as above and set

$$\varepsilon_n^{(\varphi)}(t) = \int_n^{n+t} \varphi(n+t-s)dW(s), \quad 0 \leq t \leq 1, \quad n \in \mathbb{Z}, \tag{3.3}$$

where $\varphi \in H$ and $\int_0^1 \varphi^2(u)du > 0$. Then $\left(\varepsilon_n^{(\varphi)}\right)$ is a strong white noise.

Note that in the previous examples W may be replaced by other regular enough processes with orthogonal stationary increments, for instance a compensated Poisson process (Example 1.8).

Innovation process

We now consider white noises that are naturally associated with stationary processes. Let $X = (X_n, \ n \in \mathbb{Z})$ be a zero-mean stationary H-valued process (cf. Definition 2.4). For each n, define $\mathcal{M}_n$ as the hermetically closed subspace generated by $(X_m, \ m \leq n)$ (cf. Section 1.6).

Then one may set

$$\varepsilon_n = X_n - \Pi^{\mathcal{M}_{n-1}}(X_n), \ n \in \mathbb{Z}, \tag{3.4}$$

where $\Pi^{\mathcal{M}_{n-1}}$ is the orthogonal projector of $\mathcal{M}_{n-1}$.

Now X is said to be **regular** if $E \parallel \varepsilon_n \parallel^2 = \sigma^2 > 0$, where σ^2 does not depend on n since X is stationary. If X is regular (ε_n) becomes, a white noise which is called the **innovation process** (or simply **innovation**) of X.

Rewriting (3.4) as $X_n = \Pi^{\mathcal{M}_{n-1}}(X_n) + \varepsilon_n$, one may interpret ε_n as a shock at time n, when $\Pi^{\mathcal{M}_{n-1}}(X_n)$ is the (linear) information on X_n provided by the past of X until time $n-1$.

In this chapter we are interested in the special case where there exists $\rho \in \mathcal{L}$ (the set of bounded linear operators from H to H) such that

$$\Pi^{\mathcal{M}_{n-1}}(X_n) = \rho(X_{n-1}), \quad n \in \mathbb{Z}. \tag{3.5}$$

In this context let us recall what is an H-valued Markov process.

DEFINITION 3.2

- *An H-valued second order process $X = (X_n,\ n \in \mathbb{Z})$ is a* **Markov process in the wide sense** *if*

$$\Pi^{\mathcal{M}^k_{n-1}}(X_n) = \Pi^{\mathcal{M}^1_{n-1}}(X_n), \quad n \in \mathbb{Z}, \quad k \geq 2, \tag{3.6}$$

  where $\Pi^{\mathcal{M}^k_{n-1}} (k \geq 1)$ is the orthogonal projector over the hermetically closed subspace of $L^2_H(\Omega, \mathcal{A}, P)$ generated by $X_{n-1}, \ldots, X_{n-k}$.

- *An H-valued process $X = (X_n,\ n \in \mathbb{Z})$ is a* **Markov process in the strict sense** *if*

$$P^{\mathcal{A}^k_{n-1}}(X_n \in A) = P^{\mathcal{A}^1_{n-1}}(X_n \in A), \ A \in \mathcal{B}_H\ ,\ n \in \mathbb{Z},\ k \geq 2, \tag{3.7}$$

  where $P^{\mathcal{A}^k_{n-1}}$ $(k \geq 1)$ denotes conditional probability with respect to the σ-algebra $\mathcal{A}^k_{n-1} := \sigma(X_{n-1}, \ldots, X_{n-k})$.

If X satisfies relation (3.5), it is a Markov process in the wide sense.

3.2 The ARH(1) model

DEFINITION 3.3 *A sequence $X = (X_n,\ n \in \mathbb{Z})$ of H-random variables is called an* **autoregressive Hilbertian process of order 1** *(ARH(1)) associated with (μ, ε, ρ) if it is* **stationary** *and such that*

$$X_n - \mu = \rho(X_{n-1} - \mu) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{3.8}$$

where $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ is an H-white noise, $\mu \in H$, and $\rho \in \mathcal{L}$.

Clearly definition 3.3 provides a generalization of the real AR(1) model and of the multidimensional $AR(1)$ model (cf. Section S.2). Note that we do not eliminate the special case $\rho = 0$ for which $(X_n - \mu)$ is nothing else but a white noise.

In order to study **existence** of X we introduce the following conditions:

(c_0) There exists an integer $j_0 \geq 1$ such that $\| \rho^{j_0} \|_{\mathcal{L}} < 1$

and

(c_1) There exist $a > 0$ and $0 < b < 1$ such that $\| \rho^j \|_{\mathcal{L}} \leq ab^j$, $j \geq 0$.

The next lemma is simple but somewhat surprising.

LEMMA 3.1 *(c_0) and (c_1) are equivalent.*

Proof

It is obvious that (c_1) yields (c_0). Let us show that (c_0) implies (c_1).

Clearly it suffices to prove (c_1) for $j > j_0$ and $0 < \| \rho^{j_0} \|_{\mathcal{L}} < 1$. For such a j we may write the result of its euclidian division by j_0 under the form

$$j = j_0 q + r, \tag{3.9}$$

where $q \geq 1$ and $0 \leq r < j_0$.

Now properties of $\| \cdot \|_{\mathcal{L}}$ entail

$$\| \rho^j \|_{\mathcal{L}} \leq \| \rho^{j_0} \|_{\mathcal{L}}^q \| \rho^r \|_{\mathcal{L}}$$

and since $q > \frac{j}{j_0} - 1$ and $0 < \| \rho^{j_0} \|_{\mathcal{L}} < 1$ it follows that

$$\| \rho^j \|_{\mathcal{L}} \leq ab^j, \quad j > j_0,$$

where $a = \| \rho^{j_0} \|_{\mathcal{L}}^{-1} \max\limits_{0 \leq r < j_0} \| \rho^r \|_{\mathcal{L}}$ and $b = \| \rho^{j_0} \|_{\mathcal{L}}^{1/j_0} < 1$. ■

This elementary lemma shows that the "natural" condition $\sum_{j=0}^{\infty} \|\rho^j\|_{\mathcal{L}}^2 < \infty$ for obtaining X_n under the form of a series that converges in $L_H^2(P)$ (see Theorem 3.1 below) is satisfied as soon as $\| \rho^{j_0} \|_{\mathcal{L}} < 1$ for some $j_0 \geq 1$.

Finally, observe that (c_0) (or (c_1)) does not imply $\| \rho \|_{\mathcal{L}} < 1$, contrary to the one-dimensional case (cf. Example 3.4 below).

We now may give a statement concerning existence and uniqueness of X.

THEOREM 3.1 *If (c_0) holds, then (3.8) has a unique stationary solution given by*

$$X_n = \mu + \sum_{j=0}^{\infty} \rho^j (\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{3.10}$$

where the series converges in $L^2_H(\Omega, \mathcal{A}, P)$ and almost surely. Moreover, ε is the innovation process of $(X_n - \mu)$.

Proof

We may and do assume that $\mu = 0$. Now orthogonality of the ε_n's entails

$$\Delta_m^{m'} := \left\| \sum_{j=m}^{m'} \rho^j(\varepsilon_{n-j}) \right\|^2_{L^2_H(P)} = \sum_{j=m}^{m'} \| \rho^j(\varepsilon_{n-j}) \|^2_{L^2_H(P)},$$

$1 \leq m < m'$. On the other hand,

$$\begin{aligned} \| \rho^j(\varepsilon_{n-j}) \|^2_{L^2_H(P)} &= E < \rho^j(\varepsilon_{n-j}),\ \rho^j(\varepsilon_{n-j}) > \\ &\leq \sigma^2 \| \rho^j \|^2_{\mathcal{L}}; \end{aligned}$$

hence lemma 3.1 yields

$$\Delta_m^{m'} \leq \sigma^2 \sum_{j=m}^{m'} \| \rho^j \|^2_{\mathcal{L}} \longrightarrow 0 \text{ as } m \text{ and } m' \to \infty.$$

Thus from the Cauchy criterion it follows that the series in (3.10) converges in $L^2_H(P)$. In fact, since $E\left(\sum_{j=0}^{\infty} \| \rho^j \| \, \| \varepsilon_{n-j} \|\right)^2 < \infty$, it follows that

$\sum_{j=0}^{\infty} \| \rho^j \| \, \| \varepsilon_{n-j} \| < \infty$ a.s. and the series also converges almost surely.

Let us now consider the stationary process

$$Y_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}.$$

By using boundedness of ρ we see that

$$\begin{aligned} Y_n - \rho(Y_{n-1}) &= \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}) - \sum_{j=0}^{\infty} \rho^{j+1}(\varepsilon_{n-1-j}), \\ &= \varepsilon_n, \quad n \in \mathbb{Z}, \end{aligned}$$

which means that (Y_n) is a solution of equation (3.8).

Conversely, let (X_n) be a stationary solution of (3.8). A straightforward induction gives

$$X_n = \sum_{j=0}^{k} \rho^j(\varepsilon_{n-j}) + \rho^{k+1}(X_{n-k-1}), \quad k \geq 1. \tag{3.11}$$

Therefore

$$E \parallel X_n - \sum_{j=0}^{k} \rho^j(\varepsilon_{n-j}) \parallel^2 \leq \parallel \rho^{k+1} \parallel_{\mathcal{L}}^2 E \parallel X_{n-k-1} \parallel^2 .$$

By stationarity, $E \parallel X_{n-k-1} \parallel^2$ remains constant and Lemma 3.1 yields $\parallel \rho^{k+1} \parallel_{\mathcal{L}}^2 \to 0$ as $k \to \infty$. Consequently

$$X_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}.$$

This proves uniqueness.

It remains to show that ε is the innovation of X: firstly, $\varepsilon_n = X_n - \rho(X_{n-1}) \in \mathcal{M}_n$ since $\mathcal{M}_n$ is an LCS. Secondly, $\varepsilon_n \perp \rho^j(\varepsilon_{m-j})$, $j \geq 0$, $m < n$, implies $\varepsilon_n \perp X_m = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{m-j})$, $m < n$, which in turn implies $\varepsilon_n \perp \mathcal{M}_m$, $m < n$, and the proof of Theorem 3.1 is now complete. ■

Let us give examples of such a model.

Example 3.3

Consider the Hilbert space $H = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$ and $\rho = \ell_K$, a kernel operator (cf. Example 1.18) associated with a kernel K such that

$$\int_{[0,1]^2} K^2(s,t)dsdt < 1.$$

Take as a white noise $\left(\varepsilon_n^{(\varphi)}\right)$ given by (3.3). Conditions in Theorem 3.1 are then satisfied and one obtains the ARH(1) process

$$X_n = \sum_{j=0}^{\infty} \ell_K^j \left(\varepsilon_{n-j}^{(\varphi)}\right), \quad n \in \mathbb{Z}.$$

Example 3.4

Let $\xi = (\xi_t,\ t \in \mathbb{R})$ be a measurable version of the Ornstein-Uhlenbeck process (1.15), with $\sigma^2 = 1$. Setting

$$X_n(t) = \xi_{n+t}, \quad 0 \leq t \leq 1, \quad n \in \mathbb{Z},$$

applying Theorem 1.5, and using the Karhunen-Loève expansion (1.29), one may claim that $X = (X_n,\ n \in \mathbb{Z})$ defines a sequence of random variables with values in $L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$.

For convenience we use a slightly different Hilbert space, namely $H = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda + \delta_{(1)})$, where as usual $\delta_{(1)}$ denotes Dirac measure at 1. Clearly X also defines an H-valued random sequence. Note that

$X_n(1)$ is fixed when $(X_n(t),\ 0 \leq t < 1)$ may be modified over a λ-null set.

Now let us define $\rho_\theta : H \mapsto H$ by putting

$$\rho_\theta(x)(t) = e^{-\theta t}x(1), \quad t \in [0,1], \quad x \in H, \quad \theta > 0,$$

(or equivalently $\rho_\theta(x)(t) = e^{-\theta t} < x,\ \mathbf{I}_{\{1\}} >$).

Concerning ε we set

$$\varepsilon_n(t) = \int_n^{n+t} e^{-\theta(n+t-s)} dW(s), \quad t \in [0,1[, \quad n \in \mathbb{Z},$$

and

$$\varepsilon_n(1) = X_n(1) - e^{-\theta} X_{n-1}(1), \ n \in \mathbb{Z}.$$

Then we have

$$\begin{aligned} X_n(t) &= \int_{-\infty}^{n+t} e^{-\theta(n+t-s)} dW(s) \\ &= e^{-\theta} \int_{-\infty}^{n} e^{-\theta(n-1+t-s)} dW(s) + \varepsilon_n(t), \end{aligned}$$

that is,

$$X_n = \rho_\theta(X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z}.$$

Now

$$\begin{aligned} \| \rho_\theta \|_{\mathcal{L}}^2 &= \int_0^1 e^{-2\theta t} d(\lambda + \delta_{(1)})(t) \\ &= \frac{1 - e^{-2\theta}}{2\theta} + e^{-2\theta} =: \alpha(\theta), \end{aligned}$$

and more generally

$$\| \rho_\theta^j \|_{\mathcal{L}}^2 = e^{-2\theta(j-1)} \alpha(\theta), \quad j \geq 1.$$

Thus condition (c_0) holds and (X_n) is an ARH(1) with innovation (ε_n). Note that if $0 < \theta \leq \frac{1}{2}$ we have $\| \rho_\theta \|_{\mathcal{L}} \geq 1$, hence $j_0 > 1$.

Also note that this autoregressive process is somewhat degenerated, since here the range of ρ_θ is one dimensional. This suggests that the ARH(1) model has a good amount of generality.

We will see later that ξ can also be associated with autoregressive processes in some suitable Banach spaces.

In order to state a corollary concerning uniqueness of (μ, ε, ρ), let us recall that the **support** S_Z of distribution of a random variable Z is defined by

$$S_Z = \{x : x \in H, \ \ P(\| Z - x \| < \alpha) > 0 \text{ for all } \alpha > 0\}.$$

COROLLARY 3.1 *If X is an ARH(1) associated with (μ, ε, ρ) and (c_0) holds, then (μ, ε) is unique, and ρ is unique over*

$$S = \overline{sp} \bigcup_{n \in \mathbb{Z}} \left(S_{X_n - \mu} \cup S_{\varepsilon_n} \right).$$

Proof

Uniqueness of (μ, ε) is obvious since $EX_n = \mu$ and ε is the innovation of $(X_n - \mu)$.

Now if $\rho_1 \in \mathcal{L}$ satisfies (c_0) and

$$X_n = \mu + \sum_{j=0}^{\infty} \rho_1^j (\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{3.12}$$

then (3.8) implies

$$\rho(X_{n-1} - \mu) = \rho_1(X_{n-1} - \mu), \ (a.s.), \quad n \in \mathbb{Z},$$

which in turn implies $\rho_1 = \rho$ over $S_{X_{n-1}-\mu}$ for all n.

On the other hand, (3.10) and (3.12) entail

$$(\rho - \rho_1)(\varepsilon_{n-1}) = \sum_{j=2}^{\infty} \left(\rho_1^j - \rho^j \right) (\varepsilon_{n-j}).$$

Then, from

$$(\rho - \rho_1)(\varepsilon_{n-1}) \perp \sum_{j=2}^{\infty} \left(\rho_1^j - \rho^j \right) (\varepsilon_{n-j}),$$

it follows that

$$(\rho - \rho_1)(\varepsilon_{n-1}) = 0, \quad n \in \mathbb{Z}.$$

This implies equality of ρ and ρ_1 over $S_{\varepsilon_{n-1}}$ for all n.

Finally, by linearity and continuity of ρ and ρ_1, one obtains uniqueness of ρ over S. ■

In some special cases, (c_0) is not necessary for obtaining (3.10). Let us give an example.

Example 3.5
Let ρ be defined as

$$\rho(x) = \alpha < x, e_1 > e_1 + \beta < x, e_3 > e_2, \quad x \in H,$$

where $\{e_1, e_2, e_3\}$ is an orthonormal system in H and $0 < \beta < 1 \leq \alpha$.

Here we have

$$\sum_{j=0}^{\infty} \| \rho^j \|_{\mathcal{L}}^2 = \sum_{j=0}^{\infty} \alpha^{2j} = +\infty.$$

However, if ε is a strong white noise such that

$$P(<\varepsilon_n,\ e_1>=0)=1,\ \ n\in\mathbb{Z},$$

we obtain

$$\rho(\varepsilon_n)=\beta<\varepsilon_n,e_3>e_2 \qquad (a.s.).$$

Then

$$\rho^j(\varepsilon_n)=0, \qquad j\geq 2;$$

Thus (3.8) has a stationary solution given by

$$X_n=\varepsilon_n+\rho(\varepsilon_{n-1}),\ \ n\in\mathbb{Z},$$

and (ε_n) is the innovation process of (X_n).

Here (X_n) is at the same time an ARH(1) and a so called **Hilbertian moving average of order 1**! This kind of phenomenon cannot occur in an univariate context.

Clearly the above example can be extended by taking ρ **nilpotent** (i.e. there exists $j\geq 2$ such that $\rho^j(x)=0$ for each x in S_{ε_0}); details are omitted.

3.3 Basic properties of ARH(1) processes

In the following we shall say that an ARH(1) is **standard** if $\mu=0$ and (c_0) holds. We now present some properties of (standard) ARH(1), processes. Here C denotes the covariance operator of X_0.

THEOREM 3.2 *If X is a standard ARH(1) the following relations are satisfied.*

$$C=\rho\, C\rho^*+C_\varepsilon. \tag{3.13}$$

$$C=\sum_{j=0}^{\infty}\rho^j C_\varepsilon \rho^{*j}, \tag{3.14}$$

where the series converges in the $\|\cdot\|_{\mathcal{N}}$ sense.

$$C_{X_n,X_m}=C\rho^{*(n-m)},\ \ n>m. \tag{3.15}$$

$$C_{X_m,X_n}=\rho^{(n-m)}C,\ \ n>m. \tag{3.16}$$

Proof

For each x in H, (3.8) yields

$$<C(x),x>=E(<X_n,x>^2)=E(<\rho(X_{n-1}),\ x>^2)+E(<\varepsilon_n,x>^2)$$

$$=<(\rho C\rho^*)(x),x>+<C_\varepsilon(x),x>,$$

which implies (3.13) since C and $\rho C\rho^* + C_\varepsilon$ are symmetric.

Let us now show (3.14). From (3.11) we infer that

$$C = \sum_{j=0}^{k} \rho^j C_\varepsilon \rho^{*j} + \rho^{k+1} C \rho^{*(k+1)}.$$

We now that $\rho^{k+1} C \rho^{*(k+1)}$ is nuclear since it is the covariance operator of $\rho^{k+1}(X_0)$. Using properties (1) and (3) in Section 1.5 we obtain

$$\| \rho^{k+1} C \rho^{*(k+1)} \|_{\mathcal{N}} = \| \rho^{k+1} A \|_{\mathcal{S}}^2 \leq \| \rho^{k+1} \|_{\mathcal{L}}^2 \| A \|_{\mathcal{S}}^2,$$

where A is such that $A^2 = C$.

Therefore

$$\| C - \sum_{j=0}^{k} \rho^j C_\varepsilon \rho^{*j} \|_{\mathcal{N}} \leq \| \rho^{k+1} \|_{\mathcal{L}}^2 \| A \|_{\mathcal{S}}^2 \underset{k\to\infty}{\longrightarrow} 0 .$$

It remains to prove (3.15) and (3.16). Using (3.11) again we get, for $n > m$ and x, y in H,

$$\begin{aligned} < C_{X_n,X_m}(x), y > = & \sum_{j=0}^{n-m-1} E\left(< \rho^j(\varepsilon_{n-j}), x >< X_m, y >\right) \\ & + E(< \rho^{n-m}(X_m), x >< X_m, y >). \end{aligned}$$

Since ε is the innovation, we obtain

$< C_{X_n,X_m}(x), y > = < C\rho^{*(n-m)}(x), y >$; hence (3.15) holds.

Concerning (3.16) it suffices to write, for $n > m$,

$$C_{X_m,X_n} = C^*_{X_n,X_m} = [C\rho^{*(n-m)}]^* = \rho^{n-m} C.$$

■

Theorem 3.2 shows that ρ and ρ^* may be interpreted as **autocorrelation operators** for $(X_n,\ X_{n+1})$.

The next assertion recalls the Markovian character of X.

THEOREM 3.3 *A standard ARH(1) X is a Markov process in the wide sense. If ε is a strong white noise, then X is a Markov process in the strict sense.*

Proof: clear. ■

We now study the link between the autoregressive structure of X and the fact that some specific coordinates of X are real autoregressive processes.

THEOREM 3.4 *Let X be a zero-mean ARH(1) associated with (ρ, ε), where ε is the innovation.*

Let v be an element of H such that $E(<\varepsilon_0, v>^2) > 0$. Then $(<X_n, v>,\ n \in \mathbb{Z})$ is a real Markov process in the wide sense, with innovation $(<\varepsilon_n, v>,\ n \in \mathbb{Z})$, if and only if there exists a real number α such that

$$<X_n,\ \rho^*(v) - \alpha v> = 0 \quad (a.s.),\ \ n \in \mathbb{Z}. \tag{3.17}$$

In that case $|\alpha| < 1$ and $(<X_n, v>)$ has the autoregressive representation

$$<X_n, v> = \alpha <X_{n-1}, v> + <\varepsilon_n, v>,\ \ n \in \mathbb{Z}. \tag{3.18}$$

Proof

If (3.17) holds for some α, we have (3.18), since

$$\begin{aligned} <X_n, v> &= <\rho(X_{n-1}) + \varepsilon_n, v> = <X_{n-1}, \rho^*(v)> + <\varepsilon_n, v> \\ &= \alpha <X_{n-1}, v> + <\varepsilon_n, v>. \end{aligned}$$

Therefore

$$E(<X_n, v>^2) = \alpha^2 E(<X_{n-1}, v>^2) + E(<\varepsilon_n, v>^2),$$

and by stationarity

$$(1-\alpha^2)E(<X_0, v>^2) = E(<\varepsilon_0, v>^2) > 0,$$

which yields $|\alpha| < 1$.

Consequently, the real AR(1) process $(<X_n, v>)$ is a Markov process in the wide sense.

Conversely if $(<X_n, v>)$ is Markovian (regular stationary) with innovation $(<\varepsilon_n, v>)$, there exists α and a sequence of real random variables $(\eta_n(v))$ such that

$$<X_n, v> = \alpha <X_{n-1}, v> + \eta_n(v),\ \ n \in \mathbb{Z},$$

with

$$\eta_n(v) \perp <X_{n-1}, v>,\ \ n \in \mathbb{Z}.$$

But this is possible only if $|\alpha| < 1$ and $\eta_n(v) = <\varepsilon_n, v>$, $n \in \mathbb{Z}$; thus $(<X_n, v>)$ has representation (3.18).

Comparing (3.8) and (3.18), one obtains (3.17) and the proof is complete.

■

Condition (3.17) is obviously satisfied if v is an eigenvector of ρ^* associated with the eigenvalue α. The following example shows that (3.17) may hold even if $\rho^*(v) \neq \alpha v$.

Example 3.6
Let $\{e_1, e_2\}$ be an orthogonal system of H and consider the operator ρ defined by

$$\rho(x) = \alpha x + \beta < x, e_1 > e_1, \quad x \in H,$$

where $\alpha > 0$, $\beta > 0$, and $\alpha + \beta < 1$.

Now let ε be a white noise such that $\varepsilon_0 \perp e_1$ (a.s.) and $E(< \varepsilon_0, e_2 >^2) > 0$.

Noting that $\| \rho \|_{\mathcal{L}} = \alpha + \beta < 1$ and $\rho^j(\varepsilon_0) = \alpha^j \varepsilon_0 \perp e_1$, $j \geq 0$, we may set

$$X_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}.$$

Then, if $v = \frac{1}{2}(e_1 + e_2)$, (3.17) holds, while

$$\rho^*(v) - \alpha v = \frac{\beta}{2} e_1 \neq 0.$$

■

The next example shows that $(< X_n, v >)$ is not in general a Markov process.

Example 3.7
Set

$$\rho(x) = \alpha(< x, e_1 > + < x, e_2 >)e_1 + \alpha < x, e_1 > e_2, \; x \in H,$$

where $\{e_1, e_2\}$ is an orthonormal system in H and $0 < \alpha < 1$.

Let ε be a white noise such that $E(<\varepsilon_n, e_1>^2) > 0$ and $E(<\varepsilon_n, e_2>^2)=0$.

Then, if (X_n) is the ARH(1) associated with (ρ, ε), it is easy to see that

$$< X_n, e_1 >= \alpha < X_{n-1}, e_1 > + \alpha^2 < X_{n-2}, e_2 > + < \varepsilon_n, e_1 > .$$

Thus $(< X_n, e_1 >)$ is a real AR(2) while (X_n) is an ARH(1)!

3.4 ARH(1) processes with symmetric compact autocorrelation operator

We now consider the special case where ρ is symmetric compact. As we have seen in Chapter 1, ρ admits the spectral decomposition

$$\rho = \sum_{j=1}^{\infty} \alpha_j e_j \otimes e_j, \tag{3.19}$$

where (e_j) is an orthonormal basis of H such that

$$\rho(e_j) = \alpha_j e_j, \ \ j \geq 1,$$

and where $|\alpha_j|$ is decreasing and such that $\lim\limits_{j\to\infty} |\alpha_j| = 0$.

Let us introduce an operator that connects ε with ρ. If there exists j such that $E(<\varepsilon_0, e_j>^2) > 0$, we set

$$\rho_\varepsilon = \sum_{j=j_0}^{\infty} \alpha_j e_j \otimes e_j, \tag{3.20}$$

where j_0 is the smallest j satisfying $E(<\varepsilon_0, e_j>^2) > 0$. Otherwise we set $\rho_\varepsilon = 0$. This case must be considered as degenerate since we then have $\rho(\varepsilon_0) = 0$ and

$$X_n = \varepsilon_n, \ \ n \in \mathbb{Z}.$$

Now it should be noticed that ρ_ε is associated with some specific choice of the e_j's. This choice may vary if one modifies eigenvectors that belong to the same multidimensional eigensubspace of ρ. However, $\| \rho_\varepsilon \|_{\mathcal{L}}$ does not depend on these e_j's and $\rho_\varepsilon(\varepsilon_0) = \rho(\varepsilon_0)$ (a.s.).

We are now in a position to state a necessary and sufficient condition for existence of an ARH(1) associated with (ρ, ε).

THEOREM 3.5 *Let ρ be a symmetric compact operator over H and ε be a, H-white noise. Then the equation*

$$X_n = \rho(X_{n-1}) + \varepsilon_n, \ \ n \in \mathbb{Z}, \tag{3.21}$$

has a stationary solution with innovation ε if and only if

$$\| \rho_\varepsilon \|_{\mathcal{L}} < 1. \tag{3.22}$$

Proof

First note that ε is a solution of (3.21) if and only if $\rho_\varepsilon = 0$.

Now if $\rho_\varepsilon \neq 0$ we have

$$|\alpha_{j_0}|^j = \| \rho_\varepsilon^j \|_{\mathcal{L}} = \| \rho_\varepsilon \|_{\mathcal{L}}^j, \ \ j \geq 1.$$

It follows that $\sum\limits_{j=0}^{\infty} \| \rho_\varepsilon^j \|_{\mathcal{L}}^2 < \infty$ is equivalent to $\| \rho_\varepsilon \|_{\mathcal{L}} < 1$.

Then, if condition $\| \rho_\varepsilon \|_{\mathcal{L}} < 1$ holds, one has

$$\| \sum_{j=m}^{m'} \rho^j(\varepsilon_{n-j}) \|_{L_H^2(P)}^2 \leq \sigma^2 \sum_{j=m}^{m'} \| \rho_\varepsilon \|_{\mathcal{L}}^{2j} \underset{\substack{m\to\infty \\ m'\to\infty.}}{\longrightarrow} 0$$

Then the Cauchy criterion entails that (3.21) has a stationary solution given by

$$X_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}) = \sum_{j=0}^{\infty} \rho_\varepsilon^j(\varepsilon_{n-j}) \ , \quad n \in \mathbb{Z},$$

and ε is clearly the innovation of (X_n).

Conversely let X be some stationary solution of (3.21) with innovation ε.

From (3.21) and symmetry of ρ it follows that

$$E\left(< X_n, e_{j_0} >^2\right) = \alpha_{j_0}^2 E\left(< X_{n-1}, e_{j_0} >^2\right) + E\left(< \varepsilon_n, e_{j_0} >^2\right);$$

that is to say,

$$\left(1 - \alpha_{j_0}^2\right) E\left(< X_0, e_{j_0} >^2\right) = E\left(< \varepsilon_0, e_{j_0} >^2\right) > 0,$$

which yields

$$\| \rho_\varepsilon \|_{\mathcal{L}} = |\alpha_{j_0}| < 1.$$

■

The next proposition shows that the autoregressive structure of coordinates implies the same structure for X.

THEOREM 3.6 *Let $\rho = \sum_{j=1}^{\infty} \alpha_j e_j \otimes e_j$ be a symmetric compact operator over H. Then (X_n) is a zero-mean ARH(1) associate with (ρ, ε) if and only if $(< X_n, e_j >)$ is an (eventually degenerate) AR(1) associated with $(\alpha_j, < \varepsilon_n, e_j >)$.*

Proof

If (X_n) is an ARH(1) associated with $(0, \rho, \varepsilon)$, then

$$< X_n, e_j > = < \rho(X_{n-1}), e_j > + < \varepsilon_n, e_j > .$$

Hence

$$< X_n, e_j > = \alpha_j < X_{n-1}, e_j > + < \varepsilon_n, e_j >, \quad j \geq 1. \tag{3.23}$$

If $\alpha_j \neq 0$ and $E(< \varepsilon_n, e_j >^2) > 0$, then $(< X_n, e_j >)$ is an AR(1). If $\alpha_j = 0$, $< X_n, e_j > = < \varepsilon_n, e_j >$ and $(< X_n, e_j >)$ is a degenerate AR(1).

Conversely, if (3.23) holds for all j, we may write

$$\begin{aligned} < X_n, x > &= \sum_{j=1}^{\infty} < X_n, e_j >< x, e_j > \\ &= \sum_{j=1}^{\infty} \alpha_j < X_{n-1}, e_j >< x, e_j > + \sum_{j=1}^{\infty} < \varepsilon_n, e_j >< x, e_j > \\ &= < \rho(X_{n-1}), x > + < \varepsilon_n, x > \ , \quad x \in H, \ n \in \mathbb{Z}. \end{aligned}$$

Thus

$$X_n = \rho(X_{n-1}) + \varepsilon_n \ , \quad n \in \mathbb{Z}.$$

Now, since $(< X_n, e_j >)$ is supposed to be AR(1) for each j, this implies $|\alpha_j| < 1$ and $\| \rho \|_{\mathcal{L}} = |\alpha_1| < 1$; thus

$$X_n = \sum_{j=0}^{\infty} \rho^j (\varepsilon_{n-j}), \quad n \in \mathbb{Z},$$

which proves that (X_n) is stationary with innovation (ε_n), and the proof is therefore complete. ■

Example 3.8
Let $(X_{n,j},\ n \in \mathbb{Z})$; $j = 1, 2, \ldots$; be a countable family of independent real AR(1) processes defined by

$$X_{n,j} = \alpha_j \ X_{n-1,j} + \varepsilon_{n,j}, \quad n \in \mathbb{Z}, \quad j \geq 1,$$

where $|\alpha_j| < 1$, $j \geq 1$, $|\alpha_j| \downarrow 0$ as $j \uparrow \infty$, and

$$E\left(\varepsilon_{n,j}^2\right) = \sigma_j^2 > 0 \ \text{ with } \ \sum_{j=1}^{\infty} \sigma_j^2 < \infty.$$

Now consider the Hilbert space ℓ^2 of square summable real sequences. Owing to the relations

$$\begin{aligned} E\left(\sum_{j=1}^{\infty} X_{n,j}^2\right) &= \sum_{j=1}^{\infty} E\left(\sum_k \alpha_j^k \varepsilon_{n-k,j}\right)^2 \\ &= \sum_{j=1}^{\infty} \frac{\sigma_j^2}{1 - \alpha_j^2} < \frac{1}{1 - |\alpha_1^2|} \sum_{j=1}^{\infty} \sigma_j^2 < \infty, \end{aligned}$$

we may claim that $X_n = (X_{n,j},\ j \geq 1)$, $n \in \mathbb{Z}$, defines a sequence of ℓ^2-valued random variables.

Furthermore, Theorem 3.6 implies that

$$X_n = \rho(X_{n-1}) + \varepsilon_n.$$

Here (ε_n) is the ℓ^2-valued white noise defined as $\varepsilon_n = (\varepsilon_{n,j},\ j \geq 1)$, $n \in \mathbb{Z}$, and ρ is the symmetric compact operator over ℓ^2 defined by

$$\rho = \sum_{j=1}^{\infty} \alpha_j e_j \otimes e_j,$$

where (e_j) is the standard orthonormal basis of ℓ^2:

$$e_1 = (1, 0, 0, \ldots) \ , \ e_2 = (0, 1, 0, \ldots), \ \ldots.$$

3.5 Limit theorems for ARH(1) processes

In this section we will study asymptotic behavior of $S_n = X_1 + \ldots + X_n$, $n \geq 1$, where (X_n) is ARH(1).

Let us begin with a simple but useful covariance type property.

LEMMA 3.2 *If X is a zero-mean ARH(1) with innovation ε then*

$$| E < X_0, X_h > | \leq \| \rho^h \|_{\mathcal{L}} \quad E \| X_0 \|^2 , \quad h > 0. \tag{3.24}$$

Proof

Relation (3.11) yields

$$< X_0, X_h >= \sum_{j=0}^{h-1} < X_0, \rho^j(\varepsilon_{h-j}) > + < X_0, \rho^h(X_0) > , \quad h > 0. \tag{3.25}$$

Then, since ε is the innovation, we have

$$E < X_0, X_h >= E < X_0, \rho^h(X_0) >,$$

which implies (3.24). ■

We may now state laws of large numbers.

THEOREM 3.7 *Let X be a standard ARH(1). Then, as $n \to \infty$,*

$$E \| \frac{S_n}{n} \|^2 = O\left(\frac{1}{n}\right) \tag{3.26}$$

and, for all $\beta > \frac{1}{2}$,

$$\frac{n^{1/4}}{(\text{Log } n)^\beta} \frac{S_n}{n} \to 0 \quad a.s. \tag{3.27}$$

Proof

We start from the identity

$$E \| X_n + \ldots + X_{n+p-1} \|^2 = \sum_{n \leq j, j' \leq n+p-1} E < X_j, X_{j'} > .$$

Using stationarity and Lemma 3.2 we obtain

$$E \| X_n + \ldots + X_{n+p-1} \|^2 \leq 2pE \| X_0 \|^2 \sum_{h=0}^{p-1} \| \rho^h \|_{\mathcal{L}};$$

thus Lemma 3.1 gives

$$E \| X_n + \ldots + X_{n+p-1} \|^2 = O(p). \tag{3.28}$$

Then (3.26) is obvious and (3.27) follows from Corollary 2.3. ■

The next statement gives the exact rate of convergence.

THEOREM 3.8 *If X is a standard ARH(1) process, then we have*

$$\| nC_{S_n/n} - \sum_{h=-\infty}^{+\infty} C_{X_0,X_h} \|_{\mathcal{N}} \underset{n\to\infty}{\longrightarrow} 0 , \tag{3.29}$$

and

$$nE \| \frac{S_n}{n} \|^2 \underset{n\to\infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < X_0, X_h > . \tag{3.30}$$

Proof

First note that (3.25) entails

$$C_{X_0,X_h} = C_{X_0,\rho^h(X_0)}. \tag{3.31}$$

Now from (1.62) and (3.31) it follows that

$$\| C_{X_0,X_h} \|_{\mathcal{N}} \leq E \left(\| X_0 \| \| \rho^h(X_0) \| \right).$$

Hence

$$\| C_{X_0,X_h} \|_{\mathcal{N}} \leq \| \rho^h \|_{\mathcal{L}} E \| X_0 \|^2, \tag{3.32}$$

and (3.32) yields convergence of $\sum_{h=-\infty}^{+\infty} C_{X_0,X_h}$ in the Banach space $(\mathcal{N}, \| \cdot \|_{\mathcal{N}})$ of nuclear operators over H.

On the other hand,

$$C_{S_n/n} = \frac{1}{n^2} \sum_{1\leq i,j\leq n} C_{X_i,X_j}$$

and by stationarity

$$n\, C_{S_n/n} = \sum_{|h|\leq n-1} \left(1 - \frac{|h|}{n}\right) C_{X_0,X_h}.$$

From (3.32) we deduce that

$$nC_{S_n/n} \underset{n\to\infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} C_{X_0,X_h}$$

in nuclear norm.

Finally, in order to establish (3.30), it suffices to remark that

$$n\, E \| \frac{S_n}{n} \|^2 = \sum_{|h|\leq n-1} \left(1 - \frac{|h|}{n}\right) E < X_0, X_h >$$

and then to use Lemma 3.2. ■

As a by-product of (3.29) and (3.30), and since $\| C_{S_n/n} \|_{\mathcal{N}} = E \| \frac{S_n}{n} \|^2$ (cf. (1.59)), we have

$$\| \sum_{h=-\infty}^{+\infty} C_{X_0,X_h} \|_{\mathcal{N}} = \sum_{h=-\infty}^{+\infty} E < X_0, X_h > . \tag{3.33}$$

Note also that (3.31) and some calculations lead to

$$\sum_{h=-\infty}^{+\infty} C_{X_0,X_h} = C_{X_0,(I-\rho)^{-1}X_0} + C_{(I-\rho)^{-1}X_0,X_0} - C. \tag{3.34}$$

We now turn to large deviations. For convenience we shall say that $X_0 \in \mathcal{E}$ if there exists some $\gamma > 0$ such that $E(\exp \gamma \| X_0 \|) < \infty$. The next lemma connects the conditions $X_0 \in \mathcal{E}$ and $\varepsilon_0 \in \mathcal{E}$.

LEMMA 3.3 *Let X be an ARH(1) associated with (μ, ε, ρ), where ε is a strong white noise and ρ satisfies (c_0). Then $X_0 \in \mathcal{E}$ if and only if $\varepsilon_0 \in \mathcal{E}$.*

Proof

We may suppose that $\rho \neq 0$ since the result is obvious if $\rho = 0$. We may also take $\mu = 0$, since $X_0 \in \mathcal{E}$ and $X_0 - \mu \in \mathcal{E}$ are equivalent.

Now (3.8) implies

$$\| \varepsilon_n \| \leq \| \rho \|_{\mathcal{L}} \| X_{n-1} \| + \| X_n \|,$$

so

$$E\left(e^{\alpha \| \varepsilon_n \|}\right) \leq \frac{1}{2}\left[e^{2\alpha \| \rho \|_{\mathcal{L}} \| X_{n-1} \|} + e^{2\alpha \| X_n \|}\right].$$

Then, if $X_0 \in \mathcal{E}$ and if $\alpha = \frac{\gamma}{2} \min\left(1, \| \rho \|_{\mathcal{L}}^{-1}\right)$, we have

$$E\left(e^{\alpha \| \varepsilon_n \|}\right) \leq E\left(e^{\gamma \| X_0 \|}\right) < \infty;$$

thus $\varepsilon_0 \in \mathcal{E}$.

Conversely, let us suppose that $E\left(e^{\gamma \| \varepsilon_0 \|}\right) < \infty$ for some $\gamma > 0$. Using (3.10) we get

$$\| X_n \| \leq \sum_{j=0}^{\infty} \| \rho^j \|_{\mathcal{L}} \| \varepsilon_{n-j} \| .$$

Now independence of $\| \varepsilon_n \|, \| \varepsilon_{n-1} \|, \ldots$ and monotone convergence of expectation entail, for each $\alpha > 0$,

$$E\left(e^{\alpha \| X_n \|}\right) \leq \prod_{j=0}^{\infty} E\left(e^{\alpha \| \rho^j \|_{\mathcal{L}} \| \varepsilon_{n-j} \|}\right),$$

and the choice $\alpha = \gamma \left(\max_{j \geq 0} \| \rho^j \|_{\mathcal{L}} \right)^{-1}$ ensures finiteness of each term in the infinite product.

Now, in order to verify convergence of this infinite product, it suffices to show that

$$S := \sum_{j \geq j_1} \mathrm{Log} E \left(e^{\alpha \| \rho^j \|_{\mathcal{L}} \| \varepsilon_0 \|} \right) < \infty,$$

where $(j_1 \geq j_0)$ is such that $\| \rho^j \|_{\mathcal{L}} \leq 1$ for each $j \geq j_1$.

Noting that

$$S \leq \sum_{j \geq j_1} \mathrm{Log} \left(1 + \sum_{k=1}^{\infty} \frac{\alpha^k \| \rho^j \|_{\mathcal{L}}^k}{k!} E \| \varepsilon_0 \|^k \right),$$

we obtain

$$\begin{aligned} S &\leq \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} E \| \varepsilon_0 \|^k \sum_{j \geq j_1} \| \rho^j \|_{\mathcal{L}}^k \\ &\leq E \left(e^{\gamma \| \varepsilon_0 \|} \right) \sum_{j \geq j_1} \| \rho^j \|_{\mathcal{L}} < \infty. \end{aligned}$$

Thus $E \left(e^{\alpha \| \varepsilon_0 \|} \right) < \infty$ and the proof is complete. ■

We then have the following inequality.

THEOREM 3.9 *Let X be an ARH(1) associated with (μ, ε, ρ), where ε is a strong white noise and ρ satisfies (c_0).*

Then, if $\varepsilon_0 \in \mathcal{E}$ (or, equivalently, $X_0 \in \mathcal{E}$), there exist $\alpha_0 > 0$ and $\beta_0 > 0$, which only depend on ρ and P_{ε_0}, such that

$$P \left(\| \frac{S_n}{n} - \mu \| \geq \eta \right) \leq 4 \exp \left(- \frac{n \eta^2}{\alpha_0 + \beta_0 \eta} \right), \quad \eta > 0. \tag{3.35}$$

Proof

We may suppose that $\mu = 0$. Now note that $(I - \rho)^{-1} = \sum_{j \geq 0} \rho^j$ does exist and is bounded, and consider the decomposition

$$\frac{S_n}{n} = (I - \rho)^{-1} \overline{\varepsilon_n} + \Delta_n, \quad n \geq 1, \tag{3.36}$$

where $\overline{\varepsilon_n} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i$ and

$$\Delta_n = (I - \rho)^{-1} \frac{\rho^n(\varepsilon_1) + \ldots + \rho(\varepsilon_n)}{n} + \frac{(\rho + \ldots + \rho^n)(X_0)}{n}. \tag{3.37}$$

For all $\eta > 0$ we have

$$P\left(\| \frac{S_n}{n} \| \geq \eta\right) \leq P\left(\| (I-\rho)^{-1} \|_{\mathcal{L}} \| \overline{\varepsilon_n} \| \geq \frac{\eta}{2}\right) + P\left(\| \Delta_n \| \geq \frac{\eta}{2}\right). \tag{3.38}$$

We are now in a position to apply inequality (2.21) in Theorem 2.5.

First note that $\varepsilon_0 \in \mathcal{E}$ implies (2.19) for the random variables $\varepsilon_1, \ldots, \varepsilon_n$. Actually, since

$$e^{\gamma \| \varepsilon_0 \|} \geq \frac{\gamma^k \| \varepsilon_0 \|^k}{k!}, \quad k \geq 2,$$

it follows that

$$E \| \varepsilon_0 \|^k \leq \frac{k!}{2} \ell_0^2 b^{k-2},$$

where $\ell_0^2 = 2\gamma^{-2} E\left(e^{\gamma \| \varepsilon_0 \|}\right)$ and $b = \gamma^{-1}$; hence

$$\sum_{i=1}^{n} E \| \varepsilon_i \|^k \leq \frac{k!}{2} (\ell_0 \sqrt{n})^2 b^{k-2} \quad , \quad k \geq 2.$$

Now putting, $t = n\frac{\eta}{2} \| (I-\rho)^{-1} \|_{\mathcal{L}}^{-1} := n\eta'$ and applying (2.21), we obtain

$$P\left(\| (I-\rho)^{-1} \|_{\mathcal{L}} \| \overline{\varepsilon_n} \| \geq \frac{\eta}{2}\right) \leq 2 \exp\left(-\frac{n\, \eta'^2}{2\ell_0^2 + 2b\eta'}\right).$$

Concerning $P\left(\| \Delta_n \| \geq \frac{\eta}{2}\right)$, we may again apply (2.21), since the summands are independent and belong to $\mathcal{E}$, with a γ that does not depend on n because $\sum_{j=0}^{\infty} \| \rho^j \|_{\mathcal{L}} < \infty$.

Finally (3.38) entails (3.35). ■

Note that α_0 and β_0 are somewhat complicated but explicit.

Inequality (3.35) leads to an optimal rate in the law of large numbers, namely

COROLLARY 3.2

$$\| \frac{S_n}{n} \| = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \tag{3.39}$$

Proof:

Apply (3.35) and the Borel-Cantelli Lemma with a suitable η or look at the proof of Corollary 4.1. ■

We now deal with the central limit theorem for ARH(1). We have the following.

THEOREM 3.10 *Let X be an ARH(1) associated with a strong white noise and an operator ρ satisfying (c_0). Then*

$$\sqrt{n}\left(\frac{S_n}{n}-\mu\right) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0,\Gamma), \tag{3.40}$$

where $\Gamma = (I-\rho)^{-1}C_\varepsilon(I-\rho^*)^{-1}$.

Proof

Of course we may suppose that $\mu = 0$. Now consider the decomposition

$$\sqrt{n}\frac{S_n}{n} = (I-\rho)^{-1}\sqrt{n}\overline{\varepsilon_n} + \sqrt{n}\Delta_n, \tag{3.41}$$

where Δ_n is defined by (3.37).

Applying the i.i.d. central limit theorem in Hilbert space (Theorem 2.7), we obtain

$$(I-\rho)^{-1}\sqrt{n}\overline{\varepsilon_n} \xrightarrow{\mathcal{L}} N \sim \mathcal{N}(0,\Gamma).$$

On the other hand,

$$\| \sqrt{n}\Delta_n \| \leq \| (I-\rho^{-1}) \|_{\mathcal{L}} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \| \rho^{n-j+1} \|_{\mathcal{L}} \| \varepsilon_j \| + \frac{\| X_0 \|}{\sqrt{n}} \sum_{j\geq 1} \| \rho^j \|_{\mathcal{L}} .$$

It follows that

$$\sqrt{n}\Delta_n \underset{p}{\longrightarrow} 0.$$

Then (2.9) entails the desired result. ■

We finally derive a **Berry-Esseen type bound.**

THEOREM 3.11 *If conditions in Theorem 3.10 hold and if in addition $E \| \varepsilon_0 \|^3 < \infty$ and $\Gamma \neq 0$, then there exists a positive constant c that depends only on P_X, such that*

$$\sup_{r\geq 0}\left| P\left(\| \sqrt{n}\left(\frac{S_n}{n}-\mu\right) \| < r\right) - P\left(\| N \| < r\right)\right| \leq \frac{c}{\sqrt{n}}, \quad n \geq 1. \tag{3.42}$$

Proof

First, since $(I-\rho)^{-1}\varepsilon_i$ is a sequence of i.i.d. H-valued nondegenerate zero-mean random variables satisfying $E \| (I-\rho)^{-1}\varepsilon_i \|^3 < \infty$, we may use the bound (2.31) in Theorem 2.9 to write

$$\sup_{r\geq 0}\left| P\left(\| (I-\rho)^{-1}\sqrt{n}\overline{\varepsilon_n} \| < r\right) - P(\| N \| < r)\right| \leq \frac{c_1}{\sqrt{n}}, \tag{3.43}$$

where $N \sim \mathcal{N}(0, \Gamma)$ and c_1 is a positive constant.

We now consider the expression

$$\Pi_n(r) := P(\| Y_n + \sqrt{n}\Delta_n \|< r) - P(\| Y_n \|< r),$$

where $Y_n = (I - \rho)^{-1}\sqrt{n}\overline{\varepsilon_n}, \ \ n \geq 1$.

Using elementary inequalities we get

$$\begin{aligned} |\Pi_n(r)| &\leq P(r- \| \sqrt{n}\Delta_n \|\leq Y_n < r+ \| \sqrt{n}\Delta_n \|) \\ &\leq \int_0^\infty P(r - \delta \leq\| Y_n \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta). \end{aligned}$$

Using (2.31) again we obtain

$$\begin{aligned} A_n : &= \int_0^r P(r - \delta \leq\| Y_n \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta) \\ &\leq \int_0^r P(r - \delta \leq\| N \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta) + \frac{c_1}{\sqrt{n}}. \end{aligned}$$

Now since $\Gamma \neq 0$ it can be shown that $\| N \|$ has a bounded density, say φ_N. This is a consequence of the representation $\| N \|^2 = \sum_{j=1}^{\infty} < N, g_j >^2$, where (g_j) is an orthonormal basis formed with eigenvectors of Γ. Then we have the bound

$$A_n \leq 2 \| \varphi_N \|_\infty E \| \sqrt{n}\Delta_n \| + \frac{c_1}{\sqrt{n}}.$$

But from (3.37) it follows that $E \| \sqrt{n}\Delta_n \| \leq \frac{c_2}{\sqrt{n}}$, where c_2 is a constant; thus

$$A_n \leq \frac{2 \| \varphi_N \|_\infty c_2 + c_1}{\sqrt{n}}. \tag{3.44}$$

On the other hand, consider

$$\begin{aligned} B_n : &= \int_r^\infty P(r - \delta \leq\| Y_n \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta) \\ &= \int_r^\infty P(\| Y_n \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta). \end{aligned}$$

We have

$$\begin{aligned} B_n &= \int_r^\infty P(\| Y_n \|\leq r) + P(r <\| Y_n \|< r + \delta) dP_{\sqrt{n}\Delta_n}(\delta) \\ &\leq \left[P(\| N \|\leq r) + \frac{c_1}{\sqrt{n}}\right] P(\| \sqrt{n}\Delta_n \|\geq r) \\ &+ \int_r^\infty \left[P(r \leq\| N \|< r + \delta) + \frac{2c_1}{\sqrt{n}}\right] dP_{\sqrt{n}\Delta_n}(\delta). \end{aligned}$$

Now note that

$$P(\| N \| \leq r) P(\| \sqrt{n}\Delta_n \| \geq r) \leq 2r \| \varphi_N \|_\infty \frac{c_2}{r\sqrt{n}}$$

and

$$\begin{aligned} \int_r^\infty \cdots &\leq \| \varphi_N \|_\infty \int_r^\infty \delta dP_{\sqrt{n}\Delta_n}(\delta) + \frac{2c_1}{\sqrt{n}} \\ &\leq \frac{\| \varphi_N \|_\infty c_2 + 2c_1}{\sqrt{n}}. \end{aligned}$$

Therefore

$$B_n \leq \frac{c_3}{\sqrt{n}}, \tag{3.45}$$

where c_3 is a positive constant.

Thus (3.44) and (3.45) imply

$$|\Pi_n(r)| \leq \frac{c_4}{\sqrt{n}}, \tag{3.46}$$

where c_4 is a positive constant.

Now, in order to get (3.42), it suffices to take into account the decomposition

$$\sqrt{n}\frac{S_n}{n} = (I - \rho)^{-1}\sqrt{n}\overline{\varepsilon_n} + \sqrt{n}\Delta_n$$

together with (3.43) and (3.46). ■

Finally, let us indicate that ARH(1) processes satisfy the **law of the iterated logarithm**. Details will be given in the context of Banach spaces (cf. Chapter 6).

NOTES

3.1 Definition of innovation process is taken from Fortet (1995).

3.2 The ARH model appears in Bosq (1990, 1991-a), Mourid (1993), Merlevède (1996-a), and Fortet (1995), among others. The definition given here is explicit and thus adapted to statistical applications. Some results and examples are taken from Bosq (1999-b). Lemma 3.1 allows us to improve conditions concerning ρ employed in previous works. A similar lemma appears for example in Kubrusly (1985).

3.3 Theorem 3.4 and examples are presented in Bosq (1999-b).

3.4 Theorems 3.5 and 3.6 are slight improvements of results in Bosq (1991-a).

3.5 Most of the results are improvements of Theorems that appear in the papers cited above. In particular, Theorems 3.8 and 3.11 seem to be new.

4

Estimation of Autocovariance Operators for ARH(1) Processes

Estimation of second-order parameters of an ARH(1) has a twofold interest. Firstly, it gives information about inner structure of the process. Secondly, it leads to construction of statistical linear predictors.

The main tool used in this chapter is autoregressive representation of observed covariance operators of an ARH(1). This representation allows us to mimic the methods of proof that appear in the previous chapter.

Section 1 is devoted to estimation of the covariance operator and Section 2 to estimation of its eigenelements. Section 3 deals with estimation of the cross-covariance operators. Limits in distribution appear in Section 4.

Notation is the same as in Chapter 3.

4.1 Estimation of the covariance operator

Let $X = (X_n,\ n \in \mathbb{Z})$ be a **standard ARH(1)** associated with (ρ, ε) where ε is a **strong white noise**, and such that

$$E \parallel X_0 \parallel^4 < \infty. \tag{4.1}$$

For convenience this family of conditions will be called **"Assumption** A_1**"** from now on.

This section deals with estimation of the covariance operator C defined by

$$C(x) = E(< X_0, x > X_0) \ , \quad x \in H. \tag{4.2}$$

If $X_1, \ldots, X_n$ are observed, a natural estimator of C is the so-called **empirical covariance operator**, defined as

$$C_n(x) = \frac{1}{n} \sum_{i=1}^{n} < X_i, x > X_i \ , \ \ x \in H, \tag{4.3}$$

or more compactly as

$$C_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i. \tag{4.4}$$

Since the range of C_n is finite-dimensional, it follows that C_n is nuclear and therefore Hilbert-Schmidt.

Now let us set

$$Z_i =< X_i, \cdot > X_i - C \ \ , \ \ \ i \in \mathbb{Z}.$$

Lemma 1.2 shows that $Z = (Z_i, \ i \in \mathbb{Z})$ is a sequence of $\mathcal{S}$-valued random variables. Moreover, (1.49) yields

$$\|< X_i, \cdot > X_i \|_{\mathcal{S}}^2 = \| X_i \|^4;$$

thus from (4.1) it follows that Z_i belongs to the Hilbert space $L^2_{\mathcal{S}}(\Omega, \mathcal{A}, P)$.

On the other hand it is easy to check that

$$EC_n = C, \tag{4.5}$$

where expectation is taken in $\mathcal{S}$: $\boldsymbol{C_n}$ **is an unbiased estimator of** $\boldsymbol{C}$.

In order to state a representation lemma for Z we need some notation. In the following we shall denote by $\| \cdot \|_{\mathcal{L}}^{(\mathcal{S})}$, $\| \cdot \|_{\mathcal{S}}^{(\mathcal{S})}$, and $\| \cdot \|_{\mathcal{N}}^{(\mathcal{S})}$ respectively the norms $\mathcal{L}$, $\mathcal{S}$, and $\mathcal{N}$ over the corresponding spaces of linear operators on $\mathcal{S}$.

We also set $F_i =< X_{i-1}, \cdot > \varepsilon_i$, $G_i =< \varepsilon_i, \cdot > \varepsilon_i$, and $\mathcal{B}_i = \sigma(\varepsilon_i, \varepsilon_{i-1}, \ldots)$ for all i in $\mathbb{Z}$.

Now we have the following crucial assertions

LEMMA 4.1 *If A_1 holds, the $\mathcal{S}$-valued process Z admits the* $\boldsymbol{ARS}(1)$ **representation**

$$Z_i = R(Z_{i-1}) + E_i \ , \ \ i \in \mathbb{Z}, \tag{4.6}$$

where R is the bounded linear operator over $\mathcal{S}$ defined as

$$R(s) = \rho s \rho^* \ , \ \ s \in \mathcal{S}, \tag{4.7}$$

and where

$$E_i = F_i \rho^* + \rho F_i^* + G_i - C_\varepsilon \ , \ \ i \in \mathbb{Z}. \tag{4.8}$$

Furthermore, R satisfies

$$\| R^h \|_{\mathcal{L}}^{(\mathcal{S})} \leq \| \rho^h \|_{\mathcal{L}}^2 \ , \ \ h \geq 1, \tag{4.9}$$

and E_i belongs to $L^2_{\mathcal{S}}(\Omega, \mathcal{B}_i, P)$ and satisfies

$$E^{\mathcal{B}_{i-1}}(E_i) = 0 \ , \ \ i \in \mathbb{Z}. \tag{4.10}$$

Hence (E_i) is a **martingale difference** *with respect to $(\mathcal{B}_i)$ and consequently a Hilbertian white noise. Moreover (E_i) is the innovation of (Z_i).*

Proof

Using the basic relation

$$X_i = \rho(X_{i-1}) + \varepsilon_i \ , \ \ i \in \mathbb{Z},$$

we obtain

$$\begin{aligned} Z_i(\cdot) \quad &= \quad < \rho(X_{i-1}) + \varepsilon_i, \cdot > [\rho(X_{i-1}) + \varepsilon_i] - C(\cdot) \\ &= \quad \rho\,[< X_{i-1}, \rho^*(\cdot) > X_{i-1} - C\rho^*] + E_i \end{aligned}$$

since, from (3.13), $C_\varepsilon = C - \rho C \rho^*$. Thus

$$Z_i = R(Z_{i-1}) + E_i \ , \ \ i \in \mathbb{Z}.$$

Now boundedness of R in $\mathcal{S}$ follows from

$$\| \, R(s) \, \|_{\mathcal{S}} = \| \, \rho s \rho^* \, \|_{\mathcal{S}} \leq \| \, \rho \, \|^2_{\mathcal{L}} \| \, s \, \|_{\mathcal{S}} \ .$$

More generally, an easy induction shows that

$$R^h = \rho^h s \rho^{*h} \ , \ \ h \geq 1, \tag{4.11}$$

and (4.9) follows.

On the other hand, E_i is clearly $\mathcal{B}_i$-measurable and since $E \, \| \, F_i \, \|^2_{\mathcal{S}}$ and $E \, \| \, G_i \, \|^2_{\mathcal{S}}$ are finite we thus have $E_i \in L^2_{\mathcal{S}}(\Omega, \mathcal{B}_i, P)$.

Now consider an orthonormal basis $(e_j, \ j \geq 1)$ of H. It generates an orthonormal basis $(s_{\ell k}, \ \ell \geq 1, \, k \geq 1)$ of $\mathcal{S}$ via the relations

$$s_{\ell k} =< e_\ell, \cdot > e_k, \ \ \ell \geq 1, \ k \geq 1. \tag{4.12}$$

We use $(s_{\ell k})$ for proving (4.10). First we have

$$\begin{aligned} &< E^{\mathcal{B}_{i-1}}(F_i), \ s_{\ell k} >_{\mathcal{S}} = E^{\mathcal{B}_{i-1}}(< X_{i-1}, e_\ell >< \varepsilon_i, e_k >) \\ &=< X_{i-1}, e_\ell > E^{\mathcal{B}_{i-1}}(< \varepsilon_i, e_k >) = 0 \ \ ; \quad \ell \geq 1, \ k \geq 1. \end{aligned}$$

Hence

$$E^{\mathcal{B}_{i-1}}(F_i) = 0. \tag{4.13}$$

Similarly we have

$$E^{\mathcal{B}_{i-1}}\,(F_i^*) = 0. \tag{4.14}$$

On the other hand,

$$< E^{\mathcal{B}_{i-1}}(G_i),\ s_{\ell k} >_{\mathcal{S}} = E^{\mathcal{B}_{i-1}}(< \varepsilon_i, e_\ell >< \varepsilon_i, e_k >)$$

$$= E(< \varepsilon_i, e_\ell >< \varepsilon_i, e_k >) =< C_\varepsilon(e_\ell), e_k > ;\ \ell \geq 1,\ k \geq 1.$$

So

$$E^{\mathcal{B}_{i-1}}(G_i) = C_\varepsilon. \tag{4.15}$$

Equations (4.13), (4.14), and (4.15) entail (4.10). Finally, stationarity of (E_i) and Theorem 3.1 imply that (E_i) is the innovation of (Z_i). ■

Notice that regularity properties of ρ induce similar properties of R. For example, if ρ is a symmetric compact operator with spectral decomposition

$$\rho = \sum_j a_j e_j \otimes e_j,$$

then R is a symmetric compact operator over $\mathcal{S}$ with spectral decomposition

$$R = \sum_{i,j} a_i a_j s_{ij} \otimes s_{ij}.$$

The above autoregressive representation allows us to derive asymptotic results concerning C_n by using those in Chapter 3. In the remainder, Γ_S, $\Gamma_{S,S'}$, $\Gamma^*_{S,S'}$ will denote the covariance and cross-covariance operators of the $\mathcal{S}$-valued random variables S and S'.

THEOREM 4.1 *If A_1 holds then, as $n \to \infty$,*

$$n\Gamma_{(C_n - C)} \longrightarrow \sum_{h=-\infty}^{+\infty} \Gamma_{Z_0, Z_h}, \tag{4.16}$$

where convergence takes place in $\| \cdot \|_{\mathcal{N}}^{(\mathcal{S})}$ sense, and

$$n\, E \parallel C_n - C \parallel_{\mathcal{S}}^2 \longrightarrow \sum_{h=-\infty}^{+\infty} E < Z_0, Z_h >_{\mathcal{S}}\ . \tag{4.17}$$

Moreover, for all $\beta > \frac{1}{2}$,

$$n^{1/4} (\operatorname{Log} n)^{-\beta} \parallel C_n - C \parallel_{\mathcal{S}} \longrightarrow 0 \ \ a.s. \tag{4.18}$$

Proof

From Lemma 4.1 it follows that the $AR\mathcal{S}(1)$ process Z satisfies assumptions in Theorem 3.8.

Now note that

$$C_n - C = \frac{Z_1 + \ldots + Z_n}{n}, \quad n \geq 1; \tag{4.19}$$

hence (3.29) gives (4.16) and (3.30) gives (4.17).

Concerning (4.18), note first that Lemma 3.2 applied to Z yields

$$|E < Z_0, Z_h >_{\mathcal{S}} | \leq \| R^h \|_{\mathcal{L}}^{(\mathcal{S})} E \| Z_0 \|_{\mathcal{S}}^2 .$$

Therefore

$$|E < Z_0, Z_h >_{\mathcal{S}} | \leq 2 \| \rho^h \|_{\mathcal{L}}^2 E \| X_0 \|^4, \quad h \geq 0, \tag{4.20}$$

which entails

$$E \| Z_n + \ldots + Z_{n+p-1} \|_{\mathcal{S}}^2 \leq 4E \| X_0 \|^4 \sum_{h=0}^{\infty} \| \rho^h \|_{\mathcal{L}}^2 \cdot p.$$

We now may apply Corollary 2.3 to Z; hence (4.18) holds. ■

It should be noticed that, since

$$\| (C_n - C)^2 \|_{\mathcal{N}} = \| C_n - C \|_{\mathcal{S}}^2,$$

(4.17) may be written in the form

$$n\, E \| (C_n - C)^2 \|_{\mathcal{N}} \longrightarrow \| \sum_{h=-\infty}^{+\infty} \Gamma_{Z_0, Z_h} \|_{\mathcal{N}}^{(\mathcal{S})} . \tag{4.21}$$

We now turn to **large deviations** for $C_n - C$. Let us begin with the bounded case.

THEOREM 4.2 *If A_1 holds and $\| X_0 \|$ is bounded, then for all $\eta > 0$*

$$P(\| C_n - C \|_{\mathcal{S}} \geq \eta) \leq 4 \exp\left(- \frac{n\, \eta^2}{\alpha_1 + \beta_1 \eta} \right), \tag{4.22}$$

where α_1 and β_1 are explicit positive numbers that depend only on ρ and P_{ε_0}.

Proof

From the autoregressive representation (4.6) it follows that we have

$$C_n - C = (I - R)^{-1} \frac{E_1 + \ldots + E_n}{n} + \Delta_n, \tag{4.23}$$

where

$$\Delta_n = (I-R)^{-1}\frac{R^n(E_1)+\ldots+R(E_n)}{n} + \frac{(R+\ldots+R^n)(Z_0)}{n}. \quad (4.24)$$

Therefore, for all $\eta > 0$, we may write

$$P(\| C_n - C \|_{\mathcal{S}} \geq \eta) \leq P\left(\| \sum_{i=1}^{n} E_i \|_{\mathcal{S}} \geq n\, a\, \eta\right) + P\left(\| \Delta_n \|_{\mathcal{S}} \geq \frac{\eta}{2}\right), \quad (4.25)$$

where $a = \frac{1}{2}\left[\| (I-R)^{-1} \|_{\mathcal{L}}^{(\mathcal{S})}\right]^{-1}$.

Now it is clear that boundedness of $\| X_0 \|$ yields the same property for $\| \varepsilon_0 \|$, $\| Z_0 \|_{\mathcal{S}}$ and, $\| E_0 \|_{\mathcal{S}}$. Then we may apply inequality (2.57) to the martingale difference (E_i) to obtain

$$P\left(\| \sum_{i=1}^{n} E_i \|_{\mathcal{S}} \geq n\, a\, \eta\right) \leq 2\exp\left(-\frac{na^2\eta^2}{2b^2 + \frac{2}{3}ba\eta}\right), \quad (4.26)$$

where b is any number satisfying $b \geq \| E_0 \|_{\infty,\mathcal{S}}$.

Similarly, noting that the finite sequence

$$(R+\ldots+R^n)(Z_0),\ (I-R)^{-1}R^n(E_1),\ \ldots,(I-R)^{-1}R(E_n)$$

is a bounded martingale difference with respect to the σ-algebras $\mathcal{B}_0,\ldots,\mathcal{B}_{n-1}$, we may again apply (2.57), which gives

$$P\left(\| \Delta_n \|_{\mathcal{S}} \geq \frac{\eta}{2}\right) \leq 2\exp\left(-\frac{n\eta^2}{8b'^2 + \frac{4}{3}b'\eta}\right), \quad (4.27)$$

where

$$b' \geq \max(\| (R+\ldots+R^n)(Z_0) \|_{\infty,\mathcal{S}},\ \| (I-R)^{-1}R^n(E_1) \|_{\infty,\mathcal{S}},\ldots$$
$$\| (I-R)^{-1}R(E_n) \|_{\infty,\mathcal{S}}).$$

Note that b' can be choosen independently on n, since

$$\| R^n \|_{\mathcal{L}}^{(\mathcal{S})} \to 0 \text{ and } \| R+\ldots+R^n \|_{\mathcal{L}}^{(\mathcal{S})} \longrightarrow \| (I-R)^{-1} \|_{\mathcal{L}}^{(\mathcal{S})}.$$

Finally applying (4.26) and (4.27) to (4.25) we obtain the desired inequality. ■

COROLLARY 4.1

$$\| C_n - C \|_{\mathcal{S}} = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \quad (4.28)$$

Proof

Choose $\eta = A\left(\frac{\text{Log}\, n}{n}\right)^{1/2}$; then (4.22) gives

$$\begin{aligned} P\left(\left(\frac{n}{\text{Log}\, n}\right)^{1/2} \| C_n - C \|_{\mathcal{S}} \geq A\right) &\leq 4\exp\left(-\frac{A^2 \,\text{Log}\, n}{\alpha_1 + \beta_1 \left(\frac{\text{Log}\, n}{n}\right)^{1/2} A}\right) \\ &\leq 4n^{-\frac{A^2}{\alpha_1 + \beta_1 A}}. \end{aligned}$$

Now take A such that $A^2 > \alpha_1 + \beta_1 A$ and apply the Borel-Cantelli Lemma; (4.28) follows. ■

If X is strongly mixing, the boundedness condition may be replaced by existence of some exponential moments. In the next statement, $(\alpha_k,\ k \geq 1)$ denotes the sequence of strong mixing coefficients of X and $\left(\lambda_j^Z,\ j \geq 1\right)$ the sequence of eigenvalues of the covariance operator Γ_{Z_0} of $Z_0 = < X_0, \cdot > X_0 - C$ in the Hilbert space $\mathcal{S}$.

THEOREM 4.3 *If* A_1 *holds and* $E\left(e^{\gamma \|X_0\|^2}\right) < \infty$ *for some* $\gamma > 0$, *then*

$$\begin{aligned} P(\| C_n - C \|_{\mathcal{S}} > \varepsilon) &\leq 4\nu \exp\left(-\frac{(1-\delta)^3 \varepsilon^2}{128 \nu m^2} q\right) \\ &+ 22\nu \left(1 + \frac{16 m \sqrt{\nu}}{\varepsilon (1-\delta)^{3/2}}\right)^{1/2} q \alpha_{[n/2q]} + \frac{4}{\delta (1-\delta)^2 \varepsilon^2} \sum_{j > \nu} \lambda_j^Z \\ &+ n e^{-\gamma m} E\left(e^{\gamma \|X_0\|^2}\right) e^{\gamma \|C\|_{\mathcal{S}}}; \end{aligned} \tag{4.29}$$

$\nu = 1, 2, \ldots;\ n \geq 2;\ q = 1, \ldots \left[\frac{n}{2}\right];\ m > \frac{2}{\gamma} \text{Log}\, \frac{2c'}{(1-\delta)\varepsilon}$, *where*

$$c' = 2\left[e^{\gamma \|C\|_{\mathcal{S}}} E\left(e^{\gamma \|X_0\|^2}\right) \left(E \| X_0 \|^4 + \| C \|_{\mathcal{S}}^2\right)\right]\ ;\ 0 < \delta < 1\ ;\ \varepsilon > 0.$$

Proof

The $\mathcal{S}$-valued process $(Z_i,\ i \geq 1)$ satisfies the conditions in Theorem 2.13 with mixing coefficients $\left(\alpha_k^Z\right)$ such that $\alpha_k^Z \leq \alpha_k$, $k \geq 1$, and

$$E\left(e^{\gamma \|Z_0\|_{\mathcal{S}}}\right) \leq e^{\gamma \|C\|_{\mathcal{S}}} E\left(e^{\gamma \|X_0\|^2}\right) < \infty.$$

Then (2.50) implies (4.29). ■

COROLLARY 4.2 *If there exist* $a > 0$ *and* $r \in]0, 1[$ *such that*

$$\alpha_k \leq a r^k\ ,\quad k \geq 1, \tag{4.30}$$

and

$$\lambda_j^Z \leq ar^j \ , \ \ j \geq 1, \tag{4.31}$$

then

$$P(\| C_n - C \|_{\mathcal{S}} > \varepsilon) \leq k_1' \exp\left(-k_2' n^{1/5}\right), \tag{4.32}$$

$\varepsilon > 0$, $n \geq 1$, *where* k_1' *and* k_2' *depend only on* ε *and* P_X.

Moreover, as $n \to \infty$,

$$\| C_n - C \|_{\mathcal{S}} = O\left(\frac{(\text{Log } n)^{5/2}}{n^{1/2}}\right) \ \ a.s. \tag{4.33}$$

Proof

Similar to the proof of Corollary 2.4 and therefore omitted. ■

Condition (4.30) is satisfied if X is a Gaussian $AR(1)$ process. Condition (4.31) is more restrictive and must be verified in each case. It is of course satisfied if X takes its values in H, which has a finite but "large" and unknown dimension! This kind of space appears in several practical applications (see Grenander (1981)).

4.2 Estimation of the eigenelements of C

Estimation of eigenvalues and eigenvectors of C is of great interest, since it is connected with **data analysis** of the observed process (see Ramsay and Silverman (1997)). Furthermore, we shall use estimators of these eigenelements to construct consistent estimators of ρ and statistical predictors.

Recall that C admits the spectral decomposition

$$C = \sum_{j=1}^{\infty} \lambda_j v_j \otimes v_j, \tag{4.34}$$

where (v_j) is a complete orthonormal system in H and (λ_j) is a decreasing sequence of positive real numbers such that $\sum_{j=1}^{\infty} \lambda_j < \infty$. We thus have

$$C(v_j) = \lambda_j v_j \ , \ \ j \geq 1. \tag{4.35}$$

Natural estimators of these parameters are the **empirical eigenelements** defined by

$$C_n(v_{jn}) = \lambda_{jn} \, v_{jn} \ , \ \ j \geq 1, \tag{4.36}$$

where $\lambda_{1n} \geq \ldots \geq \lambda_{nn} \geq 0 = \lambda_{n+1,n} = \lambda_{n+2,n} = \ldots$ and (v_{jn}) constitutes a complete orthonormal system in H.

In the remainder we shall use the vectors

$$v_{jn}' = sgn < v_{jn}, v_j > v_j \ , \ \ j \geq 1,$$

where $sgn\ u = \mathbb{I}_{u \geq 0} - \mathbb{I}_{u<0},\ u \in \mathbb{R}$.

Note that, since v_j and $-v_j$ are both eigenvectors associated with λ_j, the statistical parameter v_j is not well defined even if the associated eigensubspace, say $\mathcal{V}_j$, is one dimensional. Actually, the genuine parameter is $\mathcal{V}_j$. This explains introduction of (v'_{jn}).

The following technical lemma provides a crucial inequality.

LEMMA 4.2 *Let ℓ_0 be a compact linear operator on H with spectral decomposition*

$$\ell_0 = \sum_{j=1}^{\infty} \alpha_{j,0} e_{j,0} \otimes f_{j,0}.$$

Then

$$\alpha_{j,0} = \min_{\ell \in \mathcal{L}_{j-1}} \| \ell - \ell_0 \|_{\mathcal{L}},\ \ j \geq 1, \tag{4.37}$$

where $\mathcal{L}_{j-1} = \{\ell : \ell \in \mathcal{L}\ ,\ \ \dim \ell(H) \leq j-1\}$.

Further, if $\ell_1 = \sum_{j=1}^{\infty} \alpha_{j,1} e_{j,1} \otimes f_{j,1}$ *is another compact linear operator on H, we have*

$$|\alpha_{j,1} - \alpha_{j,0}| \leq \| \ell_1 - \ell_0 \|_{\mathcal{L}}\ ,\ \ j \geq 1. \tag{4.38}$$

Proof

If $j = 1$, (4.37) is obvious. We now suppose that $j > 1$ and set

$$N^{\perp} = \{x : \ell(x) = 0\}^{\perp}\ \ ,\ \ \ell \in \mathcal{L}_{j-1},\ \dim \ell(H) = j-1.$$

We thus have

$$\alpha_{j,0} \leq \sup_{x \in N} \frac{\| \ell_0(x) \|}{\| x \|}$$

and, if $x \in N$,

$$\| \ell_0(x) \| = \| (\ell_0 - \ell)(x) \| \leq \| \ell_0 - \ell \|_{\mathcal{L}} \| x \|;$$

therefore

$$\alpha_{j,0} \leq \| \ell_0 - \ell \|_{\mathcal{L}}\ .$$

On the other hand,

$$\alpha_{j,0} = \| \ell_0 - \ell_{0,j-1} \|_{\mathcal{L}},$$

where

$$\ell_{0,j-1} = \sum_{k=1}^{j-1} \alpha_{k,0} e_{k,0} \otimes f_{k,0};$$

hence (4.37) holds.

Concerning (4.38) we may write

$$\begin{aligned} \alpha_{j,0} &= \min\nolimits_{\ell \in \mathcal{L}_{j-1}} \| \ell_0 - \ell \|_{\mathcal{L}} \leq \| \ell_0 - \ell_1 \|_{\mathcal{L}} + \min_{\ell \in \mathcal{L}_{j-1}} \| \ell_1 - \ell \|_{\mathcal{L}} \\ &\leq \| \ell_0 - \ell_1 \|_{\mathcal{L}} + \alpha_{j,1}. \end{aligned}$$

Interchanging the roles of ℓ_0 and ℓ_1, we obtain $\alpha_{j,1} \leq \| \ell_1 - \ell_0 \|_{\mathcal{L}} + \alpha_{j,0}$, and (4.38) follows. ■

We are now in a position to state consistency results for empirical eigenvalues.

THEOREM 4.4 *Let us suppose that A_1 holds. We first have*

$$\limsup_{n\to\infty} n\, E\left[\sup_{j\geq 1} |\lambda_{jn} - \lambda_j|^2\right] \leq \sum_{h=-\infty}^{+\infty} E < Z_0,\ Z_h >_{\mathcal{S}}, \tag{4.39}$$

and for all $\beta > \frac{1}{2}$,

$$n^{1/4} (\mathrm{Log}\, n)^{-\beta} \sup_{j\geq 1} |\lambda_{jn} - \lambda_j| \underset{a.s.}{\longrightarrow} 0. \tag{4.40}$$

If in addition $\| X_0 \|$ is bounded, then

$$\sup_{j\geq 1} |\lambda_{jn} - \lambda_j| = O\left(\left(\frac{\mathrm{Log}\, n}{n}\right)^{1/2}\right) \quad a.s. \tag{4.41}$$

Finally, if X satisfies (4.30), (4.31), and $E\left(e^{\gamma \| X_0 \|^2}\right) < \infty$ for some $\gamma > 0$, then

$$\sup_{j\geq 1} |\lambda_{jn} - \lambda_j| = O\left(\frac{(\mathrm{Log}\, n)^{5/2}}{n^{1/2}}\right) \quad a.s. \tag{4.42}$$

Proof

Straightforward since inequality (4.38) yields

$$\sup_{j\geq 1} |\lambda_{jn} - \lambda_j| \leq \| C_n - C \|_{\mathcal{L}} \leq \| C_n - C \|_{\mathcal{S}} . \tag{4.43}$$

Then (4.17), (4.18), (4.28), and (4.33) respectively imply (4.39), (4.40), (4.41), and (4.42). ■

We now turn to estimation of $(v_j,\ j \geq 1)$. For this purpose we need some bounds.

LEMMA 4.3 *If $\mathcal{V}_j$ is one dimensional, then*

$$\| v_{jn} - v'_{jn} \| \leq a_j \| C_n - C \|_{\mathcal{L}} , \tag{4.44}$$

where

$$a_j = 2\sqrt{2} \max\left[(\lambda_{j-1} - \lambda_j)^{-1},\ (\lambda_j - \lambda_{j+1})^{-1}\right] \tag{4.45}$$

if $j \geq 2$, and

$$a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}. \tag{4.46}$$

Proof

First note that

$$C(v_{jn}) - \lambda_j v_{jn} = (C - C_n)(v_{jn}) + (\lambda_{jn} - \lambda_j)v_{jn};$$

then (4.43) yields

$$\| C(v_{jn}) - \lambda_j v_{jn} \| \leq 2 \| C - C_n \|_{\mathcal{L}} . \tag{4.47}$$

On the other hand,

$$\begin{aligned} \| v_{jn} - v'_{jn} \|^2 &= \sum_{\ell=1}^{\infty} (< v_{jn}, v_\ell > - sgn < v_{jn}, v_\ell >< v_j, v_\ell >)^2 \\ &= (< v_{jn}, v_j > - sgn < v_{jn}, v_j >)^2 + \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 \\ &= [1 - | < v_{jn}, v_j > |]^2 + \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 . \end{aligned}$$

However,

$$\sum_{\ell=1}^{\infty} < v_{jn}, v_\ell >^2 = 1;$$

hence

$$\begin{aligned} &[1 - | < v_{jn}, v_j > |]^2 = \sum_{\ell=1}^{\infty} < v_{jn}, v_\ell >^2 - 2| < v_{jn}, v_j > | + < v_{jn}, v_j >^2 \\ &= \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 + 2[< v_{jn}, v_j >^2 - | < v_{jn}, v_j > |] \\ &\leq \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 . \end{aligned}$$

Therefore

$$\| v_{jn} - v'_{jn} \|^2 \leq 2 \sum_{\ell \neq j} (< v_{jn}, v_\ell >^2). \tag{4.48}$$

Now

$$\begin{aligned} &\| C(v_{jn}) - \lambda_j v_{jn} \|^2 = \sum_{\ell=1}^{\infty} (< C(v_{jn}), v_\ell > - < \lambda_j v_{jn}, v_\ell >)^2 \\ &= \sum_{\ell \neq j} < v_{jn}, (\lambda_\ell - \lambda_j) v_\ell >^2 \geq \min_{\ell \neq j} |\lambda_\ell - \lambda_j|^2 \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 . \end{aligned}$$

Putting

$$\alpha_j = \min(\lambda_{j-1} - \lambda_j, \ \ \lambda_j - \lambda_{j+1}) \ , \ \ j \geq 2,$$

and

$$\alpha_1 = \lambda_1 - \lambda_2,$$

we get

$$\| C(v_{jn}) - \lambda_j v_{jn} \|^2 \geq \alpha_j^2 \sum_{\ell \neq j} < v_{jn}, v_\ell >^2 .$$

Then from (4.47) and (4.48) it follows that

$$\| v_{jn} - v'_{jn} \| \leq \frac{2\sqrt{2}}{\alpha_j} \| C_n - C \|_{\mathcal{L}} , \quad j \geq 1,$$

which is the desired result. ■

Now we have asymptotic results for (v_{jn}).

THEOREM 4.5 *Suppose that A_1 holds.*
If $\mathcal{V}_j$ is one dimensional, we have

$$\limsup_{n\to\infty} E \| v_{jn} - v'_{jn} \|^2 \leq a_j^2 \sum_{h=-\infty}^{+\infty} E < Z_0, \ Z_h >_{\mathcal{S}}, \tag{4.49}$$

and for all $\beta > \frac{1}{2}$,

$$n^{1/4} (\text{Log } n)^{-\beta} \| v_{jn} - v'_{jn} \| \underset{a.s.}{\longrightarrow} 0. \tag{4.50}$$

If in addition $\| X_0 \|$ is bounded, then

$$\| v_{jn} - v'_{jn} \| = O\left(\left(\frac{\text{Log } n}{n} \right)^{1/2} \right) \quad a.s. \tag{4.51}$$

Finally, if X satisfies (4.30), (4.31) and $E\left(e^{\gamma \| X_0 \|^2}\right) < \infty$ for some $\gamma > 0$, then

$$\| v_{jn} - v'_{jn} \| = O\left(\frac{(\text{Log } n)^{5/2}}{n^{1/2}} \right) \quad a.s. \tag{4.52}$$

Proof

Clear since Lemma 4.3 allows us to use the asymptotic results concerning $\| C_n - C \|_{\mathcal{S}}$. ■

To reach uniform convergence with respect to j is more difficult for the v_j's than for the λ_j's. This is because eigenvectors are parameters much more sensitive to variation of operators than eigenvalues. However, it is possible to obtain partial results.

COROLLARY 4.3 *Suppose that $\lambda_j \downarrow 0$ and set*

$$\Lambda_k = \sup_{1 \le j \le k} (\lambda_j - \lambda_{j+1})^{-1}, \; k \ge 1.$$

If (k_n) is a sequence of integers such that $\Lambda_{k_n} = o(n^{1/2})$, we have

$$E\left[\sup_{1 \le j \le k_n} \| v_{jn} - v'_{jn} \|^2\right] \to 0. \tag{4.53}$$

If $\| X_0 \|$ is bounded and if $\Lambda_{k_n} = o\left(\left(\frac{n}{\operatorname{Log} n}\right)^{1/2}\right)$, then

$$\sup_{1 \le j \le k_n} \| v_{jn} - v'_{jn} \| \underset{a.s.}{\longrightarrow} 0. \tag{4.54}$$

Proof

Lemma 4.3 yields

$$\sup_{1 \le j \le k_n} \| v_{jn} - v'_{jn} \|^2 \le 8\Lambda_{k_n}^2 \| C_n - C \|_{\mathcal{L}}^2;$$

hence

$$E\left[\sup_{1 \le j \le k_n} \| v_{jn} - v'_{jn} \|^2\right] \le 8\frac{\Lambda_{k_n}^2}{n} n\, E \| C_n - C \|_{\mathcal{S}}^2,$$

and (4.53) follows from (4.17) and the condition $\Lambda_{k_n} = o(n^{1/2})$.

Concerning pointwise convergence we may use the large deviation inequality (4.22) to obtain

$$P\left(\sup_{1 \le j \le k_n} \| v_{jn} - v'_{jn} \| \ge \eta\right) \le P\left(\| C_n - C \|_{\mathcal{S}} \ge \frac{\eta}{2\sqrt{2}\Lambda_{k_n}}\right)$$

$$\le 4 \exp\left(-\frac{n\eta^2}{8\alpha\Lambda_{k_n}^2 + \dfrac{4\eta}{\sqrt{2}}\beta\Lambda_{k_n}}\right), \quad \eta > 0.$$

Then $\Lambda_{k_n} = o\left(\left(\frac{n}{\operatorname{Log} n}\right)^{1/2}\right)$ and Borel-Cantelli entail (4.54). ■

Note that if there exists a convex function φ such that $\varphi(j) = \lambda_j, \; j \ge 1$, then Λ_k has the simpler form

$$\Lambda_k = (\lambda_k - \lambda_{k+1})^{-1}.$$

Let us now consider the case of a (possibly) multidimensional $\mathcal{V}_j$. In the following lemma, Π_a^b stands for the orthogonal projector of the eigensubspace $\mathcal{V}_a^b$ generated by $\{v_a, \ldots, v_b\}$ $(a \le b)$, where $v_a, \ldots, v_b$ are associated with the eigenvalue $\lambda_a (= \lambda_b)$.

LEMMA 4.4 *If $j \in \{a, \dots, b\}$ then*

$$\| v_{jn} - \Pi_a^b(v_{jn}) \| \leq c_j \| C_n - C \|_{\mathcal{L}}, \tag{4.55}$$

where

$$c_j = 2 \max[(\lambda_{a-1} - \lambda_j)^{-1}, \ (\lambda_j - \lambda_{b+1})^{-1}] \ \textit{if}\ a \geq 2$$

and

$$c_j = 2(\lambda_j - \lambda_{b+1})^{-1} \ \textit{if}\ a = 1.$$

Proof

Consider the decomposition

$$\begin{aligned} \| v_{jn} - \Pi_a^b(v_{jn}) \|^2 &= \sum_{\ell=1}^{\infty} < v_{jn} - \Pi_a^b(v_{jn}), v_\ell >^2 \\ &= \sum_{\ell \notin I} < v_{jn}, v_\ell >^2, \end{aligned} \tag{4.56}$$

where $I = \{a, \dots, b\}$.

On the other hand, note that

$$\| C(v_{jn}) - \lambda_j v_{jn} \|^2 = \sum_{\ell=1}^{\infty} < C(v_{jn}) - \lambda_j v_{jn}, \ v_\ell >^2$$

$$= \sum_{\ell=1}^{\infty} (\lambda_\ell - \lambda_j)^2 < v_{jn}, v_\ell >^2 = \sum_{\ell \notin I} (\lambda_\ell - \lambda_j)^2 < v_{jn}, v_\ell >^2$$

$$\geq \inf_{\ell \notin I} (\lambda_\ell - \lambda_j)^2 \sum_{\ell \notin I} < v_{jn}, v_\ell >^2$$

$$\geq \frac{4}{c_j^2} \sum_{\ell \notin I} < v_{jn}, v_\ell >^2 .$$

Then (4.56) entails

$$\| v_{jn} - \Pi_a^b(v_{jn}) \| \leq \frac{c_j}{2} \| C(v_{jn}) - \lambda_j v_{jn} \|,$$

and finally (4.47) yields (4.55). ■

From Lemma 4.4 one may easily deduce results similar to Theorem 4.5 and Corollary 4.3. This is straightforward and therefore omitted.

In order to complete the above results, we now construct estimators of the dimensions of eigensubspaces. Let $d_1, d_2, \dots$ be the sequence of these dimensions associated with the decreasing (in the wide sense) sequence $\lambda_1, \lambda_2, \dots$

We first define an **estimator** of d_1. Consider a sequence (k_n) of integers such that $(k_n) \uparrow \infty$ and $\left(\frac{k_n}{n}\right) \to 0$, and a sequence $(\eta_n) \downarrow 0$ of positive numbers, and set

$$J_{1n} = \{j : 2 \leq j \leq k_n,\ |\lambda_{jn} - \lambda_{1n}| > \eta_n\}.$$

The estimator d_{1n} of d_1 is defined as

$$\begin{aligned} d_{1n} &= (\inf J_{1n}) - 1 \quad \text{if} \quad J_{1n} \neq \emptyset, \\ &= k_n - 1 \qquad\qquad \text{if} \quad J_{1n} = \emptyset. \end{aligned}$$

The estimator d_{jn} of d_j is defined recursively by

$$\begin{aligned} d_{jn} &= (\inf J_{jn}) - 1 \quad \text{if} \quad J_{jn} \neq \emptyset, \\ &= k_n - 1 \qquad\qquad \text{if} \quad J_{jn} = \emptyset, \end{aligned}$$

where

$$J_{jn} = \left\{j : d_{j-1,n} + 2 \leq j \leq k_n,\ |\lambda_{jn} - \lambda_{d_{j-1,n}+1,n}| > \eta_n\right\}.$$

We first have

THEOREM 4.6 *If $(k_n) \uparrow \infty$, $(\eta_n) \downarrow 0$, and $\left(\frac{k_n}{n\eta_n^2}\right) \to 0$, then*

$$d_{1n} \to d_1 \ \textit{in probability.} \tag{4.57}$$

Proof

For large enough n we have $k_n > d_1$. Now $\{d_{1n} < d_1\} = \emptyset$ if $d_1 = 1$ and, if $d_1 > 1$,

$$\{d_{1n} < d_1\} = \bigcup_{j=2}^{d_1} \{|\lambda_{jn} - \lambda_{1n}| > \eta_n\}.$$

Therefore

$$P(d_{1n} < d_1) \leq \sum_{j=2}^{d_1} P(|\lambda_{jn} - \lambda_{1n}| > \eta_n)$$

$$\leq \sum_{j=2}^{d_1} \left[P\left(|\lambda_{jn} - \lambda_{1n}| > \eta_n,\ |\lambda_{1n} - \lambda_1| \leq \frac{\eta_n}{2}\right) + P\left(|\lambda_{1n} - \lambda_1| > \frac{\eta_n}{2}\right)\right]$$

and since $\lambda_j = \lambda_1$ for $j \leq d_1$, we get

$$P(d_{1n} < d_1) \leq \sum_{j=1}^{d_1} P\left(|\lambda_{jn} - \lambda_j| > \frac{\eta_n}{2}\right) + (d_1 - 1)P\left(|\lambda_{1n} - \lambda_1| > \frac{\eta_n}{2}\right).$$

Now from (4.43) it follows that

$$P(d_{1n} < d_1) \leq 2(d_1 - 1)P\left(\| C_n - C \|_{\mathcal{S}} > \frac{\eta_n}{2}\right) \leq \frac{8d_1}{\eta_n^2} E \| C_n - C \|_{\mathcal{S}}^2 .$$

Using (4.17) and noting that $n\eta_n^2 \to \infty$, we obtain

$$P(d_{1n} < d_1) \underset{n\to\infty}{\longrightarrow} 0 . \tag{4.58}$$

On the other hand,

$$\begin{aligned}\{d_{1n} > d_1\} &= \bigcup_{j=d_1+1}^{k_n-1} \{d_{1n} = j\} \\ &\subset \{\exists j \in]\, d_1, k_n] : |\lambda_{jn} - \lambda_{1n}| \leq \eta_n\}.\end{aligned}$$

Thus

$$\begin{aligned}P(d_{1n} > d_1) &\leq \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_{1n}| \leq \eta_n) \\ &\leq \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_{1n}| \leq \eta_n,\ |\lambda_{1n} - \lambda_1| < \eta_n) \\ &\quad + k_n P(|\lambda_{1n} - \lambda_1| \geq \eta_n),\end{aligned}$$

and hence

$$\begin{aligned}P(d_{1n} > d_1) &\leq \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_1| < 2\eta_n) + k_n P(|\lambda_{1n} - \lambda_1| \geq \eta_n) \\ &\leq \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_j| < \eta_n,\ |\lambda_{jn} - \lambda_1| < 2\eta_n) \\ &\quad + \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_j| \geq \eta_n,\ |\lambda_{jn} - \lambda_1| < 2\eta_n) \\ &\quad + k_n P(|\lambda_{1n} - \lambda_1| \geq \eta_n).\end{aligned}$$

The first term in the above bound vanishes for large enough n, since $\lambda_j < \lambda_1$ for $j > d_1$. Then

$$P(d_{1n} > d_1) \leq \sum_{j=d_1+1}^{k_n} P(|\lambda_{jn} - \lambda_j| \geq \eta_n) + k_n P(|\lambda_{1n} - \lambda_1| \geq \eta_n).$$

Thus, using (4.43) again we obtain

$$\begin{aligned}P(d_{1n} > d_1) &\leq (2k_n - d_1) P(\| C_n - C \|_{\mathcal{S}} \geq \eta_n) \\ &< \frac{2k_n}{\eta_n^2} E \| C_n - C \|_{\mathcal{S}}^2 .\end{aligned}$$

Now, from (4.17) and $\dfrac{k_n}{n\eta_n^2} \to 0$, it follows that

$$P(d_{1n} > d_1) \underset{n\to\infty}{\longrightarrow} 0 . \tag{4.59}$$

Finally (4.58) and (4.59) lead to the desired result. ■

COROLLARY 4.4 *For all $j \geq 1$,*

$$d_{jn} \longrightarrow d_j \ \textit{ in probability.} \tag{4.60}$$

Proof

We have

$$\begin{aligned} P(d_{2n} \neq d_2) \leq \quad & P(d_{1n} \neq d_1) + P(d_{1n} = d_1,\ d_{2n} < d_2) \\ & + P(d_{1n} = d_1,\ d_{2n} > d_2). \end{aligned}$$

Now we have seen that $P(d_{1n} \neq d_1) \to 0$ and the study of the other terms in the bound can be performed similarly as for $P(d_{1n} < d_1)$ and $P(d_{1n} > d_1)$. Therefore

$$P(d_{2n} \neq d_2) \to 0 \ ;$$

(4.60) follows by induction. ■

A special case

We conclude this section with the study of alternative estimators for eigenvalues in the particular case where (v_j) is known. These natural estimators are defined as

$$\widehat{\lambda}_{jn} = \frac{1}{n}\sum_{i=1}^{n} < X_i, v_j >^2, \ \ j \geq 1,\ n \geq 1. \tag{4.61}$$

They are **unbiased**, since

$$E\widehat{\lambda}_{jn} = E(< X_0, v_j >^2) =< Cv_j, v_j >= \lambda_j .$$

Moreover, we have

COROLLARY 4.5 *Asymptotic results in Theorem 4.4 remain valid if λ_{jn} is replaced by $\widehat{\lambda}_{jn}$.*

Proof

It suffices to note that

$$\begin{aligned} \sup_{j\geq 1}\left|\widehat{\lambda}_{jn} - \lambda_j\right| \quad &= \quad \sup_{j\geq 1} | < C_n(v_j), v_j > - < C(v_j), v_j > | \\ &\leq \quad \| C_n - C \|_{\mathcal{L}} \leq \| C_n - C \|_{\mathcal{S}} . \end{aligned}$$

■

Observe, however, that $\left(\widehat{\lambda}_{jn}\right)$ is not decreasing with respect to j, unlike (λ_{jn}). On the other hand, we have

$$\sum_{j=1}^{\infty} \widehat{\lambda}_{jn} = \sum_{j=1}^{\infty} \lambda_{jn}. \tag{4.62}$$

The asymptotic behavior of this quantity appears in Corollary 6.2.

4.3 Estimation of the cross-covariance operators

Let X be a zero-mean ARH(1). Its **cross-covariance operator of order** $\boldsymbol{h}$ is defined as

$$C_h(x) = C_{X_0,X_h}(x) = E(< X_0, x > X_h), \quad x \in H, \tag{4.63}$$

or in a more concise form,

$$C_h = E(X_0 \otimes X_h), \quad h \geq 1,$$

where expectation is taken in $\mathcal{S}$.

A natural estimator of C_h is the **empirical cross-covariance operator of order** $\boldsymbol{h}$, defined as

$$C_{h,n} = \frac{1}{n-h} \sum_{i=1}^{n-h} X_i \otimes X_{i+h} \ , \quad x \in H, \tag{4.64}$$

provided $1 \leq h \leq n-1$. If $h > n-1$, we set $C_{h,n} = 0$.

$C_{h,n}$ is clearly a nuclear operator and hence a Hilbert-Schmidt operator. Moreover,

$$E\ C_{h,n} = C_h, \quad 1 \leq h \leq n-1, \tag{4.65}$$

where expectation is taken in $\mathcal{S}$.

For convenience we write D for C_1 and D_n for $C_{1,n}$.

Let us begin with a representation statement analogous to Lemma 4.1.

LEMMA 4.5 *Suppose that A_1 holds and consider the $\mathcal{S}$-valued process*

$$Z_i' = X_{i-1} \otimes X_i - D, \quad i \in \mathbb{Z}.$$

Then it satisfies the relation

$$Z_i' = R(Z_{i-1}') + E_i' \ , \quad i \in \mathbb{Z}, \tag{4.66}$$

where R is the bounded linear operator over $\mathcal{S}$ defined by

$$R(s) = \rho s \rho^*, \quad s \in \mathcal{S}$$

and where

$$E'_i = (X_{i-2} \otimes \varepsilon_i)\rho^* + \rho(\varepsilon_{i-1} \otimes X_{i-1}) + (\varepsilon_{i-1} \otimes \varepsilon_i) - \rho C_\varepsilon, \ i \in \mathbb{Z}.$$

The $\mathcal{S}$-valued process (E'_i) has the following properties: $E'_i \in L^2_{\mathcal{S}}(\Omega, \mathcal{B}_i, P)$, where $\mathcal{B}_i = \omega(\varepsilon_i, \varepsilon_{i-1}, \ldots)$, and

$$E^{\mathcal{B}_j}(E'_i) = 0; \ \ i,j \in \mathbb{Z}; \ j \leq i-2 \ ; \tag{4.67}$$

hence

$$E'_i \perp E'_j \ \ \textit{if} \ |i-j| \geq 2.$$

Proof

Using definition 3.3 of ARH(1) we obtain

$$\begin{aligned} Z'_i &= \ < \rho(X_{i-2}) + \varepsilon_{i-1}, \cdot > (\rho(X_{i-1}) + \varepsilon_i) - D \\ &= \ \rho\,[< X_{i-2}, \rho^*(\cdot) > \ X_{i-1} - D\rho^*] \\ &\quad + < X_{i-2}, \rho^*(\cdot) > \varepsilon_i + < \varepsilon_{i-1}, \cdot > \varepsilon_i \\ &\quad + \rho(< \varepsilon_{i-1}, \cdot > X_{i-1}) + \rho D \rho^* - D. \end{aligned}$$

Noting that

$$\rho D \rho^* - D = \rho(\rho C)\rho^* - \rho C = \rho(\rho C \rho^* - C) = -\rho C_\varepsilon,$$

we obtain (4.66).

We already have seen that $\| R^h \|^{\mathcal{S}}_{\mathcal{L}} \leq \| \rho^h \|^2_{\mathcal{L}}$ (cf. Lemma 4.1).

Now it is clear that $E'_i \in L^2_{\mathcal{S}}(\Omega, \mathcal{B}_i, P)$. Concerning (4.67) it suffices to show that

$$E^{\mathcal{B}_{i-2}}(E'_i) = 0.$$

For this purpose we use a natural orthonormal basis of $\mathcal{S}$ defined as

$$s_{\ell k} = e_\ell \otimes e_k, \quad \ell \in \mathbb{Z}, \ k \in \mathbb{Z},$$

where $(e_j, \ j \in \mathbb{Z})$ is an orthonormal basis of H.

We then have

$$\begin{aligned} < \ E^{\mathcal{B}_{i-2}}(E'_i), \ s_{\ell k} >_{\mathcal{S}} &= E^{\mathcal{B}_{i-2}}\,(< X_{i-2}, \rho^*(e_\ell) >< \varepsilon_i, e_l >) \\ &+ E^{\mathcal{B}_{i-2}}(< \varepsilon_{i-1}, e_\ell >< \rho(X_{i-1}), e_k >) \\ &+ E^{\mathcal{B}_{i-2}}(< \varepsilon_{i-1}, e_\ell >< \varepsilon_i, e_k >) - < \rho C_\varepsilon(e_\ell), e_k > . \end{aligned}$$

The first and third terms on the right of the above equality vanish and, for every $(\ell, k) \in \mathbb{Z}^2$,

$$E^{\mathcal{B}_{i-2}}(< \varepsilon_{i-1}, e_\ell >< \rho(X_{i-1}),\ e_k >)$$

$$= E^{\mathcal{B}_{i-2}}\,(< \varepsilon_{i-1}, e_\ell >< \varepsilon_{i-1}, \rho^*(e_k) >)$$

$$+ < \rho(X_{i-2}),\ \rho^*(e_k) > E^{\mathcal{B}_{i-2}}(< \varepsilon_{i-1}, e_\ell >)$$

$$= < C_\varepsilon(e_\ell),\ \rho^*(e_k) >=< (\rho C_\varepsilon)(e_\ell), e_k > .$$

Thus

$$E^{\mathcal{B}_{i-2}}[< \varepsilon_{i-1}, \cdot > \rho(X_{i-1})] = \rho C_\varepsilon,$$

and finally

$$E^{\mathcal{B}_{i-2}}(E_i') = 0.$$

■

It should be noticed that E_{i-1}' and E_i' are not orthogonal in general. For example, if $H = \mathbb{R}$, we have $E(E_i'\ E_{i-1}') = \dfrac{\rho^4\sigma^4}{1-\rho^2}$.

We now may specify asymptotic behavior of $D_n - D$.

THEOREM 4.7 *If A_1 holds, the following assertions are true:*

$$n\,\Gamma_{(D_n - D)} \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} \Gamma_{(Z_0', Z_h')}\ , \tag{4.68}$$

where convergence takes place in the $\| \cdot \|_{\mathcal{N}}^{(\mathcal{S})}$ sense, and

$$n\,E \parallel D_n - D \parallel_{\mathcal{S}}^2 \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < Z_0', Z_h' >_{\mathcal{S}}\ . \tag{4.69}$$

Proof

We cannot directly apply Theorem 3.8, since (E_i') is not exactly a white noise.

Let us consider the decomposition

$$Z_h' = R^h(Z_0') + V_h\ ,\ \ h \geq 1, \tag{4.70}$$

where

$$V_h = \sum_{j=0}^{h-1} R^j(E_{h-j}').$$

Noting that $Z'_0 = \sum_{j=0}^{\infty} R^j(E'_{-j})$, we infer that $E'_j \perp Z'_0$ for $j \geq 2$. Therefore (4.70) implies

$$\Gamma_{Z'_0, Z'_h} = \Gamma_{Z'_0, R^{h-1}(E'_1)} + \Gamma_{Z'_0, R^h(Z'_0)}, \ h \geq 2.$$

Taking nuclear norms and using (1.62), we obtain

$$\begin{aligned}\| \Gamma_{Z'_0, Z'_h} \|_{\mathcal{N}}^{(\mathcal{S})} \leq \ & \| R^{h-1} \|_{\mathcal{L}}^{(\mathcal{S})} E\left(\| Z'_0 \|_{\mathcal{S}} \| E'_1 \|_{\mathcal{S}}\right) \\ & + \| R^h \|_{\mathcal{L}}^{(\mathcal{S})} E\left(\| Z'_0 \|_{\mathcal{S}}^2\right).\end{aligned}$$

Recalling that $\| R^h \|_{\mathcal{L}}^{(\mathcal{S})} \leq \| \rho^h \|_{\mathcal{L}}^2$, we get

$$\| \Gamma_{Z'_0, Z'_h} \|_{\mathcal{N}}^{(\mathcal{S})} \leq c \max\left(\| \rho^{h-1} \|_{\mathcal{L}}^2, \| \rho^h \|_{\mathcal{L}}^2\right), \ h \geq 2,$$

where c is a positive constant.

Thus

$$\sum_{h=-\infty}^{+\infty} \| \Gamma_{Z'_0, Z'_h} \|_{\mathcal{N}}^{(\mathcal{S})} < \infty, \tag{4.71}$$

which yields absolute convergence of $\sum_{h=-\infty}^{+\infty} \Gamma_{Z'_0, Z'_h}$ in the Banach space of nuclear operators over $\mathcal{S}$.

Now, by stationarity, we have

$$(n-1)\Gamma_{(D_n - D)} = \sum_{h=-(n-2)}^{n-2} \left(1 - \frac{|h|}{n-1}\right) \Gamma_{Z'_0, Z'_h},$$

and (4.68) follows from (4.71).

Similarly, we may write

$$(n-1)\, E \| D_n - D \|_{\mathcal{S}}^2 = \sum_{h=-(n-2)}^{n-2} \left(1 - \frac{|h|}{n-1}\right) E < Z'_0, Z'_h >_{\mathcal{S}} .$$

Using (4.70) again we obtain

$$|E < Z'_0, Z'_h >_{\mathcal{S}}| \leq c \max\left(\| \rho^{h-1} \|_{\mathcal{L}}^2, \| \rho^h \|_{\mathcal{L}}^2\right), \tag{4.72}$$

$h \geq 2$; hence (4.69) holds. ■

Note that from (4.68) and (4.69) it follows that

$$\| \sum_{h=-\infty}^{+\infty} \Gamma_{Z'_0, Z'_h} \|_{\mathcal{N}}^{(\mathcal{S})} = \sum_{h=-\infty}^{+\infty} E < Z'_0, Z'_h >_{\mathcal{S}} .$$

Let us now turn to pointwise consistency.

In this context, results concerning D_n are similar to those concerning C_n. However, proofs must be slightly modified since (4.66) is not a full autoregressive representation. We summarize the main facts in the following statement.

THEOREM 4.8 *We suppose that A_1 holds.*

1. *For each $\beta > \frac{1}{2}$,*

$$n^{1/4}(\operatorname{Log} n)^{-\beta} \parallel D_n - D \parallel_{\mathcal{S}} \underset{a.s.}{\longrightarrow} 0. \tag{4.73}$$

2. *If $\parallel X_0 \parallel$ is bounded, then*

$$P(\parallel D_n - D \parallel_{\mathcal{S}} > \eta) \leq 8 \exp\left(-\frac{n\eta^2}{\gamma + \delta\eta}\right), \ \eta > 0, \tag{4.74}$$

where γ and δ are explicit positive numbers that depend only on ρ and P_{ε_0}. Consequently,

$$\parallel D_n - D \parallel_{\mathcal{S}} = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \tag{4.75}$$

3. *If $E\left(e^{\gamma \|X_0\|^2}\right) < \infty$ for some $\gamma > 0$, X is GSM, and $\lambda_j^{Z'} \leq ar^j$, $j \geq 1$ $(a > 0,\ 0 < r < 1)$, where $\left(\lambda_j^{Z'}\right)$ is the sequence of eigenvalues of $\Gamma_{Z_0'}$, then*

$$\parallel D_n - D \parallel_{\mathcal{S}} = O\left(\frac{(\operatorname{Log} n)^{5/2}}{n^{1/2}}\right) \quad a.s. \tag{4.76}$$

Proof

1. From (4.72) it is easy to deduce that

$$E \parallel Z_n' + \ldots + Z_{n+p-1}' \parallel_{\mathcal{S}}^2 \leq d.p,\ p \geq 1,\ n \geq 1,$$

where d is a positive constant. Then, applying Corollary 2.3 to the process (Z_i'), one obtains (4.73).

2. Consider the following four sequences of $\mathcal{S}$-random variables:

$$(E_{2i}',\ i \geq 1),\ (E_{2i+1}',\ i \geq 0),$$

$$\left[(R + \ldots + R^n)(Z_0'),\ (I - R)^{-1}R^{n-1}(E_2'), \ldots, (I - R)^{-1}R^{n-2p+1}(E_{2p}')\right],$$

where $2p \leq n < 2p+2$, and

$$\left[(I-R)^{-1}R^{n}(E_1'),\dots,(I-R)^{-1}R^{n-2p+2}(E_{2p-1}')\right],$$

where $2p-1 \leq n < 2p+1$.
They are bounded differences of martingales with respect to suitable subsequences of $(\mathcal{B}_i,\ i \geq 0)$. Thus the idea of the proof of (4.74) is to split up $D_n - D$ under the form

$$D_n - D = (I-R)^{-1}\frac{E_2' + \ldots + E_{2p}'}{n} + (I-R)^{-1}\frac{E_1' + \ldots + E_{2p-1}'}{n}$$

$$+ (I-R)^{-1}\frac{R^n(E_1') + \ldots + R^{n-2p+2}(E_{2p-1}')}{n} + \frac{(R + \ldots + R^n)(Z_0)}{n}$$

$$+ \frac{(I-R)^{-1}R^{n-1}(E_2') + \ldots + (I-R)^{-1}R^{n-2p+1}(E_{2p}')}{n},$$

and then to apply (2.57) four times. Details are left to the reader. Finally (4.75) follows directly from (4.74).

3. Proof of (4.76) is similar to that of (4.33) and is therefore omitted.

■

Estimation of the eigenelements of D^*D

The nuclear operator D has spectral decomposition of the form

$$D = \sum_{j=1}^{\infty} \mu_j f_j \otimes g_j,$$

where $\sum_{j=1}^{\infty} \mu_j < \infty$ (see Section 1.5).
Consequently we have

$$D^*D = \sum_{j=1}^{\infty} \mu_j^2 f_j \otimes f_j;$$

thus D^*D is a positive symmetric nuclear operator.
A natural estimator of D^*D is $D_n^*D_n$, is given by the expression

$$D_n^*D_n = \frac{1}{(n-1)^2}\sum_{1\leq i,j\leq n-1} < X_{j+1}, X_{i+1} >< X_i, \cdot > X_j. \qquad (4.77)$$

We then define empirical estimators (μ_{jn}), (f_{jn}), and (g_{jn}) of (μ_j), (f_j), and (g_j), respectively, by setting

$$\left(D_n^*D_n\right)^{1/2}(f_{jn}) = \mu_{jn} f_{jn}\ ,\quad j \geq 1,$$

and

$$D_n(f_{jn}) = \mu_{jn} g_{jn} \ , \ \ j \geq 1.$$

Since consistency of $D_n^* D_n$ comes from the inequality

$$\| D_n^* D_n - D^* D \|_{\mathcal{N}} \leq (\| D_n \|_{\mathcal{S}} + \| D \|_{\mathcal{S}}) \| D_n - D \|_{\mathcal{S}}, \tag{4.78}$$

it is not difficult to derive consistency results concerning the above empirical eigenelements. Again the work is left to the reader!

Estimation of C_h

In order to estimate C_h and the eigenelements of $C_h^* C_h$, one may derive a representation of the form

$$Z_{i,h} = R^h(Z_{i-h,h}) + F_{i,h}, \ \ i \in \mathbb{Z}, \tag{4.79}$$

with

$$Z_{i,h} = X_{i-h} \otimes X_i - C_h, \ \ i \in \mathbb{Z},$$

and

$$E^{\mathcal{B}_j}(F_{i,h}) = 0 \ \ , \ \ \ j \leq i - h - 1.$$

Thus

$$F_{\ell,h} \perp F_{k,h} \ , \ \ |\ell - k| \geq h + 1.$$

From (4.79) one may obtain results analogous to those obtained for D. Methods of proof are clearly similar.

4.4 Limits in distribution

In order to study asymptotic normality of our estimators, we need a new representation.

Consider $\mathcal{S} \times \mathcal{S}$ equipped with the scalar product

$$< (s_1, s_2), \ (s_1', s_2') >_{\mathcal{S} \times \mathcal{S}} = < s_1, s_1' >_{\mathcal{S}} + < s_2, s_2' >_{\mathcal{S}} \ ;$$

$$(s_1, s_2) \in \mathcal{S} \times \mathcal{S}, \ \ (s_1', s_2') \in \mathcal{S} \times \mathcal{S}.$$

We now exhibit an autoregressive representation in $\mathcal{S} \times \mathcal{S}$.

LEMMA 4.6 *Suppose again that A_1 holds, and put*

$$T_i = (X_i \otimes X_i - C \ , \ \ X_{i-1} \otimes \varepsilon_i) \ , \ \ i \in \mathbb{Z}.$$

Then $(T_i, \ i \in \mathbb{Z})$ is an $(\mathcal{S} \times \mathcal{S})$-valued random sequence that admits autoregressive representation

$$T_i = R'(T_{i-1}) + W_i, \ \ i \in \mathbb{Z}, \tag{4.80}$$

where R' is the bounded linear operator over $\mathcal{S} \times \mathcal{S}$ defined by

$$R'(s_1, s_2) = (\rho s_1 \rho^*, 0); \quad s_1 \in \mathcal{S},\ s_2 \in \mathcal{S},$$

and

$$W_i = (E_i, F_i), \quad i \in \mathbb{Z},$$

with $E_i = F_i\rho^* + \rho F_i^* + G_i - C_\varepsilon$, $F_i = X_{i-1} \otimes \varepsilon_i$, and $G_i = \varepsilon_i \otimes \varepsilon_i$.

Moreover, (W_i) is a martingale difference with respect to $(\mathcal{B}_i)$.

Proof

Combining results in Lemma 4.1 and the fact that (F_i) is a martingale difference with respect to $(\mathcal{B}_i)$, it is easy to arrive at (4.80) and to prove that (W_i) is a martingale difference. Details are omitted. ■

We may now prove a general weak convergence result.

THEOREM 4.9 *If A_1 holds, then*

$$\Sigma_n =: \sqrt{n}(C_n - C,\ D_n - D) \xrightarrow{\mathcal{D}} N, \tag{4.81}$$

where weak convergence takes place in $\mathcal{S} \times \mathcal{S}$ and where N is a zero-mean $(\mathcal{S} \times \mathcal{S})$-valued Gaussian random variable.

N has the representation $(N_1, \rho N_1 + N_2)$, where (N_1, N_2) is Gaussian with covariance operator $(I - R')^{-1}\Gamma_{W_1}(I - R'^*)^{-1}$.

Proof

We establish (4.81) in three steps. We first derive the asymptotic distribution of $\dfrac{W_1 + \ldots + W_n}{\sqrt{n}}$.

Since (W_i) is a martingale difference, we may apply Theorem 2.16 (Section 2.4) in the Hilbert space $\mathcal{S} \times \mathcal{S}$. To this aim we have to show that (2.59), (2.60), and (2.61) hold in our context.

First we have

$$\begin{aligned} P\left(\max_{1\le i\le n} \| W_i \|_{\mathcal{S}\times\mathcal{S}} > \eta\sqrt{n}\right) \le \;& \sum_{i=1}^{n} P\left(\| W_i \|_{\mathcal{S}\times\mathcal{S}} > \eta\sqrt{n}\right) \\ & \le \frac{1}{n\eta^2}\sum_{i=1}^{n} E\left(\| W_i \|^2_{\mathcal{S}\times\mathcal{S}}\ \mathbf{I}_{\|W_i\|_{\mathcal{S}\times\mathcal{S}}>\eta\sqrt{n}}\right) \\ & \le \frac{1}{\eta^2} E\left(\| W_1 \|^2_{\mathcal{S}\times\mathcal{S}}\ \mathbf{I}_{\|W_1\|_{\mathcal{S}\times\mathcal{S}}>\eta\sqrt{n}}\right). \end{aligned}$$

Then monotone continuity of expectation yields

$$\lim_{n\uparrow\infty} P\left(\max_{1\le i\le n} \| W_i \|_{\mathcal{S}\times\mathcal{S}} > \eta\sqrt{n}\right) = 0.$$

On the other hand,

$$E\left(\max_{1\leq i\leq n} \| \frac{W_i}{\sqrt{n}} \|^2_{\mathcal{S}\times\mathcal{S}}\right) \leq E \| W_1 \|^2_{\mathcal{S}\times\mathcal{S}} < \infty;$$

thus $\left(\max_{1\leq i\leq n} \| \frac{W_i}{\sqrt{n}} \|_{\mathcal{S}\times\mathcal{S}}\ ,\ n \geq 1\right)$ is uniformly integrable and condition (2.59) is therefore satisfied.

Now let $(t_\ell,\ \ell \geq 1)$ be an orthonormal basis of $\mathcal{S} \times \mathcal{S}$ built with eigenvectors of the covariance operator Γ_{W_1} of W_1. Setting

$$r^2_N(u) = \sum_{\ell=N}^{\infty} < u,\ t_\ell >^2_{\mathcal{S}\times\mathcal{S}}\ ,\quad u \in \mathcal{S} \times \mathcal{S},$$

we get

$$P\left(\sum_{i=1}^{n} r^2_N\left(\frac{W_i}{\sqrt{n}}\right) > \eta\right) \leq \frac{n}{\eta} E\left(r^2_N\left(\frac{W_1}{\sqrt{n}}\right)\right),$$

but

$$E\left(r^2_N\left(\frac{W_1}{\sqrt{n}}\right)\right) = \frac{1}{n}\sum_{\ell=N}^{\infty}\gamma_\ell,$$

where γ_ℓ stands for the eigenvalue associated with t_ℓ.
Hence

$$\lim_{N\to\infty} \overline{\lim}_{n\to\infty} P\left(\sum_{i=1}^{n} r^2_N\left(\frac{W_i}{\sqrt{n}}\right) > \eta\right) \leq \lim_{N\to\infty} \frac{1}{\eta}\sum_{\ell=N}^{\infty}\gamma_\ell = 0$$

and (2.61) holds.

Concerning (2.60), it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} W_i \otimes W_i \underset{a.s.}{\longrightarrow} \Gamma_{W_1}, \tag{4.82}$$

where convergence takes place in the space of Hilbert-Schmidt operators over $\mathcal{S} \times \mathcal{S}$. So (4.82) is nothing but a strong law of large numbers for $(W_i \otimes W_i,\ i \geq 1)$, where (W_i) is a (special) martingale difference in a Hilbert space. Then some tedious calculations show that conditions in Corollary 2.3 are satisfied for $(W_i \otimes W_i)$. This entails (4.82).

Finally we are in a position to apply Theorem 2.16 and we may claim that, in $\mathcal{S} \times \mathcal{S}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i \xrightarrow{\mathcal{D}} N' \sim \mathcal{N}\left(0, \Gamma_{W_1}\right).$$

For the second step we use Lemma 4.6 and the canonical decomposition

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_i = (I - R')^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i + \Delta'_n\ ,$$

with

$$\Delta_n' = (I - R')^{-1} \frac{R'^n(W_1) + \ldots + R'(W_n)}{\sqrt{n}} + \frac{(R' + \ldots + R'^n)(T_0)}{\sqrt{n}}.$$

Then, as in Theorem 3.10, it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_i \xrightarrow{\mathcal{D}} N' \sim \Gamma(0, \Gamma), \tag{4.83}$$

with $\Gamma = (I - R')^{-1} \Gamma_{W_1} (I - R'^*)^{-1}$.

For the final step we utilize equalities

$$\begin{aligned} \sqrt{n}(D_n - D) &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^{n-1} (< X_i, \cdot > X_{i+1} - D) \\ &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^{n-1} [< X_i, \cdot > (\rho(X_i) + \varepsilon_{i+1}) - D] , \end{aligned}$$

which imply

$$\sqrt{n}(D_n - D) = \rho \sqrt{n}(C_{n-1} - C) + \frac{\sqrt{n}}{n-1} \sum_{i=1}^{n-1} F_{i+1}. \tag{4.84}$$

Now notice that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_i = \left(\sqrt{n}(C_n - C) \, , \, \frac{F_1 + \ldots + F_n}{\sqrt{n}} \right). \tag{4.85}$$

Then, using (4.82) and (4.84), we arrive at

$$\left(\sqrt{n}(C_{n-1} - C) \, , \, \frac{\sqrt{n}}{n-1} \sum_{i=1}^{n-1} F_{i+1} \right) \xrightarrow{\mathcal{D}} N', \tag{4.86}$$

and since weak convergence is preserved by continuity we may use (4.83) to obtain

$$(\sqrt{n}(C_{n-1} - C) \, , \, \sqrt{n}(D_n - D)) \xrightarrow{\mathcal{D}} N;$$

hence (4.81) holds. ■

As by-products we have asymptotic normality for (C_n) and (D_n). For example, it is easy to establish the following corollary.

COROLLARY 4.6 *In $\mathcal{S}$ we have*

$$\sqrt{n}(C_n - C) \xrightarrow{\mathcal{D}} N_1 \sim \mathcal{N}(0, \Gamma_1), \tag{4.87}$$

where Γ_1 *satisfies*

$$\begin{aligned} <\Gamma_1(s),\, s'>_{\mathcal{S}} = \;& Cov(<s,\, C_{(1)}>_{\mathcal{S}}, <s',\, C_{(1)}>_{\mathcal{S}}) \\ & + \sum_{k=2}^{\infty} Cov(<s,\, C_{(1)}>_{\mathcal{S}}, <s',\, C_{(k)}>_{\mathcal{S}}) \\ & + \sum_{k=2}^{\infty} Cov(<s',\, C_{(1)}>_{\mathcal{S}}, <s,\, C_{(k)}>_{\mathcal{S}}) \,; \end{aligned} \tag{4.88}$$

$s,\, s' \in \mathcal{S}$; *with* $C_{(k)} = X_k \otimes X_k,\; k \geq 1$.

If (X_i) is strongly mixing and satisfies the conditions in Theorem 2.17, we have the rate

$$\delta\left(P_{\Sigma_n}, P_N\right) = O(n^{-\eta})$$

for some $\eta > 0$, where δ is Prokhorov's metric.

We now derive asymptotic distributions for the empirical eigenvalues λ_{jn} and $\widehat{\lambda}_{jn}$.

THEOREM 4.10 *If* A_1 *holds and* $\mathcal{V}_j$ *is one dimensional, then*

$$\sqrt{n}(\lambda_{jn} - \lambda_j) \xrightarrow{\mathcal{D}} N_{(j)} = <N(v_j), v_j> . \tag{4.89}$$

$N_{(j)}$ *is a zero-mean Gaussian random variable with variance* σ_j^2 *given by*

$$\sigma_j^2 = V <X_1, v_j>^4 + 2\sum_{k=2}^{\infty} Cov(<X_1, v_j>^2, <X_k, v_j>^2). \tag{4.90}$$

Proof

Let us set

$$L_{jn} = <\sqrt{n}(C_n - C)(v_{jn}),\; v'_{j,n}> .$$

Using relations $C_n(v_{jn}) = \lambda_{jn} v_{jn}$ and $C(v'_{jn}) = \lambda_j v'_{jn}$, we obtain

$$L_{jn} = \sqrt{n}(\lambda_{jn} - \lambda_j) <v_{jn}, v'_{jn}> ,$$

and since

$$\begin{aligned} |1 - <v_{jn}, v'_{jn}>| &= |<v_{jn}, v_{jn} - v'_{jn}>| \\ &\leq \| v_{jn} - v'_{jn} \| , \end{aligned}$$

Theorem 4.5 entails

$$<v_{jn}, v'_{jn}> \underset{a.s.}{\longrightarrow} 1 .$$

Therefore L_{jn} and $\sqrt{n}(\lambda_{jn} - \lambda_j)$ have the same asymptotic behavior in distribution.

Now we have

$$| < \sqrt{n}(C_n - C)(v_j), v_j > - < \sqrt{n}(C_n - C)(v_{jn}), sgn < v_{jn}, v_j > v_j > |$$

$$= | < \sqrt{n}(C_n - C)(v_j), v_j - sgn < v_{jn}, v_j > v_{jn} > |$$

$$= | < \sqrt{n}(C_n - C)(v_j) \ , \ \ v'_{jn} - v_{jn} > |$$

$$\leq \ \| \sqrt{n}(C_n - C) \|_{\mathcal{S}} \| v'_{jn} - v_{jn} \| \ .$$

From Corollary 4.6 it follows that $\| \sqrt{n}(C_n - C) \|_{\mathcal{S}}$ is bounded in probability and, since $\| v_{jn} - v'_{jn} \| \underset{a.s.}{\longrightarrow} 0$, we conclude that L_{jn} and $< \sqrt{n}(C_n - C)(v_j), v_j >$ have the same asymptotic behaviour in distribution.

Applying Corollary 4.6 again we find

$$\sqrt{n}(\lambda_{jn} - \lambda_j) \xrightarrow{\mathcal{D}} < N_1(v_j), v_j > .$$

Now note that

$$< N_1(v_j), v_j > = < N_1, \ v_j \otimes v_j >_{\mathcal{S}},$$

so $< N_1(v_j), v_j >$ is actually a zero-mean Gaussian random variable with variance

$$E\left(< N_1, v_j \otimes v_j >_{\mathcal{S}}^2\right) = < \Gamma_{N_1}(v_j \otimes v_j), \ (v_j \otimes v_j) >_{\mathcal{S}};$$

hence (4.89) is proved by using (4.87). ■

The same result holds true for $\widehat{\lambda}_{jn}$:

COROLLARY 4.7 *If A_1 holds, then*

$$\sqrt{n}\left(\widehat{\lambda}_{jn} - \lambda_j\right) \xrightarrow{\mathcal{D}} N_j \ . \tag{4.91}$$

Proof

It is enough to remark that

$$\sqrt{n}\left(\widehat{\lambda}_{jn} - \lambda_j\right) = < \sqrt{n}(C_n - C)(v_j), \ v_j >,$$

which converges in distribution to N_j, as shown in the proof of Theorem 4.10. ■

Recall that, although $\widehat{\lambda}_{jn}$ is defined in a special case (v_j known!), it has the same asymptotic variance as λ_{jn}. However, Corollary 4.7 is valid whatever is the dimension of $\mathcal{V}_j$.

We finally indicate without proof a limit in distribution for v_{jn}.

COROLLARY 4.8 *Under the conditions in Theorem 4.10,*

$$\sqrt{n}(v_{jn} - v'_{jn}) \xrightarrow{\mathcal{D}} N'_j \sim \mathcal{N}(0, \Gamma'_j), \tag{4.92}$$

with

$$\Gamma'_j = \left(\sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} v_k \otimes v_k \right) [N_1(v_j)],$$

where N_1 appears in Corollary 4.6.

NOTES

4.1 Lemma 4.1 is taken from Bosq (1999-a). Theorems and corollaries in Section 4.1 are consequences of this lemma.

4.2 The first part of Lemma 4.2 is due to Allahverdiev (1957), whereas the second part appears in Gohberg and Krein (1971). Lemma 4.3 comes from Bosq (1991-a). For results in the i.i.d. case we refer to Dauxois, Pousse and Romain (1982) and for the general stationary case to Bosq (1989).

4.3 Lemma 4.5 is in Bosq (1999-a). Most of the results in this section appear to be new.

4.4 The useful Lemma 4.6, which is a generalization of Lemma 4.1, is based on an idea by Mas (1999-a). Limits in distribution of empirical eigenvalues are new (at least in this general context). Corollary 4.8 is due to Mas (1999-b).

5

Autoregressive Hilbertian Processes of Order p

The Markovian character of the ARH(1) model induces some limits to its efficiency for applications to statistics in continuous time. In this chapter we introduce the more flexible autoregressive model of order p.

From a pure mathematical point of view $ARH(p)$ processes may be interpreted as $ARH^p(1)$ processes. This simple remark makes their study easier in particularly in an asymptotic context.

However, we shall see that some problems are specifically associated with this model for example, properties of partial autocovariance or estimation of p.

Except where otherwise stated, notation is the same as in Chapters 3 and 4.

5.1 The $ARH(p)$ model

DEFINITION 5.1 *Let H be a separable Hilbert space. A sequence $X = (X_n,\ n \in \mathbb{Z})$ of H-random variables is said to be an* **autoregressive Hilbertian process of order p** *($ARH(p)$) associated with $(\mu, \varepsilon, \rho_1, \ldots, \rho_p)$ if it is stationary and such that*

$$X_n - \mu = \rho_1(X_{n-1} - \mu) + \ldots + \rho_p(X_{n-p} - \mu) + \varepsilon_n,\ n \in \mathbb{Z}, \qquad (5.1)$$

where $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is an H-white noise, $\mu \in H$, and $\rho_1, \ldots, \rho_p \in \mathcal{L}$, with $\rho_p \neq 0$.

Note that, if $p = 1$, Définitions 3.3 and 5.1 differ slightly since $\rho = 0$ is not prohibited in Definition 3.3. Condition $\rho_p \neq 0$ is of course necessary for identifiability of p. Actually, this condition is not sufficient. A sufficient condition would be $P\{\rho_p(X_n) \neq 0\} > 0,\ n \in \mathbb{Z}$.

Markovian representation of an $ARH(p)$

Let H^p be the cartesian product of p copies of H. If H^p is equipped with the scalar product

$$< (x_1, \dots, x_p),\ (y_1, \dots, y_p) >_p := \sum_{j=1}^{p} < x_j, y_j > ;\ x_1, \dots, x_p,\ y_1, \dots, y_p \in H, \tag{5.2}$$

it becomes a separable Hilbert space.

The norm in H^p will be denoted by $\|\cdot\|_p$, the space of bounded linear operators over H^p by $\mathcal{L}_p$, the space of Hilbert-Schmidt (resp. nuclear) operators over H^p by $\mathcal{S}_p$ (resp. $\mathcal{N}_p$), and the corresponding norms and scalar products by $\|\cdot\|_{\mathcal{L}_p}$, $\|\cdot\|_{\mathcal{S}_p}$, $<\cdot,\cdot>_{\mathcal{S}_p}$, and $\|\cdot\|_{\mathcal{N}_p}$.

Now let us set $Y = (Y_n,\ n \in \mathbb{Z})$, where

$Y_n = (X_n, \dots, X_{n-p+1})\ ,\ \ n \in \mathbb{Z}\ ;$

$\mu' = (\mu, \dots, \mu) \in H^p\ ;$

$\varepsilon' = (\varepsilon'_n,\ n \in \mathbb{Z})$ with $\varepsilon'_n = (\varepsilon_n, 0, \dots, 0)$, $n \in \mathbb{Z}$, where 0 appears $p-1$ times; and consider the operator on H^p defined as

$$\rho' = \begin{bmatrix} \rho_1 & \rho_2 & \dots & \rho_p \\ I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ 0 & \dots & I & 0 \end{bmatrix}$$

where I, as usual, denotes the identity operator.

We have the following simple but crucial lemma.

LEMMA 5.1 *If X is an $ARH(p)$ associated with $(\mu, \varepsilon, \rho_1, \dots, \rho_p)$, then Y is an $ARH^p(1)$ associated with $(\mu', \varepsilon', \rho')$.*

Proof

Clearly ε' is a H^p-white noise and $\rho' \in \mathcal{L}_p$. Now let $x = (x_0, \dots, x_{p-1})$ and $y = (y_0, \dots, y_{p-1})$ be two elements of H^p. Then

$$E(< Y_n, x >_p < Y_m, y >_p) = \sum_{j=0}^{p-1} E(< X_{n-j}, x_j >< X_{m-j}, y_j >);$$

thus

$$
\begin{aligned}
< C_{Y_n,Y_m}(x), y >_p &= \sum_{j=0}^{p-1} < C_{X_{n-j},X_{m-j}}(x_j), y_j > \\
&= \sum_{j=0}^{p-1} < C_{X_0,X_{m-n}}(x_j),\ y_j >,
\end{aligned}
$$

which shows that C_{Y_n,Y_m} depends only on $n-m$. This property, together with $EY_n = (EX_n, \dots, EX_{n-p+1})$, implies weak stationarity of Y.

Now we have

$$
Y_n - \mu' = \rho'(Y_{n-1} - \mu') + \varepsilon'_n \ , \quad n \in \mathbb{Z}, \tag{5.3}
$$

and the proof is complete. ■

We may now give a condition for existence and uniqueness of X. We will use the natural "projector" of H^p onto H defined as

$$
\pi(x_1, \dots, x_p) = x_1 \ , \quad (x_1, \dots, x_p) \in H^p.
$$

THEOREM 5.1 *Under the assumption*

$$
(c_0') \qquad \qquad \| \rho'^{j_0} \|_{\mathcal{L}_p} < 1 \ \textit{ for some } j_0 \geq 1,
$$

equation (5.1) has a unique stationary solution given by

$$
X_n = \mu + \sum_{j=0}^{\infty} (\pi \rho'^j)(\varepsilon'_{n-j}), \ n \in \mathbb{Z} \ , \tag{5.4}
$$

where the series converges in $L^2_H(\Omega, \mathcal{A}, P)$ and with probability 1. Moreover, ε is the innovation process of $(X_n - \mu)$.

Proof

We may and do suppose that $\mu = 0$. Now from Theorem 3.1 it follows immediately that (5.3) has a unique stationary solution given by

$$
Y_n = \sum_{j=0}^{\infty} \rho'^j (\varepsilon'_{n-j}) \ , \quad n \in \mathbb{Z}, \tag{5.5}
$$

where the series converges in $L^2_{H^p}(\Omega, \mathcal{A}, P)$ and with probability 1. Moreover, (ε'_n) is the innovation process of (Y_n).

Noting that

$$
X_n = \pi Y_n \ , \quad n \in \mathbb{Z}, \tag{5.6}
$$

we see that (X_n) is a stationary H-valued process that satisfies (5.4) and (5.1).

Let us now prove that (ε_n) is the innovation process of (X_n). From Theorems 1.8 and 1.9 (Section 1.6) it suffices to show that

$$E < \varepsilon_n,\ \ell(X_i) >= 0\ ,\ \ i \leq n-1,\ \ell \in \mathcal{L}. \tag{5.7}$$

If $\ell \in \mathcal{L}$, it induces an operator $\ell' \in \mathcal{L}_p$ defined by

$$\ell'(x_1, \ldots, x_p) = (\ell(x_1), 0, \ldots, 0),\ (x_1, \ldots, x_p) \in H^p,$$

and since ε' is the innovation of Y we have

$$E < \varepsilon_n',\ \ell'(Y_i) >_p = 0\ ,\ \ i \leq n-1,$$

which entails (5.7).

Concerning uniqueness, note that, if $(X_{n,1})$ is another stationary solution of (5.1), Lemma 5.1 shows that $Y_{n,1} = (X_{n,1}, \ldots, X_{n-p+1,1})$, $n \in \mathbb{Z}$, is a stationary solution of (5.3). Thus $Y_{n,1} = Y_n$ (a.s.) and it follows that $X_{n,1} = X_n$ (a.s.) for all n in $\mathbb{Z}$.

Condition (c_0') is natural but not necessary for obtaining (5.4) (recall Example 3.5).

Since (c_0') is somewhat difficult to verify, we now introduce a condition that is directly associated with the operators $\rho_1, \ldots, \rho_p$. For this purpose we put

$$Q(z) = z^p I - z^{p-1}\rho_1 - \ldots - z\rho_{p-1} - \rho_p,\ z \in \mathbb{C}.$$

For every z, $Q(z)$ is a bounded linear operator over the complex extension H' of H.

We then have

THEOREM 5.2 *Suppose that the following condition holds:*

$$Q(z) \text{ not invertible} \Rightarrow |z| < 1. \tag{5.8}$$

Then (c_0') holds. Therefore (5.1) has a unique stationary solution given by (5.4).

Proof

Let us consider the operators on H'^p defined as

$$N(z) = \begin{bmatrix} I & zI & z^2 I & \ldots & z^{p-1} I \\ 0 & I & zI & \ldots & z^{p-2} I \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & I & zI \\ 0 & \ldots & \ldots & 0 & I \end{bmatrix},$$

and

$$M(z) = \begin{bmatrix} 0 & -I & 0 & \dots & 0 \\ 0 & 0 & -I & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & \dots & 0 & -I \\ Q_0(z) & Q_1(z) & \dots & \dots & Q_{p-1}(z) \end{bmatrix}$$

where

$Q_0(z) = I$
and

$Q_j(z) = zQ_{j-1}(z) - \rho_j \ ; \ j = 1, \dots, p.$
It is easy to see that

$$M(z)(zI_p - \rho')N(z) = \begin{bmatrix} I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & \vdots & \vdots \\ \vdots & & & I & 0 \\ 0 & \dots & \dots & 0 & Q(z) \end{bmatrix}, \tag{5.9}$$

where I_p denotes the identity of H^p.

Now, due to their special form, $N(z)$ and $M(z)$ are invertible for all z. Then from (5.9) it follows that

$$E := \{z, zI - \rho' \text{ is not invertible}\} \subset \{z, Q(z) \text{ is not invertible}\}$$

and using (5.8) we get

$$E \subset \{z,\ |z| < 1\}. \tag{5.10}$$

E is the so-called spectrum of ρ' (over H'). It is a closed set and

$$r_{\rho'} := \sup_{z \in E} |z| = \lim_{j \to \infty} \| \rho'^j \|_{\mathcal{L}_p}^{1/j} \tag{5.11}$$

(see Dunford and Schwartz (1958)).

From (5.10) and (5.11) we then may deduce that there exist an integer j_1, $\alpha \in]0, 1[$ and $k > 0$ such that

$$\| \rho'^j \|_{\mathcal{L}_p} \leq k\alpha^j\ ,\ \ j \geq j_1.$$

Thus (c_0') holds and the proof is complete. ■

Note that it is possible to show that $r_{\rho'} \leq \| \rho' \|_{\mathcal{L}_p}$, but $r_{\rho'} < 1$ does not entail $\| \rho' \|_{\mathcal{L}_p} < 1$. On the other hand, if H is finite dimensional, (5.8) is equivalent to "determinant of $Q(z) = 0$ implies $|z| < 1$."

COROLLARY 5.1 *If* $\sum_{j=1}^{p} \| \rho_j \|_{\mathcal{L}} < 1$*, then (5.8) holds and* $Q(z) = 0$ *implies* $|z| < 1$.

Proof

Let $z \in \mathbb{C}$ such that $|z| \geq 1$. Then

$$Q(z) = z^p \left(I - \frac{\rho_1}{z} - \ldots - \frac{\rho_p}{z_p} \right),$$

but

$$\| \frac{\rho_1}{z} + \ldots + \frac{\rho_p}{z^p} \|_{\mathcal{L}} \leq \| \rho_1 \|_{\mathcal{L}} + \ldots + \| \rho_p \|_{\mathcal{L}} < 1.$$

Thus $I - \frac{\rho_1}{z} - \ldots - \frac{\rho_p}{z^p}$ is invertible, so $Q(z)$ is also invertible and (5.8) holds.

Suppose now that there exists $z_0 \in C$ such that $Q(z_0) = 0$ when $|z_0| \geq 1$. We then have

$$z_0^{-p} \rho_p = I - z_0^{-1} \rho_1 - \ldots - z_0^{-p+1} \rho_{p-1};$$

hence

$$\begin{aligned} \| \rho_p \|_{\mathcal{L}} &\geq \| I - z_0^{-1} \rho_1 - \ldots - z_0^{-p+1} \rho_{p-1} \|_{\mathcal{L}} \\ &\geq 1 - \| \rho_1 \|_{\mathcal{L}} - \ldots - \| \rho_{p-1} \|_{\mathcal{L}}, \end{aligned}$$

which implies $\sum_{j=1}^{p} \| \rho_j \|_{\mathcal{L}} \geq 1$, contrary to the assumption. ■

Let us now give an example of $ARH(p)$.

Example 5.1

Take H and ε as in Example 3.2 and let $\rho_j = \ell_{K_j}$; $j = 1, \ldots, p$ be kernel operators (cf. Example 1.18 in Section 1.5) associated with kernels $K_1, \ldots, K_p$ such that

$$\sum_{j=1}^{p} \left(\int_{[0,1]^2} K_j^2(s,t) ds dt \right)^{1/2} < 1.$$

Then Corollary 5.1 shows that (5.1) has a unique stationary solution (X_n), which satisfies

$$X_n(t) = \int_0^1 \left[\sum_{j=1}^{p} K_j(s,t) X_{n-j}(s) \right] ds + \int_n^{n+t} \varphi(n+t-s) dW(s), \tag{5.12}$$

$0 \leq t \leq 1$, $n \in \mathbb{Z}$.

We finally indicate a result concerning projection of an $ARH(p)$ process.

THEOREM 5.3 *Let* (X_n) *be an* $ARH(p)$ *zero-mean process associated with* $(\rho_1, \ldots, \rho_p,\ \varepsilon)$. *Suppose that there exist* $v \in H$ *and* $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ $(\alpha_p \neq 0)$ *such that*

$$\rho_j^*(v) = \alpha_j v; \ \ j = 1, \ldots, p;$$

and

$$E(< \varepsilon_0, v >^2 > 0.$$

Then $(< X_n, v >,\ n \in \mathbb{Z})$ *is an* $AR(p)$ *process that satisfies*

$$< X_n, v > = \sum_{j=1}^{p} \alpha_j < X_{n-j}, v > + < \varepsilon_n, v >, \ \ n \in \mathbb{Z}. \tag{5.13}$$

Proof: clear. ■

5.2 Second order moments of $ARH(p)$

In this section we will suppose that X is a **standard** $ARH(p)$, that is, $EX_0 = 0$ and (c_0') is satisfied. Note that if X is standard then Y is a standard $ARH^p(1)$.

Autocovariance

The autocovariance of an $ARH(p)$ process is the sequence $(C_h,\ h \in \mathbb{Z})$ of operators defined by

$$C_h = C_{X_0, X_h}, \quad h \in \mathbb{Z}. \tag{5.14}$$

We then have $C_{-h} = C_h^*$, $h \in \mathbb{Z}$, and (C_h) satisfies the so-called **Yule-Walker equations**:

THEOREM 5.4 *If* X *is a standard* $ARH(p)$, *then*

$$C_h = \sum_{j=1}^{p} \rho_j C_{h-j} \ ; \ \ h = 1, 2, \ldots; \tag{5.15}$$

and

$$C_0 = \sum_{j=1}^{p} \rho_j C_j + C_\varepsilon, \tag{5.16}$$

where C_ε *is the covariance operator of* ε_0.

Proof

From (5.1) it follows that for every natural integer h

$$\begin{aligned} E(< X_h, x >< X_0, y >) &= \sum_{j=1}^{p} E(< \rho_j(X_{h-j}), x >< X_0, y >) \\ &\quad + E(< \varepsilon_h, x >< X_0, y >) \; ; \; x, y \in H. \end{aligned}$$

Since ε is the innovation of X, we have

$$E(< \varepsilon_h, x >< X_0, y >) = \delta_{0h} E(< \varepsilon_h, x >< \varepsilon_0, y >).$$

Therefore

$$C_{-h} = \sum_{j=1}^{p} C_{-h+j} \, \rho_j^* \; , \quad h \geq 1,$$

which is equivalent to (5.15).

If $h = 0$, one obtains

$$C_0 = \sum_{j=1}^{p} C_{-j} \rho_j^* + C_\varepsilon,$$

which is equivalent to (5.16). ■

Partial autocovariance

Given a stationary process (X_n), the partial cross-covariance operator of order k is the cross-covariance operator between X_0 and X_k after elimination of the (linear) influence of $X_1, \ldots, X_{k-1}$ over X_0 and X_k. The idea is clarified by the following definition.

DEFINITION 5.2 *Let $X = (X_n, \ n \in \mathbb{Z})$ be a (weakly) stationary zero-mean H-valued process. The* **partial autocovariance** *is the sequence $(P_k, \ k \geq 2)$ of operators given by*

$$P_k = C_{X_n - \pi^{\mathcal{M}^{k-1}_{n+k-1}}(X_n), \; X_{n+k} - \pi^{\mathcal{M}^{k-1}_{n+k-1}}(X_{n+k})} \, , \; k \geq 2 \, , \tag{5.17}$$

where $\pi^{\mathcal{M}^{k-1}_{n+k-1}}$ denotes the orthogonal projector of the $\mathcal{L}$-closed subspace of $L^2_H(\Omega, \mathcal{A}, P)$ generated by $X_{n+1}, \ldots, X_{n+k-1}$.

Stationarity of X entails that P_k does not depend on n. Note that from Theorem 1.8 the above LCS is the closure of the linear space

$\left\{ \sum_{j=1}^{k-1} \ell_j(X_{n+j}), \; \ell_j \in \mathcal{L}; \; 1 \leq j \leq k-1 \right\}$ but it is **not** finite dimensional in general.

Concerning partial autocovariance of an $ARH(p)$ we have the following extension of a well-known one-dimensional property. In the statement we write π_i^j, for $\pi^{\mathcal{M}_i^j}$.

THEOREM 5.5 *Let X be a standard $ARH(1)$. If π_{p-1}^{p-1} and ρ_p commute, then*

$$P_p = C_{X_0 - \pi_{p-1}^{p-1}(X_0)} \rho_p^*. \tag{5.18}$$

If $k > p$, then

$$P_k = 0. \tag{5.19}$$

Proof

Equation (5.1) gives

$$X_p = \sum_{j=1}^{p} \rho_j(X_{p-j}) + \varepsilon_p. \tag{5.20}$$

Since ε is innovation, this implies

$$\pi_{p-1}^{p-1}(X_p) = \sum_{j=1}^{p-1} \rho_j(X_{p-j}) + \pi_{p-1}^{p-1}\rho_p(X_0). \tag{5.21}$$

From (5.20) and (5.21) it follows that

$$X_p - \pi_{p-1}^{p-1}(X_p) = \rho_p(X_0) - \pi_{p-1}^{p-1}\rho_p(X_0) + \varepsilon_p$$

and, since π_{p-1}^{p-1} and ρ_p commute,

$$X_p - \pi_{p-1}^{p-1}(X_p) = \rho_p\left(I - \pi_{p-1}^{p-1}\right)(X_0) + \varepsilon_p.$$

Therefore

$$E\left(< X_0 - \pi_{p-1}^{p-1}(X_0), x >< X_p - \pi_{p-1}^{p-1}(X_p),\ y >\right)$$

$$= E < X_0 - \pi_{p-1}^{p-1}(X_0), x >< X_0 - \pi_{p-1}^{p-1}(X_0),\ \rho_p^*(y) >,$$

$x, y \in H$, which is equivalent to (5.18).

On the other hand, if $k > p$ we have

$$\pi_{p-1}^{k-1}(X_p) = \sum_{j=1}^{p} \rho_j(X_{p-j}). \tag{5.22}$$

Then (5.20) and (5.22) yield

$$X_p - \pi_{p-1}^{k-1}(X_p) = \varepsilon_p;$$

hence

$$C_{X_{p-k} - \pi_{p-1}^{k-1}(X_{p-k}),\ X_p - \pi_{p-1}^{k-1}(X_p)} = 0,$$

which is (5.19). ■

COROLLARY 5.2 *If X is a Gaussian standard $ARH(p)$ process, then*

$$\pi_{n+k-1}^{k-1} = E^{\sigma(X_{n+1},\ldots,X_{n+k-1})} \tag{5.23}$$

and (5.18) (5.19) hold.

Proof

From Theorem 1.11 it is easy to realize that (5.23) is valid. Now conditional expectation and ρ_p commute (see Section 1.3); thus (5.18) holds as well as (5.19). ■

5.3 Limit theorems for $ARH(p)$ processes

The representation Lemma 5.1 allows us to use the limit theorems for $ARH(1)$ in order to derive the corresponding results for $ARH(p)$.

We begin with a covariance inequality analogous to (3.24).

LEMMA 5.2 *If X is a standard $ARH(p)$ then*

$$|E < X_0, X_h > | \ \leq \| \rho'^h \|_{\mathcal{L}_p}^2 \ E \| X_0 \|^2, \ \ h \geq 1. \tag{5.24}$$

Proof

$Y_n = (X_n, \ldots, X_{n-p+1})$, $n \in \mathbb{Z}$ is a standard $ARH^p(1)$. Therefore it satisfies (3.24) under the form

$$|E < Y_0, Y_h >_p | \leq \| \rho'^h \|_{\mathcal{L}_p} \ E \| Y_0 \|_p^2, \ \ h \geq 1. \tag{5.25}$$

Now we have

$$E \| Y_0 \|_p^2 = p \, E \| X_0 \|^2$$

and

$$\begin{aligned} E < Y_0, Y_h >_p &= \sum_{j=0}^{p-1} E < X_j, X_{j+h} > \\ &= p \, E < X_0, X_h >; \end{aligned}$$

thus (5.25) yields (5.24). ■

We may now state laws of large numbers.

THEOREM 5.6 *If X is a standard $ARH(p)$ then, as $n \to \infty$;*

$$\frac{n^{1/4}}{(\text{Log } n)^\beta} \frac{S_n}{n} \underset{a.s.}{\longrightarrow} 0 \, , \ \ \beta > \frac{1}{2}. \tag{5.26}$$

Proof

Similar to the proof of Theorem 3.7. ■

An exact rate of convergence in $L^2_H(P)$ is given by the following proposition.

THEOREM 5.7 *Let X be a standard $ARH(p)$. Then*

$$n\, E \parallel \frac{S_n}{n} \parallel^2 \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < X_0, X_h > . \tag{5.27}$$

Proof

From Theorem 3.8 it follows that

$$n\, E \parallel \frac{Y_1 + \ldots + Y_n}{n} \parallel_p^2 \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < Y_0, Y_h >_p < \infty.$$

Then, using (5.2) and stationarity of X, we obtain

$$n\, p\, E \parallel \frac{S_n}{n} \parallel^2 \underset{n \to \infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} p\, E < X_0, X_h > .$$

■

Concerning large deviations we can extend Theorem 3.9.

THEOREM 5.8 *Let X be a standard $ARH(p)$. Then if $\varepsilon_0 \in \mathcal{E}$ (or equivalently $X_0 \in \mathcal{E}$), there exist $\alpha_0' > 0$ and $\beta_0' > 0$, which only depend on $\rho_1, \ldots, \rho_p$ and P_{ε_0}, such that*

$$P\left(\parallel \frac{S_n}{n} \parallel \geq \eta \right) \leq 4 \exp\left(-\frac{n\eta^2}{\alpha_0' + \beta_0' \eta} \right), \quad \eta > 0. \tag{5.28}$$

Proof

Note first that, since $\exp(\gamma \parallel \varepsilon_0' \parallel_p) = \exp(\gamma \parallel \varepsilon_0 \parallel)$, $\gamma > 0$, Lemma 3.3 implies that ε_0' and Y_0 have some exponential moments.

Thus, applying Theorem 3.9 we obtain for every $\eta > 0$ and a suitable choice of α_0' and β_0'

$$P\left(\parallel \frac{Y_1 + \ldots + Y_n}{n} \parallel_p \geq \eta \right) \leq 4 \exp\left(-\frac{n\eta^2}{\alpha_0' + \beta_0' \eta} \right). \tag{5.29}$$

Now noting that $\dfrac{S_n}{n} = \pi \left(\dfrac{Y_1 + \ldots + Y_n}{n} \right)$ and that $\parallel \pi \parallel_{\mathcal{L}(H^p, H)} = 1$, we get

$$P\left(\parallel \frac{S_n}{n} \parallel \geq \eta \right) \leq P\left(\parallel \frac{Y_1 + \ldots + Y_n}{n} \parallel_p \geq \eta \right)$$

and (5.28) follows from (5.29). ■

COROLLARY 5.3

$$\| \frac{S_n}{n} \| = O\left(\left(\frac{\text{Log}n}{n}\right)^{1/2}\right) \quad a.s. \tag{5.30}$$

Proof: clear. ■

We now study asymptotic normality of $\frac{S_n}{n}$. To this aim we need a technical lemma.

LEMMA 5.3 *$(I_p - \rho')$ is invertible in H^p if and only if $\left(I - \sum_{j=1}^{p} \rho_j\right)$ is invertible in H and the first block in $(I_p - \rho')^{-1}$ is given by*

$$(I_p - \rho')^{-1}_{11} = \left(I - \sum_{j=1}^{p} \rho_j\right)^{-1}. \tag{5.31}$$

Proof

Let us consider the case $p = 2$. We have

$$I_2 - \rho' = \begin{bmatrix} I - \rho_1 & -\rho_2 \\ -I & I \end{bmatrix}$$

and invertibility of $I_2 - \rho'$ means existence of α, β, γ, δ in $\mathcal{L}$ such that

$$(I_2 - \rho')\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}(I_2 - \rho') = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{5.32}$$

Now easy calculations show that

$$\alpha(I - \rho_1 - \rho_2) = (I - \rho_1 - \rho_2)\alpha = I$$

thus $I - \rho_1 - \rho_2$ is invertible. Actually, it can be proved that

$$(I_2 - \rho')^{-1} = \begin{bmatrix} (I - \rho_1 - \rho_2)^{-1} & -(I - \rho_1 - \rho_2)^{-1}\rho_2 \\ (I - \rho_1 - \rho_2)^{-1} & I + (I - \rho_1 - \rho_2)^{-1}\rho_2 \end{bmatrix}.$$

Conversely, if $I - \rho_1 - \rho_2$ is invertible, one may set

$$\alpha = \gamma = (I - \rho_1 - \rho_2)^{-1},\ \beta = \rho_2(I - \rho_1 - \rho_2)^{-1},\ \delta = I + (I - \rho_1 - \rho_2)^{-1}\rho_2.$$

Then it is easy to verify that (5.32) and (5.31) hold.

If $p > 2$, a recursive argument leads to the general statement. Details are omitted. ■

From the above lemma it follows that condition $\sum_{j=1}^{p} \| \rho_j \|_{\mathcal{L}} < 1$ yields invertibility of $I_p - \rho'$. More generally, (5.8) gives existence of $(I_p - \rho')^{-1}$.

We now state the CLT.

THEOREM 5.9 *Let X be a standard $ARH(p)$ associated with a strong white noise ε and such that $I - \sum_{j=1}^{p} \rho_j$ is invertible. Then*

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0, \Gamma), \tag{5.33}$$

where

$$\Gamma = \left(I - \sum_{j=1}^{p} \rho_j \right)^{-1} C_\varepsilon \left(I - \sum_{j=1}^{p} \rho_j^* \right)^{-1}. \tag{5.34}$$

Proof

Using Lemma 5.1 and Theorem 3.10 we obtain

$$\frac{Y_1 + \ldots + Y_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N' \sim \mathcal{N}(0, \Gamma'), \tag{5.35}$$

where

$$\Gamma' = (I_p - \rho')^{-1} C_{\varepsilon'} (I_p - \rho'^*)^{-1}.$$

Noting that

$$\frac{S_n}{\sqrt{n}} = \pi \left(\frac{Y_1 + \ldots + Y_n}{\sqrt{n}} \right)$$

and using continuity of weak convergence with respect to continuous transformations (cf. Section 2.2) we get

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \pi N' \sim \mathcal{N}(0, \pi \Gamma' \pi).$$

Now we have

$$\Gamma := \pi \Gamma' \pi^* = \pi (I_p - \rho')^{-1} C_{\varepsilon'} (I_p - \rho'^*)^{-1} \pi,$$

which leads to (5.34) by using (5.31) and the definition of π. ■

5.4 Estimation of autocovariance of an $ARH(p)$

As in the previous section, the Markov representation of $ARH(p)$ plays a central role in this section.

We suppose that (X_n) is a standard $ARH(p)$ associated with a strong white noise and such that $E \parallel X_0 \parallel^4 < \infty$. From here on these assumptions will be denoted A_p. Clearly if A_p holds then (Y_n) satisfies A_1 in H^p (cf. Section 4.1).

Notation concerning (X_n) is the same as in Chapter 4. Concerning (Y_n), we note "Y" as an exponent. For example, the empirical covariance operator of (Y_n) is defined as

$$C_n^Y = \frac{1}{n}\sum_{i=1}^{n} Y_i \otimes Y_i.$$

We first deal with convergence in the $L^2_{\mathcal{S}}$ sense.

THEOREM 5.10 *If A_p holds we have*

$$\limsup_{n\to\infty} n\, E \parallel C_n - C \parallel_{\mathcal{S}}^2 \le \sum_{h=-\infty}^{+\infty} E < Y_0, Y_h >_{\mathcal{S}_p}, \tag{5.36}$$

and, for $1 \le k \le p-1$,

$$\limsup_{n\to\infty} n\, E \parallel C_{h,n} - C_h \parallel_{\mathcal{S}}^2 \le \sum_{h=-\infty}^{+\infty} E < Y_0, Y_h >_{\mathcal{S}_p}. \tag{5.37}$$

Proof

For convenience we prove Theorem 5.10 only for $p = 2$. Proof for $p > 2$ is similar but tedious.

First note that C^Y and C_n^Y may be written as "blocks" operators. Actually,

$$C^Y = \begin{bmatrix} C & D \\ D^* & C \end{bmatrix} \quad \text{and} \quad C_n^Y = \begin{bmatrix} C_n & D_n \\ D_n^* & C_n \end{bmatrix}.$$

Then it is straightforward that

$$\parallel C_n^Y - C^Y \parallel_{\mathcal{S}_2}^2 = 2 \parallel C_n - C \parallel_{\mathcal{S}}^2 + 2 \parallel D_n - D \parallel_{\mathcal{S}}^2. \tag{5.38}$$

It now suffices to apply Theorem 4.1 to Y, to obtain

$$n\, E \parallel C_n^Y - C^Y \parallel_{\mathcal{S}_2}^2 \underset{n\to\infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E < Y_0, Y_h >_{\mathcal{S}_2},$$

which implies (5.36) and (5.37). ■

Estimation of C_h for $h \ge p$ may be treated similarly as a by-product of estimation of C_h^Y, $h \geq 1$. This topic is omitted.

Concerning pointwise convergence and large deviations the results are the same as in the case $p = 1$. We may resume them in the following (elliptical !) statement.

THEOREM 5.11 *Under similar conditions as for an* $ARH(1)$, $\| C_n - C \|_{\mathcal{S}}$ *satisfies (4.18), (4.22), (4.28), (4.33) and* $\| D_n - D \|_{\mathcal{S}}$ *satisfies the properties in Theorem 4.8.*

Proof $(p = 2)$

It suffices to remark that Y is an $ARH(1)$ such that

$$\| C_n - C \|_{\mathcal{S}} \leq \| C_n^Y - C^Y \|_{\mathcal{S}_2},$$

and

$$\| D_n - D \|_{\mathcal{S}} \leq \| C_n^Y - C^Y \|_{\mathcal{S}_2} .$$

■

Estimation of the eigenelements of C^Y and C

THEOREM 5.12 *Results in Section 4.2 are valid for the eigenelements of* C^Y *and* C.

Proof

This a consequence of Theorem 5.11 and the inequalities

$$\begin{aligned} \sup_{j \geq 1} |\lambda_{jn}^Y - \lambda_j^Y| &\leq \| C_n^Y - C^Y \|_{\mathcal{S}_p}, \\ \sup_{j \geq 1} |\lambda_{jn} - \lambda_j| &\leq \| C_n - C \|_{\mathcal{S}_p}, \end{aligned}$$

and

$$\begin{aligned} \| v_{jn}^Y - v_{jn}'^Y \| &\leq a_j^Y \| C_n^Y - C^Y \|_{\mathcal{L}_p}, \\ \| v_{jn} - v_{jn}' \| &\leq a_j \| C_n - C \|_{\mathcal{L}} \end{aligned}$$

if the corresponding eigensubspace is one dimensional. ■

Weak convergence

THEOREM 5.13 *If* A_p *holds then*

$$\sqrt{n}(C_n - C,\ D_n - D) \xrightarrow{\mathcal{D}} M, \tag{5.39}$$

where M *is a zero-mean Gaussian* $(\mathcal{S} \times \mathcal{S})$*-valued random variable.*

Proof $(p = 2)$

Recall first that

$$C_n^Y - C^Y = \begin{bmatrix} C_n - C & D_n - D \\ D_n^* - D^* & C_n - C \end{bmatrix}.$$

Now from Corollary 4.6 applied to Y it follows that

$$\sqrt{n}\left(C_n^Y - C^Y\right) \xrightarrow{\mathcal{D}} N^Y. \tag{5.40}$$

Here weak convergence takes place in $\mathcal{S}_p$ and $N^Y \sim \mathcal{N}(0, \Gamma^Y)$, where

$$\begin{aligned} < \Gamma^Y(s),\ s' >_{\mathcal{S}_2} = \quad & \operatorname{Cov}\left(< s,\ C_{(1)}^Y >_{\mathcal{S}_2}, < s',\ C_{(1)}^Y >_{\mathcal{S}_2}\right) \\ & + \sum_{k=2}^{\infty} \operatorname{Cov}\left(< s,\ C_{(1)}^Y >_{\mathcal{S}_2}, < s',\ C_{(k)}^Y >_{\mathcal{S}_2}\right) \\ & + \operatorname{Cov}\left(< s',\ C_{(1)}^Y >_{\mathcal{S}_2}, < s, C_{(k)}^Y >_{\mathcal{S}_2}\right); \end{aligned} \tag{5.41}$$

$s, s' \in \mathcal{S}_2$; with $C_{(k)}^Y = Y_k \otimes Y_k,\ k \geq 1$.

We now define a linear bounded transformation $T : \mathcal{S}_2 \mapsto \mathcal{S} \times \mathcal{S}$ in the following way: given $s \in \mathcal{S}_2$ we write it in the form

$$s = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix},$$

where for every $(x, y) \in H \times H$

$$\begin{aligned} s_{11}(x) &= s(x, 0)_1, \\ s_{21}(x) &= s(0, x)_2, \\ s_{12}(y) &= s(0, y)_1, \\ s_{22}(y) &= s(0, y)_2. \end{aligned}$$

We then have

$$s \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_{11}(x) + s_{12}(y) \\ s_{21}(x) + s_{22}(y) \end{bmatrix},$$

and we set

$$T(s) = (s_{11}, s_{12}) \quad , \quad s \in \mathcal{S}_2.$$

Now, from (5.40) and continuity of weak convergence with respect to continuous transformations (Section 2.2) it follows that

$$T\left(\sqrt{n}\left(C_n^Y - C^Y\right)\right) \xrightarrow{\mathcal{D}} T(N^Y).$$

This is (5.39) with $M = T(N^Y)$. ■

Note that the covariance operator of M is intricate but explicit since it has the form

$$C_M = T\,\Gamma^Y T^*. \tag{5.42}$$

COROLLARY 5.4

$$\sqrt{n}\,(\lambda_{jn}^{Y} - \lambda_j^{Y}) \xrightarrow{\mathcal{D}} N_j^{Y}$$

$$\sqrt{n}\left(\widehat{\lambda}_{jn}^{Y} - \lambda_j^{Y}\right) \xrightarrow{\mathcal{D}} N_j^{Y}$$

and

$$\sqrt{n}\,(\lambda_{jn} - \lambda_j) \xrightarrow{\mathcal{D}} N_j$$

$$\sqrt{n}\left(\widehat{\lambda}_{jn} - \lambda_j\right) \xrightarrow{\mathcal{D}} N_j,$$

where N_j^Y and N_j are zero-mean Gaussian random variables.

Proof

Similar to the proofs of Theorem 4.10 and Corollary 4.7. ■

An adaptation of Corollary 4.8 also holds.

5.5 Estimation of the autoregression order

Let X be a standard $ARH(p_0)$ with p_0 unknown. In this section we give some indications about estimation of p_0 in a special case. For the general case an empirical method will be proposed in Chapter 9.

Let us suppose that the assumptions in Theorem 5.3 hold, with $\boldsymbol{v}$ **known**. Then $(< X_n, v >)$ is an $AR(p_0)$ process that satisfies

$$< X_n, v >= \sum_{j=1}^{p_0} \alpha_j < X_{n-j}, v > + < \varepsilon_n, v >, \quad n \in \mathbb{Z}.$$

In order to estimate p_0 we now may use the **Shibata-Mourid** statistics

defined as

$$(SM)_n(p) = (n + \gamma(n)p\, n^{\beta})\widehat{\sigma}_v^2(p)\ , \quad p \geq 1, \tag{5.43}$$

with $\gamma(x) > 0$, $0 \leq \beta < 1$, and

$$\widehat{\sigma}_v^2(p) = \frac{1}{n}\sum_{i=p+1}^{n}\left(< X_i, v > - \sum_{j=1}^{p}\widehat{\alpha}_j(p) < X_{i-j}, v >\right)^2, \quad n > p,$$

where $(\widehat{\alpha}_1(p), \ldots, \widehat{\alpha}_p(p))$ is the estimator of $(\alpha_1, \ldots, \alpha_p)$ obtained from the empirical Yule-Walker equations

$$\widehat{c}_k = \sum_{j=1}^{p}\widehat{\alpha}_j\widehat{c}_{k-j}\ , \quad k = 1, \ldots, p,$$

with

$$\widehat{c}_h = \frac{1}{n-h} \sum_{\ell=1}^{n-h} < X_\ell, v >< X_{\ell+h}, v > ,$$

$h = 0, 1, \dots, p$.

Suppose now that $p_0 \in \{1, \dots, p_{max}\}$, where p_{max} is a given integer. The **estimator** of p_0 is then defined by

$$\widehat{p}_n = \min \left\{ p : 1 \leq p \leq p_{max},\ (SM)_n(p) = \min_{1 \leq j \leq p_{max}} (SM)_n(j) \right\}.$$

Note that if $j \mapsto (SM)_n(j)$ has only one minimum over $\{1, \dots, p_{max}\}$ then $\widehat{p}_n$ has the simpler form

$$\widehat{p}_n = \arg\min_{1 \leq j \leq p_{max}} (SM)_n(j).$$

Under the above assumptions one obtains a consistency result for $\widehat{p}_n$ which we present without proof:

THEOREM 5.14 *If $\gamma(n) \uparrow \infty$ and $\frac{\gamma(n)}{n} \to 0$ as $n \to \infty$, then*

$$P(\widehat{p}_n = p_0) \underset{n \to \infty}{\longrightarrow} 1 .$$

NOTES

5.1 The $ARH(p)$ model is introduced in Bosq (1990). Most of the results in this section are taken from Mourid (1993, 1995), with slight improvements.

5.2 The general definition of partial autocovariance and Theorem 5.5 are original.

5.3 Lemma 5.3 comes from Mourid (1995).

5.4 Theorem 5.13 is new.

5.5 Theorem 5.14 appears in Mourid (1995).

6
Autoregressive Processes in Banach Spaces

Observation of processes with continuous or differentiable sample paths takes place in physics, chemistry, finance, meteorology, and many other fields. In order to construct a random model adapted to such a situation, it is natural to use Banach spaces whose elements are regular functions, instead of general Hilbert spaces. The drawback of that choice is intricacy due to weakness of the geometrical properties of Banach spaces.

In this chapter we develop the study of autoregressive processes with values in a separable Banach space (ARB processes). Strong $ARB(1)$'s are considered in Section 1 and some ARB representations of real continuous-time processes in Section 2. Limit theorems are derived in Section 3.

The more general weak autoregressive process appears in Section 4 and is used in Section 5 to study empirical autocovariance operators.

Section 6 is devoted to the important special case of autoregressive processes in $C[0,1]$, while Section 7 considers applications to real continuous time processes.

6.1 Strong autoregressive processes in Banach spaces

Let B be a separable Banach space with its norm $\|\cdot\|$, Borel σ-field $\mathcal{B}_B$ and dual space B^*. The natural uniform norm on B^* is also denoted by $\|\cdot\|$. We use $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ to denote the set of bounded linear operators over B equipped with this uniform norm.

All random variables that appear below are defined on the same complete probability space $(\Omega, \mathcal{A}, P)$.

A sequence $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ of B-valued random variables is called a **strong B-white noise** (SBWN) if it is i.i.d. and such that $0 < E \parallel \varepsilon_0 \parallel^2 = \sigma^2 < \infty$ and $E\varepsilon_0 = 0$.

Let us give examples of SBWN.

Example 6.1
Take $B = C[0,1]$, consider a bilateral Wiener process W, and set

$$\varepsilon_n^{(\varphi)}(t) = \int_n^{n+t} \varphi(n+t-s)dW(s), \quad 0 \le t \le 1, \quad n \in \mathbb{Z}, \tag{6.1}$$

where φ is a square integrable real function such that $\displaystyle\int_0^1 \varphi^2(u)du > 0$. Then a continuous version of $\left(\varepsilon_n^{(\varphi)}\right)$ defines an SBWN (cf. Examples 1.10, 3.1 and 3.2).

We now define ARB processes.

DEFINITION 6.1 *Let $X = (X_n,\ n \in \mathbb{Z})$ be a strictly stationary sequence of B-random variables such that*

$$X_n - \mu = \rho(X_{n-1} - \mu) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{6.2}$$

where $\mu \in B$, $\rho \in \mathcal{L}$ and $\varepsilon = (\varepsilon_n)$ is a strong B-white noise.

Then X is said to be a (strong) autoregressive process of order 1 in B ($ARB(1)$) associated with (μ, ε, ρ).

If $B = H$, a separable Hilbert space, then X becomes a strictly stationary $ARH(1)$.

The next theorem deals with existence and uniqueness of X.

THEOREM 6.1 *Suppose that:*

(c_0) *There exists $j_0 \ge 1$ such that $\parallel \rho^{j_0} \parallel_{\mathcal{L}} < 1$.*

Then (6.2) has a unique strictly stationary solution given by

$$X_n = \mu + \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{6.3}$$

where the series converges in $L_B^2(\Omega, \mathcal{A}, P)$ and with probability 1.

Proof

We may take $\mu = 0$. Now, as in Lemma 3.1, it can be proved that

$$\parallel \rho^j \parallel_{\mathcal{L}} \le ab^j, \quad j \ge 1, \tag{6.4}$$

where $a \geq 1$ and $0 < b < 1$.

On the other hand,

$$E \parallel \sum_{j=m}^{m'} \rho^j(\varepsilon_{n-j}) \parallel^2 \leq \sum_{m \leq j,j' \leq m'} \parallel \rho^j \parallel_{\mathcal{L}} \parallel \rho^{j'} \parallel_{\mathcal{L}} E(\parallel \varepsilon_{n-j} \parallel \parallel \varepsilon_{n-j'} \parallel),$$

and the Cauchy-Schwarz inequality entails

$$\begin{aligned} E \parallel \sum_{j=m}^{m'} \rho^j(\varepsilon_{n-j}) \parallel^2 &\leq \left[\sum_{j=m}^{m'} \parallel \rho^j \parallel_{\mathcal{L}} \left(E \parallel \varepsilon_{n-j} \parallel^2\right)^{1/2} \right]^2 \\ &\leq \sigma^2 \left(\sum_{j=m}^{m'} \parallel \rho^j \parallel_{\mathcal{L}} \right)^2, \end{aligned}$$

which, by (6.4), implies convergence of the series $\sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j})$ in $L_B^2(\Omega, \mathcal{A}, P)$.

Almost sure convergence comes from GIN Theorem 2.1 ; uniqueness may be proved as in Theorem 3.1. ■

Immediate are the following properties of $ARB(1)$ processes. We suppose that (c_0) holds.

1) X_0 is integrable and $EX_0 = \mu$.

2) X_0 belongs to $L_B^2(\Omega, \mathcal{A}, P)$. Now suppose that $\mu = 0$ and define the covariance operator of X_0 by

$$C(x^*) = E(x^*(X_0)X_0) \ , \quad x^* \in B^*,$$

and the cross covariance operator of (X_0, X_1) by

$$D(x^*) = E(x^*(X_0)X_1) \ , \ \ x^* \in B^*.$$

Then

$$D = \rho\, C. \tag{6.5}$$

Thus it is natural to call ρ the **autocorrelation operator** of X.

3) Let $\mathcal{B}_n = \sigma\ (X_i,\ i \leq n) = \sigma(\varepsilon_i,\ i \leq n)$. Then

$$E^{\mathcal{B}_{n-1}}(X_n) = \rho(X_{n-1}) \ , \ \ n \in \mathbb{Z}, \tag{6.6}$$

so

$$\varepsilon_n = X_n - E^{\mathcal{B}_{n-1}}(X_n) \ , \ \ n \in \mathbb{Z}, \tag{6.7}$$

and (ε_n) may be called the **innovation process** of (X_n).

4) If $x^* \in B^*$ is an eigenvalue of the adjoint ρ^* of ρ associated with an eigenvalue $\alpha \ \in] - 1, +1[$, then $(x^*(X_n - \mu), n \in \mathbb{Z})$ is an $AR(1)$ (possibly degenerate), which satisfies

$$x^*(X_n - \mu) = \alpha\, [x^*(X_{n-1} - \mu)] + x^*(\varepsilon_n) \ , \ \ n \in \mathbb{Z}. \tag{6.8}$$

6.2 Autoregressive representation of some real continuous-time processes

We now show that some classical real continuous-time processes may be associated with particular autoregressive models in suitable Banach spaces.

Example 6.2
Let $\xi = (\xi_t,\ t \in \mathbb{R})$ be a continuous version of the **Ornstein-Uhlenbeck process** (1.15) with $\sigma^2 = 1$. Taking $B = C[0,h]$, where $h > 0$ is fixed, we set

$$X_n(t) = \xi_{nh+t}\ ,\ \ 0 \leq t \leq h\ ,\ \ n \in \mathbb{Z}. \tag{6.9}$$

By using Example 1.10 one sees that $X_n(\cdot)$ is a B-random variable.
Now consider the operator $\rho_\theta : C[0,h] \mapsto C[0,h]$ defined by

$$\rho_\theta(x)(t) = e^{-\theta t}x(h)\ ,\ \ 0 \leq t \leq h\ ;\ \ x \in C[0,h].$$

Then $\rho_\theta \in \mathcal{L}$, $\| \rho_\theta \|_{\mathcal{L}} = 1$, and more generally

$$\| \rho_\theta^j \|_{\mathcal{L}} = e^{-\theta(j-1)h}\ ;\ \ h \geq 1,$$

since $\rho_\theta^j = e^{-\theta(j-1)}\rho$.
On the other hand, define a B-white noise by putting

$$\varepsilon_n(t) = \int_{nh}^{nh+t} e^{-\theta(nh+t-s)}dW(s)\ ,\ \ t \in [0,h],\ \ n \in \mathbb{Z},$$

where $t \mapsto \varepsilon_n(t)$ is a continuous version of the stochastic integral.
Then (X_n) is clearly an $ARB(1)$ process that satisfies

$$X_n = \rho_\theta(X_{n-1}) + \varepsilon_n\ ,\ \ n \in \mathbb{Z}.$$

Therefore

$$\left(E^{\mathcal{B}_{n-1}}X_n\right)(t) = e^{-\theta t}X_{n-1}(h) = e^{-\theta t}\xi_{nh}.$$

Let us observe that the adjoint ρ_θ^* of ρ_θ is given by

$$\rho_\theta^*(m)(x) = x(h)\int_0^h e^{-\theta t}dm(t),$$

$m \in B^*$, $x \in C[0,h]$.
Then the Dirac measure $\delta_{(h)}$ appears as the one and only eigenvector of ρ_θ^* and the corresponding eigenvalue is $e^{-\theta h}$. From (6.8) it follows that $(X_n(h),\ n \in \mathbb{Z})$ is an $AR(1)$ such that

$$X_n(h) = e^{-\theta h}X_{n-1}(h) + \int_{nh}^{(n+1)h} e^{-\theta[(n+1)h-s]}dW(s)\ ,\ \ n \in \mathbb{Z}.$$

Compare with Example 3.4.

Example 6.3
Consider the following stochastic differential equation of order $q \geq 2$:

$$\sum_{\ell=0}^{q} a_\ell d\xi^{(\ell)}(t) = dW(t), \tag{6.10}$$

where $a_0, \ldots, a_q$ are real coefficients ($a_q \neq 0$) and W is a bilateral Wiener process.

In (6.10) differentiation up to order $q-1$ is ordinary, whereas the order q derivative is taken in Ito sense.

Let us suppose that the roots $-r_1, \ldots, -r_q$ of the polynomial equation $\sum_{\ell=0}^{q} a_\ell r^\ell = 0$ are real and such that $-r_q < \ldots < -r_1 < 0$.

Then it can be checked that the only stationary solution of (6.10) is the Gaussian process

$$\xi_t = \int_{-\infty}^{+\infty} g(t-s)dW(s)\ , \quad t \in \mathbb{R},$$

where g is a Green function associated with (6.10); that is, $g(t) = 0$ for $t < 0$ and, for $t \geq 0$, g is the unique solution of the problem

$$\sum_{\ell=0}^{q} a_\ell x^{(\ell)}(t) = 0, \tag{6.11}$$

$$x(0) = \ldots = x^{(q-2)}(0) = 0\ , \quad x^{(q-1)}(0) = a_q^{-1}. \tag{6.12}$$

By choosing a version of (ξ_t) such that every sample path possesses $q-1$ continuous derivatives, one may define a sequence of B-random variables via formula (6.9).

Here $B = C_{q-1}[0,h]$, that is, the family of real functions on $[0,h]$ having $q-1$ continuous derivatives, equipped with the norm

$$\| x \| = \sum_{\ell=0}^{q-1} \sup_{0 \leq t \leq h} |x^{(\ell)}(t)|.$$

This space is complete and each element y^* of B^* is uniquely representable in the form

$$y^*(x) = \sum_{\ell=0}^{q-2} \alpha_\ell x^{(\ell)}(0) + \int_0^h x^{(q-1)}(t)dm(t), \tag{6.13}$$

where $\alpha_0, \ldots, \alpha_{q-2}$ are real coefficients and m is a bounded signed measure over $[0,h]$ (see Dunford and Schwartz (1957)).

By using this representation and Lemma 1.2 one can easily show that X_n is actually a B-random variable.

Now, noting that

$$E(\xi_{nh+t} \mid \xi_s,\ s \le nh) = \sum_{j=0}^{q-1} \xi^{(j)}(nh)\varphi_j(t)\ ,\ \ 0 \le t \le h,\ n \in \mathbb{Z},$$

where φ_j is the unique solution of (6.11) satisfying the conditions $\varphi_j^{(\ell)}(0) = \delta_{j\ell}$; $\ell = 0, \ldots, q-1$, we are led to define ρ by

$$\rho(x)(t) = \sum_{j=0}^{q-1} x^{(j)}(h)\varphi_j(t)\ ,\ \ 0 \le t \le h,\ x \in B. \tag{6.14}$$

It is easy to check that the only eigenelements of ρ are the functions $t \mapsto e^{r_i t}$ associated with the eigenvalues $e^{-r_i h}$; $i = 1, \ldots, q$. Hence

$$\| \rho^j \|_{\mathcal{L}} = O(e^{r_1(j-1)\ell}).$$

Finally, (6.2) holds, with $\mu = 0$ and

$$\varepsilon_n(t) = \int_{nh}^{nh+t} g(nh + t - u)\, dW(u)\ ,\ \ 0 \le t \le h.$$

It is noteworthy that (ξ_t) is not Markovian while (X_n) is a Markov process.

Finally note that Bergström (1982) has shown that, if t_0 is any real number, $(\xi_{ih} + t_0,\ i \in \mathbb{Z})$ turns to be an $ARMA(q, q-1)$ real process.

Example 6.4 (process with seasonal component)
Let $(\eta_t,\ t \in \mathbb{R})$ be a real process such that

$$\eta_t = \mu(t) + \xi_t\ ,\ \ t \in \mathbb{R}, \tag{6.15}$$

where (ξ_t) is a zero-mean process admitting an $ARB(1)$ representation in B, a separable Banach space of real functions defined over $[0, h]$, and where $\mu(\cdot)$ is a nonconstant real function with period h such that $t \mapsto \mu(t)$, $0 \le t \le h$, belongs to B.

Then (η_t) clearly has an $ARB(1)$ representation given by

$$X_n(t) = \mu(t) + \xi_{nh+t},\ \ 0 \le t \le h,\ \ n \in \mathbb{Z},$$

and we have

$$EX_n = \mu.$$

Note that (X_n) is stationary, although it comes from a nonstationary real process.

6.3 Limit theorems

In order to derive limit theorems for *ARB* processes we first state a useful technical lemma.

LEMMA 6.1 (Utev's Lemma) *Let (ζ_n) be an equidistributed sequence of positive random variables with finite second moment, and let (a_n) be a summable positive sequence. Then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{n-j}\zeta_j \underset{n\to\infty}{\longrightarrow} 0 \quad a.s. \tag{6.16}$$

Proof

We observe that there exists a positive function f such that $f(x) \uparrow \infty$ as $x \uparrow \infty$ and

$$E\left[\zeta_1^2 f(\zeta_1)\right] < \infty.$$

Hence there exists a positive sequence $(u_n) \downarrow 0$ that satisfies

$$\sum_{n=1}^{\infty} P(\zeta_1 \geq u_n\sqrt{n}) < \infty.$$

Then, setting $\zeta_n' = \zeta_n \mathbf{I}_{\zeta_n < u_n\sqrt{n}}$ and applying Borel-Cantelli Lemma, we obtain

$$P(\limsup\{\zeta_n' \neq \zeta_n\}) = 0.$$

Thus it remains to prove (6.16) with (ζ_j') instead of (ζ_j). By construction,

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{n-j}\zeta_j' &\leq \frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{n-j}u_j\sqrt{j} \\ \leq \quad & \frac{1}{\sqrt{n}}(1+n^{1/4})\left(\sum_{j=0}^{\infty} a_j\right) + \frac{1}{\sqrt{n}}(\sqrt{n})\left(\sum_{j=0}^{\infty} a_j\right) u_{[\sqrt{n}]} \end{aligned}$$

and the bound tends to zero as n tends to infinity. ■

From now on we will say that X is a **standard** $ARB(1)$ if $\mu = 0$ and condition (c_0) holds ; in this case we set $r = \sum_{j=0}^{\infty} \| \rho^j \|_{\mathcal{L}}$. As usual we put

$S_n = \sum_{i=1}^{n} X_i$ and $\overline{\varepsilon}_n = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$.

We first deal with the strong law of large numbers.

THEOREM 6.2 (SLLN)
If X is a standard $ARB(1)$, then

$$\frac{S_n}{n} \underset{a.s.}{\longrightarrow} 0. \tag{6.17}$$

Proof

From condition (c_0) it follows that $(I-\rho)^{-1}$ exists and is bounded. We then have the decomposition

$$\frac{S_n}{n} = (I-\rho)^{-1}\overline{\varepsilon}_n + \frac{1}{n}(I-\rho)^{-1}\left[\rho^n(\varepsilon_1) + \ldots + \rho(\varepsilon_n)\right] + \frac{1}{n}(\rho + \ldots + \rho^n)(X_0). \tag{6.18}$$

We treat each term separately. Using the classical SLLN in Banach spaces (Theorem 2.4) and boundedness of $(I-\rho)^{-1}$, we obtain

$$(I-\rho)^{-1}\overline{\varepsilon}_n \underset{a.s.}{\longrightarrow} 0.$$

Now Utev's Lemma entails

$$\| \frac{1}{n}(I-\rho)^{-1}\left[\sum_{j=1}^{n}\rho^{n-j+1}(\varepsilon_j)\right] \| \leq \| (I-\rho)^{-1} \|_{\mathcal{L}} \frac{1}{n}\left[\sum_{j=1}^{n} \| \rho^{n-j+1} \|_{\mathcal{L}} \| \varepsilon_j \| \right] \underset{a.s.}{\longrightarrow} 0.$$

Finally,

$$\| \frac{1}{n}(\rho + \ldots + \rho^n)(X_0) \| \leq \frac{r \| X_0 \|}{n} \to 0.$$

Collecting the above results we get $\frac{S_n}{n} \underset{a.s.}{\longrightarrow} 0.$ ■

The next statement provides rates.

THEOREM 6.3 *Let B be a separable Banach space of type 2 with constant c, and X a standard $ARB(1)$ process. Then*

$$\frac{n^{1/4}}{(\text{Log} n)^{\beta}}\frac{S_n}{n} \underset{a.s.}{\longrightarrow} 0 \ , \ \beta > \frac{1}{2}. \tag{6.19}$$

If in addition $E\left(e^{\gamma\|X_0\|}\right) < \infty$ for some $\gamma > 0$, then for each $\eta > 0$ there exists an explicit integer $n_0 = \left[\frac{m_0(c,\rho,\sigma)}{\eta^2}\right]$ such that $n \geq n_0$ implies

$$P\left(\| \frac{S_n}{n} \| \geq \eta\right) \leq \alpha \exp\left(-\frac{n\eta^2}{\beta + \delta\eta}\right), \tag{6.20}$$

where α, β, and δ are explicit constants.

Finally, if $\| X_0 \|$ is bounded we have

$$\| \frac{S_n}{n} \| = O\left(\frac{\text{Log } n}{n}\right)^{1/2} \quad a.s. \tag{6.21}$$

Proof

Since B has type 2 we may use (2.26) to obtain the bound

$$E \parallel \varepsilon_n + \ldots + \varepsilon_{n+p-1} \parallel^2 \leq cp\sigma^2, \; p \geq 1, \; n \geq 1.$$

Then Corollary 2.3 implies

$$\frac{n^{1/4}}{(\mathrm{Log} n)^\beta} \overline{\varepsilon}_n \underset{a.s.}{\longrightarrow} 0 \;, \;\; \beta > \frac{1}{2}.$$

On the other hand,

$$\parallel \frac{1}{n} \sum_{j=1}^{n} \rho^j (\varepsilon_{n+1-j}) \parallel \leq \frac{1}{n} \sum_{j=1}^{n} \parallel \rho^j \parallel_{\mathcal{L}} \parallel \varepsilon_{n+1-j} \parallel$$

and from Utev's Lemma it follows that

$$n^{1/2} \frac{\sum_{j=1}^{n} \rho^j (\varepsilon_{n+1-j})}{n} \underset{a.s.}{\longrightarrow} 0.$$

Lastly

$$n^{1/2} \parallel \frac{1}{n} \left(\sum_{j=1}^{n} \rho^j \right) (X_0) \parallel \leq \frac{r \parallel X_0 \parallel}{n^{1/2}} \to 0.$$

Collecting the above results and using (6.18), one obtains (6.19).

Concerning (6.20), note first that

$$E \parallel \overline{\varepsilon}_n \parallel \leq (E \parallel \overline{\varepsilon}_n \parallel^2)^{1/2} \leq \; c^{1/2} n^{-1/2} \sigma$$

and that (6.2) and $E\left(e^{\gamma \parallel X_n \parallel}\right) < \infty$ yield

$$E\left(e^{\gamma \parallel \varepsilon_0 \parallel}\right) < \infty.$$

Now, using (6.18) again, we get

$$P\left(\parallel \frac{S_n}{n} \parallel \geq \eta \right) \leq P\left(\parallel \overline{\varepsilon}_n \parallel \geq \frac{\eta}{3} \parallel (I-\rho)^{-1} \parallel^{-1} \right)$$

$$+ \quad P\left(\frac{1}{n} \parallel \sum_{j=1}^{n} \rho^j (\varepsilon_{n+1-j}) \parallel \geq \frac{\eta}{3} (\parallel (I-\rho)^{-1} \parallel^{-1}) \right) + P\left(\parallel X_0 \parallel \geq \frac{n\eta}{3r} \right). \tag{6.22}$$

Let us set $\eta' = \frac{\eta}{3} \parallel (I-\rho)^{-1} \parallel^{-1}$ and observe that $\parallel \varepsilon_1 + \ldots + \varepsilon_n \parallel \geq n\eta'$ implies $\parallel \varepsilon_1 + \ldots + \varepsilon_n \parallel - E \parallel \varepsilon_1 + \ldots + \varepsilon_n \parallel \geq n\eta' - n^{1/2} c^{1/2} \sigma \geq n\frac{\eta'}{2}$, as soon as $n \geq \dfrac{4c\sigma^2}{\eta'^2}$.

We now may apply (2.20) to obtain

$$P(\| \overline{\varepsilon}_n \| \geq \eta') \leq 2\exp\left(-\frac{n\eta'^2}{8n\ell^2 + 4b\eta'}\right),$$

where ℓ and b can be identified as in the proof of Theorem 3.9.

In order to bound the second term in (6.22), one may again use (2.20) when the third term is directly bounded by using the condition $E\left(e^{\gamma\|X_0\|}\right) < \infty$. Details are left to the reader. Collecting these bounds one arrives at (6.20).

The proof of (6.21) is similar to that of Corollary 2.1 and is therefore omitted. ■

In order to establish weak laws we need some notation and definitions.

In the sequel $(\mathcal{L}', \|\cdot\|')$ will denote the space of bounded linear operators from B^* to B equipped with uniform norm

$$\| \ell \|' = \sup_{\|x^*\| \leq 1} \| \ell'(x^*) \|, \quad \ell' \in \mathcal{L}'.$$

Now $\ell \in \mathcal{L}'$ is called **nuclear** if it admits the representation

$$\ell(x^*) = \sum_{k=1}^{\infty} x_k^{**}(x^*) y_k, \quad x^* \in B^*, \tag{6.23}$$

where $(x_k^{**}) \subset B^{**}$ (the dual of B^*) and $(y_k) \subset B$ with $\sum_{k=1}^{\infty} \|x_k^{**}\| \|y_k\| < \infty$ (where $\|\cdot\|$ also denotes uniform norm in B^{**}). The infimum of sums $\sum_{k=1}^{\infty} \|x_k^{**}\| \|y_k\|$ for all such representations of ℓ is a norm (denoted $\|\cdot\|_{\mathcal{N}'}$) for which the space $\mathcal{N}'$ of nuclear operators from B^* to B becomes a Banach space.

Covariance and cross-covariance operators are nuclear. In the special case of a covariance operator C_X (with $E\|X\|^2 < \infty$ and $EX = 0$) we have the simpler representation

$$C_X(x^*) = \sum_{k=1}^{\infty} x^*(x_k) x_k, \quad x^* \in B^*, \tag{6.24}$$

where $(x_k) \subset B$ and $\sum_{k=1}^{\infty} \| x_k \|^2 < \infty$. Moreover,

$$\| C_X \|_{\mathcal{N}'} \leq E \| X \|^2. \tag{6.25}$$

Similarly, a cross-covariance operator satisfies

$$\| C_{X,Y} \|_{\mathcal{N}'} \leq E \| X \| \| Y \| . \tag{6.26}$$

After these preliminaries we state a general weak law of large numbers (WLLN) with exact rate.

THEOREM 6.4 *Let X be a standard $ARB(1)$. Then*

$$\| nC_{S_n/n} - \sum_{h=-\infty}^{+\infty} C_{X_0,X_h} \|_{\mathcal{N}'} \to 0. \tag{6.27}$$

Proof

First note that

$$X_h = \sum_{j=0}^{h-1} \rho^j(\varepsilon_{n-j}) + \rho^h(X_0), \quad h \geq 1. \tag{6.28}$$

Thus

$$C_{X_0,X_h} = C_{X_0,\rho^h(X_0)} \tag{6.29}$$

and (6.26) yields

$$\| C_{X_0,X_h} \|_{\mathcal{N}'} \leq \| \rho^h \|_{\mathcal{L}} E \| X_0 \|^2 , \quad h \geq 1, \tag{6.30}$$

so we may claim that $\sum_{h=-\infty}^{+\infty} C_{X_0,X_h}$ converges in $\mathcal{N}'$.

Now we have

$$\begin{aligned} y^* \left[C_{S_n/n}(x^*)\right] &= E\left[x^*\left(\frac{S_n}{n}\right) y^*\left(\frac{S_n}{n}\right)\right] \\ &= \frac{1}{n^2} \sum_{1 \leq i,j \leq n} E(x^*(X_i)y^*(X_i)), \end{aligned}$$

$x^*, y^* \in B^*$.

By using stationarity we obtain

$$y^* \left[\left(nC_{S_n/n}\right)(x^*)\right] = \sum_{|h| \leq n-1} \left(1 - \frac{|h|}{n}\right) E(x^*(X_0)y^*(X_h));$$

hence

$$nC_{S_n/n} = \sum_{|h| \leq n-1} \left(1 - \frac{|h|}{n}\right) C_{X_0,X_h},$$

and (6.27) follows from (6.30). ∎

Since (6.25) is only an inequality (it is an equality in a Hilbert space) one cannot directly deduce asymptotic properties of $E \| \frac{S_n}{n} \|^2$ from (6.27). The next statement gives partial results in special cases. Recal that $r = \sum_{j=0}^{\infty} \| \rho^j \|_{\mathcal{L}}$.

THEOREM 6.5 *Let X be a standard $ARB(1)$.*

(i) If B has type p with constant c, then

$$E \| \frac{S_n}{n} \|^p \leq \frac{cr^p}{n^{p-1}} E(\| \varepsilon_0 \|^p + \| \rho(X_0) \|^p). \tag{6.31}$$

(ii) If B has cotype q with constant c' and the operators $\sum_{j=0}^{n} \rho^j$, $n \geq 0$, are uniformly coercive (i.e., there exists $c_0 > 0$ such that $\| \sum_{j=0}^{n} \rho^j(x) \| \geq c_0 \| x \|$, $x \in B$, $n \geq 0$), then

$$E \| \frac{S_n}{n} \|^q \geq \frac{c' c_0^q}{n^{q-1}} E(\| \varepsilon_0 \|^q). \tag{6.32}$$

(iii) In particular, if B is a Hilbert space and ρ is symmetric compact then $\| \rho \|_{\mathcal{L}} < 1$ and

$$\frac{a_1}{n} \leq E \| \frac{S_n}{n} \|^2 \leq \frac{a_2}{n}, \tag{6.33}$$

where $a_1 = \left(\frac{1- \| \rho \|_{\mathcal{L}}}{1+ \| \rho \|_{\mathcal{L}}} \right)^2 E \| \varepsilon_0 \|^2$ and $a_2 = \frac{E \| X_0 \|^2}{1- \| \rho \|_{\mathcal{L}}^2}$.

Moreover, $a_1 = a_2$ if and only if ρ vanishes.

Proof

Consider the decomposition

$$\frac{S_n}{n} = \frac{1}{n} \sum_{j=0}^{n-1} (I + \rho + \ldots + \rho^j) \varepsilon_{n-j} + \left(\frac{\rho + \ldots + \rho^n}{n} \right) (X_0). \tag{6.34}$$

If B has type p, independence of $X_0, \varepsilon_1, \ldots, \varepsilon_n$ and Definition 2.3 imply

$$\begin{aligned} E \| \frac{S_n}{n} \|^p &\leq \frac{c}{n^p} \left[\sum_{j=0}^{n-1} E \| \left(\sum_{\ell=0}^{j} \rho^\ell \right) (\varepsilon_{n-j}) \|^p + E \| \left(\sum_{\ell=1}^{n} \rho^\ell \right) (X_0) \|^p \right], \\ &\leq \frac{cr^p}{n^p} [nE \| \varepsilon_0 \|^p + E \| \rho(X_0) \|^p] \end{aligned}$$

which gives (6.31).

Suppose now that B has cotype q, if $E \parallel \frac{S_n}{n} \parallel^q = \infty$. Then (6.32) is obvious. If not, we have

$$E \parallel \frac{S_n}{n} \parallel^q \geq \frac{c'}{n^q} \left[\sum_{j=0}^{n-1} E \parallel \left(\sum_{\ell=0}^{j} \rho^\ell \right) (\varepsilon_{n-j}) \parallel^q + E \parallel \left(\sum_{\ell=1}^{n} \rho^\ell \right) (X_0) \parallel^q \right],$$

and by uniform coercivity

$$E \parallel \frac{S_n}{n} \parallel^q \geq \frac{c' c_0^q}{n^q} \left[nE \parallel \varepsilon_0 \parallel^q + E \parallel \rho(X_0) \parallel^q \right],$$

which gives (6.32).

Now, if B is a Hilbert space and ρ is symmetric compact, then $\parallel \rho^{j_0} \parallel_{\mathcal{L}} = \parallel \rho \parallel_{\mathcal{L}}^{j_0}$ imply $\parallel \rho \parallel_{\mathcal{L}} < 1$. Furthermore, from (3.19) we get

$$\begin{aligned} \parallel \left(\sum_{j=0}^{n} \rho^j \right) (x) \parallel &= \sum_{k=0}^{\infty} (1 + \alpha_k + \ldots + \alpha_k^n)^2 < x, e_j >^2 \\ &\geq \left(\frac{1 - \parallel \rho \parallel_{\mathcal{L}}}{1 + \parallel \rho \parallel_{\mathcal{L}}} \right)^2 \parallel x \parallel^2, \;\; n \geq 0, \; x \in B, \end{aligned}$$

which means that the operators $\sum_{j=0}^{n} \rho^j$, $n \geq 0$ are uniformly coercive with constant $c_0 = \frac{1 - \parallel \rho \parallel_{\mathcal{L}}}{1 + \parallel \rho \parallel_{\mathcal{L}}}$.

Since a Hilbert space is of type and cotype 2, we may apply (6.31) and (6.32) (with $c = c' = 1$). Noting that, in this case, $E \parallel X_0 \parallel^2 = E \parallel \varepsilon_0 \parallel^2 + E \parallel \rho(X_0) \parallel^2$ we obtain, (6.33).

Finally, $\rho = 0$ yields $X_0 = \varepsilon_0$ and inequalities (6.33) are then equalities. Conversely, equality in (6.33) implies

$$1 \geq \frac{(1 - \parallel \rho \parallel_{\mathcal{L}})^3}{(1 + \parallel \rho \parallel_{\mathcal{L}})^2} = \frac{E(\parallel \rho(X_0) \parallel^2 + \parallel \varepsilon_0 \parallel^2)}{E \parallel \varepsilon_0 \parallel^2} \geq 1,$$

which is possible only if $\rho = 0$. ■

Let us now turn to CLT and LIL. We shall see that their validity depends on the structure of C_{ε_0}.

THEOREM 6.6 (CLT)
A standard $ARB(1)$ process X satisfies the CLT if and only if the innovation ε does, and in that case

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N \sim \mathcal{N}\left(0, (I - \rho)^{-1} C_{\varepsilon_0} (I - \rho^*)^{-1}\right). \tag{6.35}$$

Proof

Consider the decomposition

$$\frac{S_n}{\sqrt{n}} = (I-\rho)^{-1}\sqrt{n}\,\overline{\varepsilon}_n + \Delta_n, \tag{6.36}$$

where

$$\Delta_n = (I-\rho)^{-1}\frac{\rho^n\varepsilon_1 + \ldots + \rho\varepsilon_n}{\sqrt{n}} + \frac{(\rho+\ldots+\rho^n)(X_0)}{\sqrt{n}}.$$

By using the triangle inequality and Utev's Lemma we arrive at

$$\| \Delta_n \| \underset{a.s.}{\longrightarrow} 0.$$

Then, if

$$\sqrt{n}\,\overline{\varepsilon}_n \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0, C_{\varepsilon_0}),$$

continuity of $(I-\rho)^{-1}$ implies

$$(I-\rho)^{-1}\sqrt{n}\,\overline{\varepsilon}_n \xrightarrow{\mathcal{D}} N \sim \mathcal{N}\left(0, I-\rho)^{-1}C_{\varepsilon_0}(I-\rho^*)^{-1}\right)$$

and (6.35) follows from (2.9) and (6.36).

Conversely, if the CLT holds for $\dfrac{S_n}{\sqrt{n}}$, asymptotic normality of $(I-\rho)^{-1}\sqrt{n}\overline{\varepsilon}_n$ follows by using the same method, and since $(I-\rho)$ is continuous, we may conclude that $\sqrt{n}\,\overline{\varepsilon}_n$ satisfies the CLT. ■

Recall that Theorem 2.8 provides sufficient conditions for the CLT in the i.i.d. case, while Example 2.2 gives a counterexample.

Concerning the LIL, we use the notation in Theorem 2.10. We will say that ε satisfies the LIL if (2.32) and (2.33) hold, with $\sum_{i=1}^{n}\varepsilon_i$ instead of S_n.

THEOREM 6.7 (LIL)

A standard $ARB(1)$ X satisfies the LIL if and only if ε satisfies it. In this case we have

$$\lim_{n\to\infty} d\left(\frac{S_n}{\sqrt{2n\mathrm{LogLog}n}}, (I-\rho)^{-1}K\right) = 0 \ a.s., \tag{6.37}$$

and

$$c\left(\frac{S_n}{\sqrt{2n\mathrm{LogLog}n}}\right) = (I-\rho)^{-1}K \ a.s. \tag{6.38}$$

Proof

Similar to the proof of Theorem 6.6 and therefore omitted. ■

Theorem 2.10 supplies a sufficient condition for LIL in a Banach space of type 2.

6.4 Weak Banach autoregressive processes

We now give some ideas about a twofold extension of $ARB(1)$.

First we define a **weak B-white noise** *(WBWN)* as a sequence $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ of B-valued random variables such that

(i) $0 < E \parallel \varepsilon_n \parallel^2 = \sigma^2 < \infty$, $E\varepsilon_n = 0$, $n \in \mathbb{Z}$,

(ii) C_{ε_n} does not depend on n and $C_{\varepsilon_n, \varepsilon_m} = 0$; $n, m \in \mathbb{Z}$; $n \neq m$.

Now, given a *WBWN* ε, $\rho_1, \ldots, \rho_p \in \mathcal{L}$ (with $\rho_p \neq 0$), and $\mu \in B$, we will say that $X = (X_n,\ n \in \mathbb{Z})$ is a **weak autoregressive B-valued process of order p** ($WARB(p)$) associated with $(\varepsilon, \rho_1, \ldots, \rho_p, \mu)$ if it is weakly stationary and such that

$$X_n - \mu = \rho_1(X_{n-1} - \mu) + \ldots + \rho_p(X_{n-p} - \mu) + \varepsilon_n,\ n \in \mathbb{Z}. \tag{6.39}$$

Note that if the cartesian product B^p is equipped with the norm

$$\parallel (x_1, \ldots, x_p) \parallel_p = \sum_{j=1}^{p} \parallel x_j \parallel \ ; \quad x_1, \ldots, x_p \in B;$$

then $Y_n = (X_n, \ldots, X_{n-p+1})$, $n \in \mathbb{Z}$ defines a $WARB^p(1)$ process associated with the WB^pWN $\varepsilon'_n = (\varepsilon_n, 0, \ldots, 0)$, $n \in \mathbb{Z}$, the operator

$$\rho' = \begin{bmatrix} \rho_1 & \rho_2 & \ldots & \rho_p \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ 0 & \ldots & I & 0 \end{bmatrix},$$

and expectation $\mu' = (\mu, \ldots, \mu)$.

Owing to this remark we may restrict our study to the case $p = 1$ (as usual we replace ρ_1 by ρ).

Now, under condition (c_0), we can mimic the proof of Theorem 6.1 to obtain

$$X_n = \mu + \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}),\ \ n \in \mathbb{Z},$$

where the series converges in $L^2_B(\Omega, \mathcal{A}, P)$. Concerning pointwise convergence, notice that

$$\begin{aligned} E\left(\sum_0^{\infty} \parallel \rho^j \parallel \parallel \varepsilon_{n-j} \parallel\right) &= \lim_{n \uparrow \infty} \uparrow \sum_0^{n} \parallel \rho^j \parallel \ E \parallel \varepsilon_{n-j} \parallel \\ &\leq \sigma \left(\sum_0^{\infty} \parallel \rho^j \parallel\right) \\ &< \infty. \end{aligned}$$

Thus $\sum_0^\infty \| \rho^j \| \| \varepsilon_{n-j} \| < \infty$ a.s., and hence $\sum_0^\infty \rho^j(\varepsilon_{n-j})$ converges (a.s.).

Concerning **innovation** we first consider the case where (ε_n) is a **martingale difference** with respect to $\mathcal{B}_n = \sigma(\varepsilon_i,\ i \le n)$, $n \in \mathbb{Z}$ (see 2.55). Then we have

$$E^{\mathcal{B}_{n-1}}(X_n) = \rho(X_{n-1})$$

and we may say that (ε_n) is the innovation process of (X_n).

In the general case the lack of scalar product in $L^2_B(\Omega, \mathcal{A}, P)$ makes the problem of innovation more complicated. Suppose that $\mu = 0$ and consider the family of closed subspaces of $L^2(\Omega, \mathcal{A}, P)$ defined as

$$\mathcal{X}_n = \overline{sp}(y^*(X_i),\ \ i \le n,\ y^* \in B^*),\ \ n \in \mathbb{Z}.$$

Now, if x^* is a fixed element of B^*, the real process $(x^*(X_n),\ n \in \mathbb{Z})$ is zero-mean and weakly stationary, and satisfies

$$x^*(X_n) = (x^*\rho)(X_{n-1}) + x^*(\varepsilon_n),\ \ n \in \mathbb{Z}.$$

Noting that $(x^*\rho)(X_{n-1}) \in \mathcal{X}_{n-1}$ and $x^*(\varepsilon_n) \perp \mathcal{X}_n$ in $L^2(\Omega, \mathcal{A}, P)$, we may consider $(x^*(\varepsilon_n))$ as a generalized innovation process for $(x^*(X_n))$. In the particular case where $x^*(\rho)(X_n) \in \overline{sp}(x^*(X_i),\ i \le n)$ (for example, if x^* is an eigenvector of ρ^*), $(x^*(\varepsilon_n))$ is the innovation process of $(x^*(X_n))$ in the usual sense. In view of the above properties, (ε_n) may be considered as a **weak innovation** for (X_n).

We now turn to **limit theorems** for *WARB* processes. We first give two weak type results.

THEOREM 6.8 *Let X be a standard $WARB(1)$ process. Then*

$$\frac{n^{1/4}}{(\mathrm{Log}\ n)^\beta}\ x^*\left(\frac{S_n}{n}\right) \xrightarrow[a.s.]{} 0,\ \ x^* \in B^*,\ \beta > \frac{1}{2}; \tag{6.40}$$

and

$$n\ E\left(\frac{x^*(S_n)}{n}\right)^2 \to \sum_{h=-\infty}^{+\infty} E\left(x^*(X_0)x^*(X_h)\right),\ x^* \in B^*, \tag{6.41}$$

moreover,

$$\sup_{\|x^*\| \le 1} E\left(\frac{x^*(S_n)}{n}\right)^2 \le \frac{2E \| X_0 \|^2 r}{n},\ \ n \ge 1. \tag{6.42}$$

Proof

From the decomposition

$$x^*(X_h) = \sum_{j=0}^{h-1} x^*\left(\rho^j(\varepsilon_{h-j})\right) + (x^*\rho^h)(X_0),\ \ x^* \in B^*$$

it follows that

$$E(x^*(X_0)x^*(X_h)) = E\left((x^*\rho^h)(X_0)x^*(X_0)\right).$$

Therefore

$$|E\left(x^*(X_0)x^*(X_h)\right)| \leq \| x^* \|^2 \| \rho^h \|_{\mathcal{L}} E \| X_0 \|^2,$$

and weak stationarity implies

$$E\left|x^*(X_n) + \ldots + x^*(X_{n+p-1})\right|^2 \leq 2p \| x^* \|^2 \sum_{h=0}^{p-1} \| \rho^h \|_{\mathcal{L}} E \| X_0 \|^2 . \tag{6.43}$$

Applying Corollary 2.3 to the real random sequence $(x^*(X_i),\ i \geq 1)$ one obtains (6.40).

Finally, (6.43) yields (6.41) and (6.42) . ■

Concerning CLT we have

THEOREM 6.9 *If X is a standard $WARB(1)$ and (ε_n) is a martingale difference such that $\sup_{i\geq 1} E \| \varepsilon_i \|^{2+\delta} < \infty$ for some $\delta > 0$, then*

$$x^*\left(\frac{S_n}{\sqrt{n}}\right) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0, E(x^*(I-\rho)^{-1}(\varepsilon_0))^2), \quad x^* \in B^*. \tag{6.44}$$

Proof

From (6.36) it follows that

$$x^*\left(\frac{S_n}{\sqrt{n}}\right) = \frac{\sum_{i=1}^{n} x^*(I-\rho)^{-1}(\varepsilon_i)}{\sqrt{n}} + x^*(\Delta_n), \quad n \geq 1,$$

and Utev's Lemma implies $x^*(\Delta_n) \underset{a.s.}{\longrightarrow} 0$.

Now $\left(x^*(I-\rho)^{-1}(\varepsilon_i),\ i \geq 1\right)$ is clearly a martingale difference such that $\sup_{i\geq 1} E|x^*(I-\rho^{-1})(\varepsilon_i)|^{2+\delta} < \infty$. Then it suffices to apply Theorem 2.16 with $H = \mathbb{R}$ to obtain (6.44). ■

Concerning "strong type" results we have the following principle already used for $ARB(1)$: $\boldsymbol{X}$ **satisfies a limit theorem if and only if** $\boldsymbol{\varepsilon}$ **satisfies it**. This principle comes from (6.36) and Utev's Lemma. It applies to $SLLN$, CLT and LIL.

For example, if ε is **strongly mixing** and satisfies assumptions in Theorem 2.17, it follows that the CLT holds for X.

As another example let us quote Theorem 2.15, which allows to obtain, strong law of large numbers for $WARB(1)$.

6.5 Estimation of autocovariance

Let X be a standard $WARB(1)$. Its autocovariance is the sequence $(C_{X_0,X_h},\ h \in \mathbb{Z})$.

In order to estimate $C := C_{X_0,X_0}$ we use the **empirical covariance operator** defined as

$$C_n(x^*) = \frac{1}{n}\sum_{i=1}^{n} x^*(X_i)X_i, \quad x^* \in B^*, \quad n \geq 1. \tag{6.45}$$

This estimator is **unbiased**, i.e., $EC_n = C$, where expectation is taken in $\mathcal{L}'$.

The following lemma is useful for asymptotics.

LEMMA 6.2 *Let X be a standard $WARB(1)$ such that $E\,\|X_n\|^4 = E \parallel X_0 \parallel^4 < \infty$ and ε is a martingale difference satisfying*

$$E^{\mathcal{B}_0}\left(u^*(\varepsilon_k)v^*(\varepsilon_k)\right) = E\left(u^*(\varepsilon_k)v^*(\varepsilon_k)\right); \tag{6.46}$$

$u^*, v^* \in B^*$; $k \geq 1$.

Then there exist two explicit constants $c_1 > 0$ and $c_2 \in]0,1[$ such that

$$|Cov(x^*(X_0)y^*(X_0),\ x^*(X_h)y^*(X_h)| \leq \parallel x^* \parallel^2 \parallel y^* \parallel^2 c_1 c_2^h; \tag{6.47}$$

$x^*, y^* \in B^*$; $h \geq 1$.

Proof

Let us set

$$e_h = \sum_{j=0}^{h-1} \rho^j(\varepsilon_{h-j}) = X_h - \rho^h(X_0), \quad h \geq 1.$$

First we have

$$\begin{aligned}
E\left(x^*(X_0)y^*(X_0)x^*(X_h)y^*(X_h)\right) = \\
&= E\left(x^*(X_0)y^*(X_0)x^*(e_h)y^*(e_h)\right) \\
&\quad + E\left(x^*(X_0)y^*(X_0)x^*(e_h)y^*\left(\rho^h(X_0)\right)\right) \\
&\quad + E\left(x^*(X_0)y^*(X_0)y^*(e_h)x^*\left(\rho^h(X_0)\right)\right) \\
&\quad + E\left(x^*(X_0)y^*(X_0)x^*(\rho^h(X_0))y^*\left(\rho^h(X_0)\right)\right) \\
&:= A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

The martingale difference property yields

$$A_2 = A_3 = 0. \tag{6.48}$$

On the other hand,

$$|A_4| \leq \parallel \rho^h \parallel_{\mathcal{L}}^2 \parallel x^* \parallel^2 \parallel y^* \parallel^2 E \parallel X_0 \parallel^4 . \tag{6.49}$$

Now, using again the martingale property of ε, we obtain

$$A_1 = \sum_{j=0}^{h-1} E\left(x^*(X_0)y^*(X_0)(x^*\rho^j)(\varepsilon_{h-j})(y^*\rho^j)(\varepsilon_{h-j})\right)$$

and from (6.46) it follows that

$$A = \sum_{j=0}^{h-1} E\left(x^*(X_0)y^*(X_0)\right) E\left((x^*\rho^j)(\varepsilon_{h-j})(y^*\rho^j)(\varepsilon_{h-j}\right).$$

Noting that

$$\begin{aligned} y^*C(x^*) &= \sum_{j=0}^{\infty} E\left((x^*\rho^j)(\varepsilon_{h-j})(y^*\rho^j)(\varepsilon_{h-j})\right) \\ &= E\left(x^*(X_0)y^*(X_0)\right), \end{aligned}$$

we obtain

$$(y^*C(x^*))^2 - A_1 = y^*C(x^*) \sum_{j=h}^{\infty} \left((x^*\rho^j)(\varepsilon_{h-j})(y^*\rho^j)(\varepsilon_{h-j})\right);$$

therefore

$$\left|(y^*C(x^*))^2 - A_1\right| \le \| y^* \|^2 \| x^* \|^2 \| C \|_{\mathcal{L}} \sum_{j=h}^{\infty} \| \rho^j \|_{\mathcal{L}}^2 E \| \varepsilon_0 \|^2 . \tag{6.50}$$

Now, by using Lemma 3.1 together with (6.48), (6.49), and (6.50), we get

$$|Cov\left(x^*(X_0)y^*(X_0),\ x^*(X_h)y^*(X_h)\right)| \le a_1 a^2 b^{2h} + \frac{b_1 a^2 b^{2h}}{1-b^2},$$

where $a_1 = \| x^* \|^2 \| y^* \|^2 E \| X_0 \|^4$

and $b_1 = \| x^* \|^2 \| y^* \|^2 \| C \|_{\mathcal{L}} E \| \varepsilon_0 \|^2$. Hence (6.47) holds. ■

Condition (6.46) is rather heavy but classical in this context; it is obviously valid if (ε_n) is an independent sequence of random variables.

Lemma 6.2 allows us to derive weak asymptotic results concerning $C_n - C$.

THEOREM 6.10 *If the assumptions in Lemma 6.2 hold then, for all x^*, $y^* \in B^*$,*

$$n\ E\left[y^*(C_n - C)(x^*)\right]^2 \longrightarrow \sum_{h=-\infty}^{+\infty} Cov\left(x^*(X_0)y^*(X_0), x^*(X_h)y^*(X_h)\right) \tag{6.51}$$

and

$$\limsup_{n\to\infty} n \sup_{\substack{\| x^* \| \leq 1 \\ \| y^* \| \leq 1}} E\left[y^*(C_n - C)(x^*)\right]^2 < \infty. \tag{6.52}$$

Furthermore,

$$\frac{n^{1/4}}{(\text{Log } n)^\beta}\, y^*(C_n - C)(x^*) \underset{a.s.}{\longrightarrow} 0, \ \ \beta > \frac{1}{2}. \tag{6.53}$$

Proof

(6.52) and (6.53) are straightforward consequences of Lemma 6.2. Statement (6.51) comes from Corollary 2.3 applied to the real random variables $y^*(X_i)\, x^*(X_i) - y^*C(x^*)$, $i \geq 1$. ■

We now give an autoregressive representation of the sequence of random operators defined as

$$Z_i(x^*) = x^*(X_i)X_i - C \ \ , \ i \in \mathbb{Z}.$$

$Z = (Z_i)$ defines a sequence of $\mathcal{L}'$-valued random variables. Below, $\mathcal{L}''$ will denote the space of bounded operators from $\mathcal{L}'$ to $\mathcal{L}'$ with its usual norm $\| \cdot \|_{\mathcal{L}''}$.

LEMMA 6.3 *Let X be a standard $ARB(1)$ process with $E \| X_0 \|^4 < \infty$. Then $(Z_i,\ i \geq 1)$ is a $WAR\mathcal{L}(1)$ process such that*

$$Z_i = R(Z_{i-1}) + E_i \ , \ \ i \in \mathbb{Z}, \tag{6.54}$$

where R is an operator over $\mathcal{L}'$ defined by

$$R(\ell) = \rho\ell\rho^* \ \ , \ \ \ell \in \mathcal{L}',$$

with

$$\| R^h \|_{\mathcal{L}''} \leq \| \rho^h \|_{\mathcal{L}}^2 \ , \ \ h \geq 1,$$

and where $E = (E_i) \subset L^2_{\mathcal{L}'}(\Omega, \mathcal{B}_i, P)$ is a $W\mathcal{L}'WN$ and a martingale difference in $\mathcal{L}'$ with respect to $(\mathcal{B}_i)$.

Proof

Similar to the proof of Lemma 4.1. ■

Here E_i is given by

$$E_i(x^*) = x^*(\varepsilon_i)\varepsilon_i + (x^*\rho(X_{i-1}))\, \varepsilon_i + (x^*(\varepsilon_i))\, X_{i-1} - C_{\varepsilon_0}(x^*), \ x^* \in B^*.$$

The above representation allows us to improve Theorem 6.9, provided X is an $ARB(1)$ (instead of a $WARB(1)$). Actually, Theorem 6.7 may be applied to study the asymptotic behavior of $\ell^*(C_n - C)$, where ℓ^* is any element in $\mathcal{L}'^*$. This is more general than Theorem 6.9 since $\ell \mapsto y^*[\ell(x^*)]$, $x^*, y^* \in B^*$, defines a particular family in $\mathcal{L}'^*$. Details are left to the reader.

Another application is the following:

COROLLARY 6.1 *If X is a standard $ARB(1)$ and (E_i) satisfies $E \| E_i \|_{\mathcal{L}'}^{2+\delta} < \infty$ for some $\delta > 0$, then*

$$\ell^*[\sqrt{n}(C_n - C)] \xrightarrow{\mathcal{D}} N_1(\ell^*) \,, \quad \ell^* \in \mathcal{L}'^*, \tag{6.55}$$

where $N_1(\ell^*)$ is a real zero-mean Gaussian random variable with variance $E\left(\ell^*(I-R)^{-1}(E_0)\right)^2$.

Proof

Apply Theorem 6.8 to the $WAR\mathcal{L}'(1)$ process $(Z_i,\ i \in \mathbb{Z})$. ■

Finally we again have the principle: ***Z* satisfies a limit theorem if and only if *E* satisfies it**. Thus under strong mixing conditions one may derive asymptotic normality of $\sqrt{n}(C_n - C)$.

Let us now give an application of this principle to Hilbert spaces.

COROLLARY 6.2 *Let X be a standard $ARH(1)$ associated to a Hilbert-Schmidt operator ρ and a strong H-white noise ε, and such that $E \| X_0 \|^4$ is finite.*

Then

$$\| C_n - C \|_{\mathcal{N}} \underset{a.s.}{\longrightarrow} 0, \tag{6.56}$$

and consequently

$$\frac{1}{n}\sum_{i=1}^{n} \| X_i \|^2 \underset{a.s.}{\longrightarrow} E \| X_0 \|^2 \,. \tag{6.57}$$

Proof

A straightforward verification shows that representation (6.54) takes place in the Banach space $\mathcal{N}$ of nuclear operators over H with

$$\| R(\ell) \|_{\mathcal{N}} \leq \| \rho \|_{\mathcal{L}}^2 \| \ell \|_{\mathcal{N}}, \quad \ell \in \mathcal{N}.$$

This implies (6.56) by using Theorem 2.15.

Now, since

$$\| C_n \|_{\mathcal{N}} \underset{a.s.}{\longrightarrow} \| C \|_{\mathcal{N}},$$

we obtain (6.57). ■

Note that (6.57) may be written in the alternative form

$$\sum_{j=1}^{\infty} \widehat{\lambda}_{jn} = \sum_{j=1}^{\infty} \lambda_{jn} \underset{a.s.}{\longrightarrow} \sum_{j=1}^{\infty} \lambda_j .$$

Concerning estimation of $D := C_{X_0,X_1}$, the natural estimator is

$$D_n(x^*) = \frac{1}{n-1}\sum_{i=1}^{n-1} x^*(X_i)X_{i+1} \ , \ x^* \in B^*, \tag{6.58}$$

which has properties very similar to those of C_n.

6.6 The case of $C[0,1]$

The case $B = C[0,1]$ is special, since this space has neither type nor cotype apart from the trivial ones. It is very important for applications since many continuous-time observed processes possess continuous sample paths. Standard autoregressive processes in this space will be denoted $WARC(1)$ (instead of $WARB(1)$) and $ARC(1)$ (instead of $ARB(1)$).

Note first that various results quoted above remain valid for $C[0,1]$. Actually, this is tune of all the results that do not use type or cotype such as the $SLLN$ (6.17) or the $WLLN$ (6.27).

The weak type results in Section 6.4 induce pointwise convergences, since for each t the Dirac measure $\delta(t)$ belongs to the dual space of $C[0,1]$. Specifically, (6.40) yields

$$\frac{n^{1/4}}{(\mathrm{Log}n)^{\beta}}\frac{S_n(t)}{n} \underset{a.s.}{\longrightarrow} 0 \ , \quad t \in [0,1], \ \beta > \frac{1}{2},$$

and (6.41) implies

$$n\,E\left(\frac{S_n(t)}{n}\right)^2 \longrightarrow \sum_{h=-\infty}^{+\infty} E(X_0(t)X_h(t)).$$

On the other hand, under the hypothesis in Theorem 6.9, it is easy to show that $\frac{1}{\sqrt{n}}(S_n(t_1),\ldots,S_n(t_k))$ converges in distribution to a Gaussian random vector for all $k-$tuples $t_1,\ldots,t_k$ in $[0,1]^k$.

Similarly, Theorem 6.10 and Corollary 6.1 imply, for each $(s,t) \in [0,1]^2$,

$$n\,E[c_n(s,t) - c(s,t)]^2 \longrightarrow \sum_{h=-\infty}^{+\infty} Cov\left(X_0(s)X_0(t), X_h(s)X_h(t)\right),$$

$$\frac{n^{1/4}}{(\mathrm{Log}n)^{\beta}}[c_n(s,t) - c(s,t)] \underset{a.s.}{\longrightarrow} 0,$$

and

$$\sqrt{n}[c_n(s,t) - c(s,t)] \xrightarrow{\mathcal{D}} N_1,$$

where

$$\begin{aligned} c_n(s,t) &= \delta_{(t)}[C_n(\delta_{(s)})] &= \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t), \\ c(s,t) &= \delta_{(t)}[C(\delta_{(s)})] &= E(X_s X_t), \end{aligned}$$

and N_1 is Gaussian.

Now, in order to obtain uniform results, we are led to ask for more regularity of observed sample paths. For this purpose we introduce the following assumption.

$$(L_\alpha) \qquad |X_n(t) - X_n(s)| \leq M_n|t-s|^\alpha \ ; \ \ s,t \in [0,1], \ \ n \in \mathbb{Z},$$

where $\alpha \in]0,1]$ and M_n is a square integrable real random variable.

This condition allows us to evaluate the level of information associated with observations of X at discrete instants.

Note that one may set

$$M_n = \sup_{s\neq t} \frac{|X_n(t) - X_n(s)|}{|t-s|^\alpha}. \tag{6.59}$$

In the remainder of the text we will always suppose that the sequence (M_n) associated with an $ARC(1)$ process is equidistributed and we will put $V = EM_0^2$. Stationarity makes these conditions natural.

THEOREM 6.11 *Let X be a $WARC(1)$. If (L_α) holds for some $\alpha \in]0,1]$, then*

$$E\left[\sup_{0\leq t\leq 1}\left|\frac{S_n(t)}{n}\right|^2\right] = O\left(n^{-\frac{2\alpha}{2\alpha+1}}\right). \tag{6.60}$$

Proof

Let us set $t_j = \dfrac{j}{\nu_n}$, $1 \leq j \leq \nu_n$. From (L_α) we get

$$\left|\frac{S_n(t)}{n} - \frac{S_n(t_j)}{n}\right| \leq \left(\frac{1}{n}\sum_{i=1}^{n} M_i\right)\nu_n^{-\alpha} \ ,$$

$\dfrac{j-1}{\nu_n} \leq t \leq \dfrac{j}{\nu_n}$; $i = 1, \ldots, \nu_n$. Therefore

$$\left|\sup_{0\leq t\leq 1}\left|\frac{S_n(t)}{n}\right| - \sup_{1\leq j\leq \nu_n}\left|\frac{S_n(t_j)}{n}\right|\right| \leq \left(\frac{1}{n}\sum_{i=1}^{n} M_i\right)\nu_n^{-\alpha}, \tag{6.61}$$

which yields

$$\begin{aligned} E \parallel \frac{S_n}{n} \parallel^2 &\leq 2E\left(\sup_{1\leq i\leq \nu_n}\left|\frac{S_n(t_j)}{n}\right|\right)^2 + 2E\left(\frac{1}{n}\sum_{i=1}^{n} M_i\right)^2 \nu_n^{-2\alpha} \\ &\leq 2\sum_{i=1}^{\nu_n} E\left(\left|\frac{S_n(t_j)}{n}\right|^2\right) + 2E\left(\frac{1}{n}\sum_{i=1}^{n} M_i^2\right)\nu_n^{-2\alpha}. \end{aligned}$$

Then, using (6.42), one gets,

$$E \parallel \frac{S_n}{n} \parallel^2 \leq (4rE \parallel X_0 \parallel^2)\frac{\nu_n}{n} + 2V\nu_n^{-2\alpha} \tag{6.62}$$

and the choice $\nu_n = \left[n^{\frac{1}{2\alpha+1}}\right] + 1$ gives (6.60). ■

The next statement deals with large deviations.

THEOREM 6.12 *Let X be an ARC(1) such that $E\left(e^{\gamma\|X_0\|}\right) < \infty$ for some $\gamma > 0$. Then*

$$P\left(\left|\frac{S_n(t)}{n}\right| \geq \eta\right) \leq 4\exp\left(-\frac{n\eta^2}{d_1 + d_2\eta}\right) + 2\exp(-d_3 n\eta)E\left(e^{\gamma\|X_0\|}\right), \tag{6.63}$$

$\eta > 0$, $0 \leq t \leq 1$, $n \geq 1$, where d_1, d_2, and d_3 are strictly positive real coefficients that depend on ρ and P_ε but not on t and η.

If in addition (L_α) holds for some $\alpha \in]0,1]$, with M_0 bounded, then

$$P\left(\sup_{0\leq t\leq 1}\left|\frac{S_n(t)}{n}\right| \geq \eta\right) \leq a(\eta)\exp(-nb(\eta)), \quad \eta > 0, \tag{6.64}$$

where $a(\eta) > 0$ and $b(\eta) > 0$ are explicit.

Proof

First note that if $\gamma' = \frac{\gamma}{2}\min(1, \parallel \rho \parallel_{\mathcal{L}}^{-1})$ then

$$\begin{aligned}
E\left(e^{\gamma'\|\varepsilon_n\|}\right) &\leq E\left(e^{\gamma'\|X_n\|} \,.\, e^{\gamma'\|\rho\|_{\mathcal{L}}\,\|X_{n-1}\|}\right)\\
&\leq \left(E\left(e^{2\gamma'\|X_n\|}\right)\right)^{1/2}\left(E\left(e^{2\gamma'\|\rho\|_{\mathcal{L}}\|X_{n-1}\|}\right)\right)^{1/2}\\
&\leq E\left(e^{\gamma\|X_0\|}\right) < \infty.
\end{aligned}$$

Therefore

$$E\left(e^{\lambda\|\phi_n\|}\right) < \infty,$$

where $\phi_n = (I-\rho)^{1/2}(\varepsilon_n)$ and $\lambda = \gamma' \parallel (I-\rho) \parallel_{\mathcal{L}}^{-1}$.

Now, for each t, we may apply inequality (2.21) to $\overline{\phi}_n(t) = \frac{1}{n}\sum_{i=1}^{n}\phi_i(t)$. We obtain

$$P(|\overline{\phi}_n(t)| \geq \eta) \leq 2\exp\left(-\frac{n\eta^2}{2\ell_0^2 + 2b\eta}\right), \quad \eta > 0, \tag{6.65}$$

where ℓ_0 and b are constants. In fact, since

$$E\left(e^{\lambda\|\phi_i\|}\right) \geq \frac{\lambda^k|\phi_i(t)|^k}{k!}, \quad k \geq 2,$$

we have

$$\sum_{i=1}^{n} E|\phi_i(t)|^k \leq \frac{k!}{2}\, n\ell_0^2 b^{k-2},$$

where $\ell_0 = 2\lambda^{-2}E\left(e^{\lambda\|\phi_0\|}\right)$ and $b = \lambda^{-1}$.

Thus ℓ_0 and b are explicit and do not depend on (t,η).

Similarly we may apply (2.21) to $\overline{\psi}_n(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \rho^j(\varepsilon_{n+1-j})(t)$. Here we have

$$\begin{aligned} E\left(e^{\delta|\rho^j(\varepsilon_{n+1-j})(t)|}\right) &\leq E\left(e^{\delta\|\rho^j\|_{\mathcal{L}}\|\varepsilon_0\|}\right) \\ &\leq E\left(e^{\lambda\|\varepsilon_0\|}\right) < \infty \end{aligned}$$

provided $\delta = \lambda R^{-1}$ where $R = \max_{j\geq 0} \| \rho^j \|_{\mathcal{L}}$, and as above we may write

$$\sum_{j=1}^{n} E\left|\rho^j(\varepsilon_{n+1-j})(t)\right|^k \leq \frac{k!}{2} n\ell_0'^2 b'^{k-2},$$

with $\ell_0' = 2(\delta R)^{-2}E\left(e^{\delta R\|\varepsilon_0\|}\right)$ and $b' = \delta^{-1}$.

Hence

$$P\left(\left|\overline{\phi}_n(t)\right| \geq \eta\right) \leq 2\exp\left(-\frac{n\eta^2}{2\ell_0'^2 + 2b'\eta}\right). \tag{6.66}$$

Finally,

$$P\left(\left|\frac{(\rho + \ldots + \rho^n)(X_0)(t)}{n}\right| \geq \eta\right) \leq 2e^{-dn\eta}E\left(e^{dr\|X_0\|}\right), \tag{6.67}$$

where $d = \gamma r^{-1}$.

Collecting (6.65), (6.66), and (6.67) and using (6.18) one easily obtains (6.63).

Now suppose that (L_α) holds, with $\|M_0\|_\infty < \infty$, and consider the integer

$$\nu_\eta := \left[\left(2\,\|M_0\|_\infty\,\eta^{-1}\right)^{1/\alpha}\right] + 1.$$

From (L_α) it follows that

$$\begin{aligned} \left|\frac{S_n(t) - S_n\left(\dfrac{j}{\nu_\eta}\right)}{n}\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\left|X_i(t) - X_i\left(\frac{j}{\nu_\eta}\right)\right| \\ &\leq \| M_0 \|_\infty\, \nu_\eta^{-\alpha}\,,\quad \frac{j-1}{\nu_\eta} \leq t \leq \frac{j}{\nu_\eta}\,,\ j = 1,\ldots,\nu_\eta. \end{aligned}$$

Consequently

$$\| \frac{S_n}{n} \| \leq \max_{1\leq j\leq \nu_\eta}\left|\frac{S_n(j/\nu_\eta)}{n}\right| + \| M_0 \|_\infty\, \nu_\eta^{-\alpha},$$

and

$$P\left(\| \frac{S_n}{n} \| \geq \eta\right) \leq \sum_{j=1}^{\nu_\eta} P\left(\left|\frac{S_n(j/\nu_\eta)}{n}\right| \geq \frac{\eta}{2\nu_\eta}\right)$$

since $\nu_\eta^\alpha > 2 \| M_0 \|_\infty \eta^{-1}$.

Now we are in a position to use (6.63) at each point j/ν_η. We obtain

$$\begin{aligned} P\left(\| \frac{S_n}{n} \| \geq \eta\right) &\leq 4\nu_\eta \exp\left(-\frac{n}{4\nu_\eta^2}\left(d_1 + d_2\frac{\eta}{2\nu_\eta}\right)\right) \\ &\quad + 2\nu_\eta \exp\left(-\frac{d_3 n\eta}{2\nu_\eta}\right) E\left(e^{\gamma\|X_0\|}\right), \end{aligned} \tag{6.68}$$

which gives (6.64). ■

Concerning CLT we have the following.

THEOREM 6.13 *If X is an $ARC(1)$ and ε satisfies condition (L_1), then X satisfies the CLT.*

Proof

Apply Theorem 2.8 to (ε_n) and then apply Theorem 6.6. ■

We now turn to uniform asymptotic properties for c_n. Note first that if (L_α) holds, then

$$\begin{aligned} \left|X_i(s)X_i(t) - X_i\left(\ell\nu_n^{-1}\right)X_i\left(\ell'\nu_n^{-1}\right)\right| &\leq |X_i(s)|\left|X_i(t) - X_i\left(\ell'\nu_n^{-1}\right)\right| \\ &\quad + \left|X_i\left(\ell'\nu_n^{-1}\right)\right|\left|X_i(s) - X_i\left(\ell\nu_n^{-1}\right)\right| \\ &\leq 2 \| X_i \| M_i\nu_n^{-\alpha}, \end{aligned} \tag{6.69}$$

provided $\ell\nu_n^{-1} \leq s \leq (\ell+1)\nu_n^{-1}$ and $\ell'\nu_n^{-1} \leq t \leq (\ell'+1)\nu_n^{-1}$.

Similarly,

$$\left|E(X_i(s)X_i(t)) - E\left(X_i\left(\ell\nu_n^{-1}\right)X_i\left(\ell'\nu_n^{-1}\right)\right)\right| \leq 2E(\| X_i \| M_i)\nu_n^{-\alpha},$$

and hence

$$\| c_n - c \|_2 \leq N_n\nu_n^{-\alpha} + \max_{1\leq \ell,\ell'\leq \nu_n}\left|c_n\left(\ell\nu_n^{-1},\, \ell'\nu_n^{-1}\right) - c\left(\ell\nu_n^{-1},\, \ell'\nu_n^{-1}\right)\right|, \tag{6.70}$$

where $N_n = \frac{2}{n}\sum_{i=1}^{n}[\| X_i \| M_i + E(\| X_i \| M_i)]$ and $\| \cdot \|_2$ denotes the uniform norm over bivariate bounded real functions over $[0,1]^2$.

Since (6.70) is similar to (6.61) it is then possible to derive the following results.

THEOREM 6.14 *If X is an $ARC(1)$ that satisfies (L_α) and is such that $E\left(\| X_0 \|^2 M_0^2\right) < \infty$, then*

$$E \| c_n - c \|_2^2 = O\left(n^{-\frac{\alpha}{1+\alpha}}\right). \tag{6.71}$$

Proof

From Theorem 6.10 it follows that

$$\sup_{0 \le s,t \le 1} E(c_n(s,t) - c(s,t))^2 \le \frac{c}{n}, \tag{6.72}$$

where c is constant.

Then (6.70) entails

$$E \| c_n - c \|^2 \le 2EN_n^2 \nu_n^{-2\alpha} + 2\nu_n^2 \frac{c}{n}.$$

Noting that $EN_n^2 \le 4E[\| X_0 \| M_0 + E \| X_0 \| M_0]^2 < \infty$ and choosing $\nu_n = O\left(n^{\frac{1}{2+2\alpha}}\right)$, one obtains (6.71). ■

THEOREM 6.15 *If X is a bounded $ARC(1)$, then*

$$P(|c_n(s,t) - c(s,t)| \ge \eta) \le 2\exp\left(-\frac{n\eta^2}{\gamma + \delta\eta}\right),\ \eta > 0, \tag{6.73}$$

where $\gamma > 0$ and $\delta > 0$ depend only on ρ and P_X.

If in addition (L_α) holds, with M_0 bounded, then

$$P(\| c_n - c \|_2 \ge \eta) \le d\, \eta^{-2/\alpha} \exp\left(-\frac{n\eta^2}{4\gamma + 2\delta\eta}\right), \tag{6.74}$$

where $d = 4 \| X_0 \|_{\infty,C} \| M_0 \|_\infty$.

Proof

From Lemma 6.3 it follows that

$$\begin{aligned} C_n - C &= \frac{1}{n}\left[\sum_{j=0}^{n-1}(I + R + \ldots + R^j)(E_{n-j}) + (R + \ldots + R^n)(Z_0)\right] \\ &:= \frac{1}{n}\left[\sum_{j=0}^{n-1} F_{n,j} + F_{n,n}\right]. \end{aligned}$$

Now $F_{n,0}, \ldots, F_{n,n}$ is an $\mathcal{L}'$-valued martingale difference with respect to $\mathcal{B}_0, \ldots, \mathcal{B}_n$.

Hence $(c_n - c)(s,t) = \delta_{(s)}(C_n - C)(\delta_{(t)}$ is a real martingale and one may apply inequality (2.57) to it, which gives (6.73).

The proof of (6.74) is somewhat similar to that of (6.64) and is therefore omitted. ■

Again we leave to the reader similar results for $D_n - D$.

Sampled data

A usual situation in practice is to dispose of observations of X at discrete instants. Suppose that $\left(X_i\left(j\nu_n^{-1}\right),\ 1 \le j \le \nu_n,\ 1 \le i \le n\right)$ are data and that (L_α) holds. It is then possible to construct good approximations of $X_1, \ldots, X_n$ by interpolating these data.

In Chapter 9 we will consider **spline smoothing**. Here we only deal with **linear interpolation**, defined for each i as

$$
\begin{aligned}
X_i'(t) &= X_i\left(\nu_n^{-1}\right) \quad , \quad O \le t \le \nu_n^{-1}, \\
X_i'(t) &= X_i\left(j\nu_n^{-1}\right) + (\nu_n t - j)\left[X_i\left((j+1)\nu_n^{-1}\right) - X_i\left(j\nu_n^{-1}\right)\right],
\end{aligned}
\tag{6.75}
$$

$j\nu_n^{-1} \le t \le (j+1)\nu_n^{-1}$, $1 \le j \le \nu_n - 1$.

Then, using (L_α), one obtains

$$\| X_i - X_i' \| \le 2M_i \nu_n^{-\alpha} \quad , \quad 1 \le i \le n. \tag{6.76}$$

Thus, if $S_n' = X_1' + \ldots + X_n'$, we have

$$\| \frac{S_n}{n} - \frac{S_n'}{n} \| \le 2\frac{M_1 + \ldots + M_n}{n}\nu_n^{-\alpha}, \tag{6.77}$$

and hence Theorem 6.11 implies

$$E \| \frac{S_n'}{n} \|^2 = O(n^{-\frac{2\alpha}{2\alpha+1}} + \nu_n^{-2\alpha}) . \tag{6.78}$$

Now if M_0 is bounded (6.64) yields

$$P\left(\| \frac{S_n'}{n} \| \ge \eta\right) \le a\left(\frac{\eta}{2}\right)\exp\left(-nb\left(\frac{\eta}{2}\right)\right) \ , \ \eta > 0, \tag{6.79}$$

provided $\nu_n > \left(\dfrac{4 \| M_0 \|_\infty}{\eta}\right)^{1/\alpha}$.

Concerning the CLT we have

COROLLARY 6.3 *If X is an $ARC(1)$ and ε satisfies (L_1), then $\dfrac{S_n'}{\sqrt{n}}$ converges in distribution to the same Gaussian random variable as $\dfrac{S_n}{\sqrt{n}}$, provided $n^{1/2}\nu_n^{-1} \underset{n\to\infty}{\longrightarrow} 0$.*

Proof

From (6.75) it follows that

$$E \parallel \frac{S_n}{\sqrt{n}} - \frac{S'_n}{\sqrt{n}} \parallel \leq 2EM_0 n^{1/2} \nu_n^{-1} \underset{n\to\infty}{\longrightarrow} 0 \ ,$$

but X satisfies the CLT (Theorem 6.13), and the result follows by using (2.9).

■

We now consider the empirical covariance c'_n associated with $X'_1, \ldots, X'_n$, defined by

$$c'_n(s,t) = \frac{1}{n}\sum_{i=1}^{n} X'_i(s)X'_i(t).$$

Inequalities (6.69) entail

$$\parallel c_n - c'_n \parallel_2 \leq \frac{2}{n}\sum_{i=1}^{n} \parallel X_i \parallel M_i \nu_n^{-\alpha}. \tag{6.80}$$

Therefore, if $E\left(\parallel X_0 \parallel^2 M_0^2\right) < \infty$, we have

$$E \parallel c'_n - c \parallel_2^2 = O(n^{-\frac{\alpha}{1+\alpha}} + \nu_n^{-2\alpha}).$$

If X_0 and M_0 are bounded, (6.74) implies

$$P(\parallel c'_n - c \parallel_2 \geq \eta) \leq a\left(\frac{\eta}{2}\right) \exp\left(-nb\left(\frac{\eta}{2}\right)\right), \ \eta > 0, \tag{6.81}$$

for large enough n and provided $\nu_n \underset{n\to\infty}{\longrightarrow} \infty$.

6.7 Some applications to real continuous-time processes

In this section we provide some examples of applications to real processes that admit an autoregressive representation.

In the sequel $\xi = (\xi_t,\ t \in \mathbb{R})$ will be a measurable zero-mean real process with representation

$$X_n(t) = \xi_{n+t},\ 0 \leq t \leq 1,\ n \in \mathbb{Z},$$

and λ will denote Lebesgue measure over $[0,1]$. Recall that $\lambda \in (C[0,1])^*$.

COROLLARY 6.4 *If ξ has an $ARC(1)$ representation, then*

$$\frac{1}{T}\int_0^T \xi_t dt \underset{a.s.}{\longrightarrow} 0 \quad as\, T \to \infty. \tag{6.82}$$

Proof

Theorem 6.2 implies $\dfrac{X_0 + \ldots + X_{[T-1]}}{[T]} \to 0$ a.s., and consequently

$$\frac{1}{T}\int_0^{[T]} \xi_t dt = \frac{[T]}{T}\lambda\left(\frac{X_0 + \ldots + X_{[T-1]}}{[T]}\right) \to 0 \text{ a.s.}$$

On the other hand,

$$\left|\frac{1}{T}\int_{[T]}^{T} \xi_t dt\right| \leq \frac{\| X_{[T]} \|}{[T]} \cdot \frac{[T]}{T}$$

and the bound tends to 0 a.s., again by Theorem 6.2. Hence (6.82) holds. ■

COROLLARY 6.5 *If ξ has an $ARC(1)$ representation and*

(i) $E\left(\exp\left[\gamma \sup_{0\leq t\leq 1} \xi_t\right]\right) < \infty$ *for some* $\gamma > 0$,

(ii) $|\xi_t - \xi_s| \leq M_0|t-s|^{\alpha}$, $0 \leq s, t \leq 1$, *for some* $\alpha \in]0,1]$ *and for a bounded random variable* M_0.

Then

$$P\left(\left|\frac{1}{T}\int_0^T \xi_t dt\right| \geq \eta\right) \leq c_1'(\eta)\exp(-Tc_2'(\eta)), \quad T > 1, \tag{6.83}$$

where $c_1'(\eta)$ and $c_2'(\eta)$ are strictly positive constants.

Proof

First observe that

$$\begin{aligned} P\left(\left|\frac{1}{T}\int_{[T]}^{T} \xi_t dt\right| \geq \frac{\eta}{2}\right) &\leq P\left(\| X_{[T]} \| \geq T\frac{\eta}{2}\right) \\ &\leq \exp\left(-\gamma\frac{\eta}{2}T\right) E(\exp\gamma \| X_0 \|). \end{aligned}$$

Now, if $[T] \geq 1$,

$$\left|\frac{1}{T}\int_0^{[T]} \xi_t dt\right| \leq \left\| \frac{X_0 + \ldots + X_{[T-1]}}{[T]} \right\|$$

and from Theorem 6.12 we have

$$P\left(\left\| \frac{X_0 + \ldots + X_{[T-1]}}{[T]} \right\| \geq \frac{\eta}{2}\right) \leq a\left(\frac{\eta}{2}\right)\exp\left(-[T]b\left(\frac{\eta}{2}\right)\right).$$

Hence (6.83) holds. ■

The next statement deals with the CLT in continuous time.

COROLLARY 6.6 *If ξ has an $ARC(1)$ representation and*

$$|\xi_t - \xi_s| \le M_0|t-s| \quad ; \quad 0 \le s,t \le 1 \tag{6.84}$$

with $EM_0^2 < \infty$, then

$$\frac{1}{\sqrt{T}}\int_0^T \xi_t dt \xrightarrow[T\to\infty]{\mathcal{D}} N, \tag{6.85}$$

where N is a zero-mean Gaussian variable.

Proof

First observe that

$$\left|\frac{1}{\sqrt{T}}\int_{[T]}^T \xi_t dt\right| \le \frac{\| X_{[T]} \|}{\sqrt{[T]}},$$

and the bound tends to zero in probability.

Let us now prove that (X_n) satisfies the CLT. To this aim we write

$$\varepsilon_0(t) - \varepsilon_0(s) = (\xi_0(t) - \xi_0(s)) - E^{\mathcal{B}_0}(\xi_0(t) - \xi_0(s)).$$

Then from (6.84) and monotonicity of conditional expectation it follows that

$$|\varepsilon_0(t) - \varepsilon_0(s)| \le \left[M_0 + E^{\mathcal{B}_0}(M_0)\right] |t-s|,$$

$0 \le s,t \le 1$.

Recalling that (ε_n) is an i.i.d. sequence in $C[0,1]$, we may apply Theorem 2.8 and we conclude that (ε_n) satisfies the CLT. Now Theorem 6.6 gives the CLT for (X_n).

Finally, since

$$\frac{1}{\sqrt{[T]}}\int_0^{[T]} \xi_t dt = \lambda\left(\frac{X_0 + \ldots + X_{[T-1]}}{\sqrt{[T]}}\right),$$

(2.11) and (2.9) allow us to complete the proof. ■

A subsidiary calculus gives $EN^2 = 2\int_0^\infty E(\xi_0\xi_u)du$.

We finally give an LIL. Notation is as in Theorem 2.10.

COROLLARY 6.7 *If ξ is Gaussian and has an $ARC(1)$ representation, then*

$$\lim_{n\to\infty} d\left(\frac{1}{\sqrt{2n\mathrm{LogLog}n}}\int_0^n \xi_t dt, K'\right) = 0 \quad a.s. \tag{6.86}$$

and

$$c\left(\frac{1}{\sqrt{2n\mathrm{LogLog}n}}\int_0^n \xi_t dt\right) = K' \quad a.s., \tag{6.87}$$

where

$$K' = \left\{ \int_0^1 y(t)dt, \;\; y \in (I-\rho)^{-1}(K) \right\}.$$

Proof (sketch)

The Gaussian character of ξ implies that (ε_n) is a Gaussian process in $C[0,1]$; hence it satisfies the LIL in this space (cf. Ledoux-Talagrand (1991) p 217-218 and 288). It is then easy to derive (6.86) and (6.87) from (6.37) and (6.38) by using the relation

$$\frac{1}{\sqrt{2n \text{ Log Log } n}} \int_0^n \xi_t dt = \lambda \left((2n \text{ Log Log } n)^{-1/2} \sum_{i=0}^{n-1} X_i \right) \qquad \blacksquare$$

Example 6.5
Consider the Ornstein-Uhlenbeck process (Example 6.2) and take $h = 1$. Then the LIL holds and some calculations show that

$$K' = \left[-\frac{1}{\theta}, \frac{1}{\theta} \right].$$

Replacing λ by the Dirac measure $\delta_{(0)}$ one may check that

$$\lim_{n\to\infty} d\left(\frac{\xi_0 + \ldots + \xi_{n-1}}{\sqrt{2n\text{LogLog}n}}, K'' \right) = 0 \quad \text{a.s.}$$

and

$$c\left(\frac{\xi_0 + \ldots + \xi_{n-1}}{\sqrt{2n\text{LogLog}n}} \right) = K'' \quad \text{a.s.},$$

where

$$K'' = \left[-\frac{1}{\sqrt{2\theta}}, \; \frac{1}{\sqrt{2\theta}} \right].$$

This is the LIL for Gaussian $AR(1)$ processes.

Example 6.6
Suppose ξ is the stationary Gaussian process that satisfies equation (6.10). Then (6.86) and (6.87) hold.

In order to determine K', we write each φ_j that appears in (6.14) under the form

$$\varphi_j(t) = \sum_{\ell=1}^{k} c_{\ell j} e^{-r_\ell t} \;\; t \in \mathbb{R}, \; j = 0, \ldots, k-1,$$

where $c_{1j}, \ldots, c_{kj}$ are constants.

Then for $h = 1$ we have

$$K' = \left\{ \int_0^1 x(t)dt + \sum_{j=0}^{k-1} x^{(j)}(1) \sum_{\ell=1}^{k} \frac{c_{\ell_j}}{r_\ell}, \;\; x \in K \right\}.$$

Again replacing λ by $\delta_{(0)}$ we obtain results similar to those in Example 6.5, with

$$K'' = \left\{ x(0) + \sum_{j=0}^{k-1} x^{(j)}(1) \sum_{\ell=1}^{k} \frac{c_{\ell j}}{r_\ell} \,, \quad x \in K \right\}.$$

NOTES

6.1 Strong ARB processes appear in Pumo (1992), Mourid (1993), and Bosq (1996). Theorem 1 is an improvement of their existence results.

6.2 The examples come from Bosq (1996). Solution of (6.10) is derived in Ash-Gardner (1975).

6.3 Lemme 6.1 has been proved by Utev (1995). Theorems 6.3 and 6.4 are new. Theorems 6.5, 6.6 and 6.7 are in Bosq (1996). Details about nuclear operators over Banach spaces may be found in Vakhania et al. (1987).

6.4 Definition and results are new.

6.5 The representation lemma 6.3 is a novelty, as are the asymptotic results associated with weak topologies.

6.6 Results are either new or improvements of those in Pumo (1992).

6.7 These applications appear in Bosq (1996). Stoica (1992) has given LIL for the Ornstein-Uhlenbeck process.

7

General Linear Processes in Function Spaces

We now consider the linear process in all its generality. Recall that in the real case this model has the form

$$X_n = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{n-j}, \quad n \in \mathbb{Z}, \tag{7.1}$$

where $\mu \in \mathbb{R}$, (ε_n) is a white noise, $a_0 = 1$ and $\sum_{j=0}^{\infty} a_j^2 < \infty$.

In a Banach space B, $a_0 = I$ and each a_j becomes a bounded linear operator, thus

$$X_n = \mu + \sum_{j=0}^{\infty} a_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{7.2}$$

which suitable conditions concerning (a_j) (Section 1).

A crucial desirable property of such a process is **invertibility**. This topic is studied in Section 2. A linear process (X_n) is said to be invertible if it admits the decomposition

$$X_n - \mu = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j(X_{n-j} - \mu), \quad n \in \mathbb{Z}, \tag{7.3}$$

where (ρ_j) is a family of bounded linear operators.

If (7.3) holds, then (7.2) may be interpreted as the **Wold decomposition** of (X_n) in B.

In Section 3 it is shown that a linear process in a Hilbert space H has a

Markovian representation in a subspace of H^∞, say H_w, and that the obtained process (Y_n) is actually an $ARH_w(1)$. So (Y_n) satisfies the limit theorems and large deviation inequalities in Chapters 3 and 4 and, as by-products, similar properties hold for (X_n) (Section 4). General limit theorems for linear processes in Banach spaces are also pointed out in this Section. Section 5 is devoted to some technical proofs.

7.1 Existence and first properties of linear processes

Keeping the notation used in Chapter 6 we first state a simple lemma.

LEMMA 7.1 *Let $(U_n,\ n \in \mathbb{Z})$ be a sequence of B-random variables, where B is a separable Banach space, with $\sup_n E \| U_n \| < \infty$. Let $(a_j,\ j \geq 0)$ be a sequence of bounded linear operators from B to B. Then*

1) *If $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} < \infty$, $\sum_{j=0}^{\infty} a_j(U_{n-j})$ converges almost surely and in $L^1_B(\Omega, \mathcal{A}, P)$.*

2) *If B has type p $(1 < p \leq 2)$, $\sum_{j=0}^{\infty} \| a_j \|^p_{\mathcal{L}} < \infty$ and (U_n) is an independent sequence such that $\sup_n E \| U_n \|^p < \infty$, then $\sum_{j=0}^{\infty} a_j(U_{n-j})$ converges almost surely and in $L^p_B(\Omega, \mathcal{A}, P)$.*

Proof

1) By monotone convergence one gets

$$
\begin{aligned}
E\left(\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} \| U_{n-j} \|\right) &= \lim_{n\uparrow\infty} \uparrow E\left(\sum_{j=0}^{n} \| a_j \|_{\mathcal{L}} \| U_{n-j} \|\right) \\
&\leq \lim_{n\uparrow\infty} \left(\sum_{j=0}^{n} \| a_j \|_{\mathcal{L}}\right) \cdot \sup_k E \| U_k \| \\
&< \infty.
\end{aligned}
$$

It follows that $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} \| U_{n-j} \| < \infty$ a.s.; hence $\sum_{j=0}^{\infty} a_j(U_{n-j})$ converges absolutely, with probability 1.

On the other hand,

$$E \parallel \sum_{j=q_1}^{q_2} a_j(U_{n-j}) \parallel \leq \left(\sum_{j=q_1}^{q_2} \parallel a_j \parallel_{\mathcal{L}} \right) \sup_k E \parallel U_k \parallel; \quad q_1, q_2 \in \mathbb{N}, \ q_1 < q_2;$$

and the Cauchy criterion yields convergence in $L^1_B(\Omega, \mathcal{A}, P)$.

2) If B has type p, there exists a constant c_p such that

$$E \parallel \sum_{j=q_1}^{q_2} a_j(U_{n-j}) \parallel^p \leq c_p \left(\sum_{j=q_1}^{q_2} \parallel a_j \parallel^p_{\mathcal{L}} \right) \sup_k E \parallel U_k \parallel^p,$$

which gives convergence in $L^p_B(\Omega, \mathcal{A}, P)$. The GIN Theorem 2.1 entails almost sure convergence. ■

We now may define linear processes in B.

DEFINITION 7.1 *Let $X = (X_n, \ n \in \mathbb{Z})$ be a B-random process. It is said to be a* **linear process** *in B (LPB) if*

$$X_n = \mu + \sum_{j=0}^{\infty} a_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{7.4}$$

where $\mu \in B$, $a_0 = I$, (a_j) is a sequence in $\mathcal{L}$, and (ε_n) is a weak white noise. The series in (7.4) converges in probability.

Lemma 7.1 provides sufficient conditions for existence of (X_n): in a general (separable) Banach space $\sum_{j=0}^{\infty} \parallel a_j \parallel_{\mathcal{L}} < \infty$ implies convergence in L^2_B and a.s.

If (ε_n) is a strong white noise and B has type 2 (in particular, B is a Hilbert space), one obtains convergence in L^2_B and a.s. as soon as $\sum_{j=0}^{\infty} \parallel a_j \parallel^2_{\mathcal{L}} < \infty$.

In the sequel we will say that X is **standard** if $\mu = 0$, (ε_n) is a strong white noise, and $\sum_{j=0}^{\infty} \parallel a_j \parallel_{\mathcal{L}} < \infty$.

An LPB has the following properties:

1) If B has type 2 and $\sum_{j=0}^{\infty} \parallel a_j \parallel^2_{\mathcal{L}} < \infty$, then X is weakly stationary.

2) If (ε_n) is a strong white noise then X is strictly stationary.

3) If $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} < \infty$ then X_n is integrable and

$$E\, X_n = \mu, \quad n \in \mathbb{Z}. \tag{7.5}$$

4) Let X be a standard LPB. Then

$$y^* C_{X_n, X_{n+h}}(x^*) = \sum_{j=0}^{\infty} y^* \left(a_{j+h} C_{\varepsilon_0} a_j^* \right) (x^*); \tag{7.6}$$

$x^*,\ y^* \in B^*;\ h \in \mathbb{Z}$.

7.2 Invertibility of linear processes

We now study conditions that ensure invertibility of linear processes in a separable Hilbert space (LPH).

In the remainder of the text we will use the complex operator

$$A(z) = \sum_{j=0}^{\infty} z^j a_j \tag{7.7}$$

for all z such that the series converges absolutely in the $\| \cdot \|_{\mathcal{L}}$ sense. Note that $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} < \infty$ yields convergence at least if $|z| \leq 1$.

If $e = (e_i,\ i \geq 1)$ is an arbitrary orthonormal basis of H, one may associate to A the complex matrix

$$(A^e(z))_{\ell, \ell'} = \sum_{j=0}^{\infty} z^j < a_j(e_\ell),\ e_{\ell'} >;\ \ell, \ell' \geq 1. \tag{7.8}$$

The following statement deals with the finite dimensional case.

THEOREM 7.1 *Let X be a standard LPH, where H is finite dimensional.*

If C_{ε_0} is invertible and $\det A^e(z)$ *has no root on the unit circle, then X is invertible:*

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j(X_{n-j}), \quad n \in \mathbb{Z}, \tag{7.9}$$

where $(\rho_j) \subset \mathcal{L}$ and $\sum_{j=1}^{\infty} \| \rho_j \|_{\mathcal{L}} < \infty$.

The series converges almost surely and in $L^2_H(\Omega, \mathcal{A}, P)$.

Proof See Nsiri and Roy (1993). ∎

Note that the roots of $\det A^e(z)$ do not depend on the orthonormal basis e.

For an extension to the infinite-dimensional case some notation is needed. Here we again deal with standard LPH.

We first write the spectral decomposition (1.58) of C_{ε_0} under the form

$$C_{\varepsilon_0} = \sum_{j=1}^{\infty} \gamma_j g_j \otimes g_j, \tag{7.10}$$

with

$$< g_j, g_{j'} >= \delta_{jj'} \text{ and } \gamma_1 \geq \gamma_2 \geq \ldots \geq 0.$$

Now set $E_k = \text{sp}(g_1, \ldots, g_k)$, $k \geq 1$, denote by P^k the orthogonal projection of E_k and consider the E_k-valued process

$$Y_{n,k} = P^k(X_n) - Z_{n,k}, \quad n \in \mathbb{Z}, \tag{7.11}$$

where

$$Z_{n,k} = P^k \left(\sum_{j=1}^{\infty} a_j(\varepsilon_{n-j}) - \sum_{j=1}^{\infty} a_j P^k(\varepsilon_{n-j}) \right). \tag{7.12}$$

It is easy to see that $(Y_{n,k},\ n \in \mathbb{Z})$ has the representation

$$Y_{n,k} = P^k(\varepsilon_n) + \sum_{j=1}^{\infty} \left(P^k a_j\right) P^k(\varepsilon_{n-j}),\ n \in \mathbb{Z}, \tag{7.13}$$

which defines a linear process over E_k provided $P^k(\varepsilon_0)$ is not degenerate.

Notice that if P^k and a_j commute for all j, we simply have $Y_{n,k} = P^k(X_n)$.

Now let A_k be the complex operator associated with $Y_{n,k}$ (see (7.7) and A_k^g its matrix form corresponding to the basis $g = (g_j,\ j \geq 1)$. We have

$$A_k(z) = \sum_{j=0}^{\infty} z^j P^k(a_j)$$

and

$$\begin{aligned}
(A_k^g(z))_{\ell,\ell'} &= \sum_{j=0}^{\infty} z^j < P^k(a_j)\ (g_\ell),\ g_{\ell'} > \\
&= \sum_{j=0}^{\infty} z^j < a_j\ (g_\ell),\ g_{\ell'} >,
\end{aligned}$$

$1 \leq \ell,\ \ell' \leq k$.

We may now state

LEMMA 7.2 *If X is a standard LPH, $C_{P^k(\varepsilon_0)}$ is invertible over E_k and there exists k_0 such that $k \geq k_0$ implies* $\det A_k^g(z) \neq 0$ *for $|z| \leq 1$, then for each $k \geq k_0$*

$$Y_{n,k} = P^k(\varepsilon_n) + \sum_{j=1}^{\infty} \rho_{j,k}(Y_{n-j,k}),\ n \in \mathbb{Z}, \tag{7.14}$$

where $(\rho_{j,k},\ j \geq 1)$ is a sequence in $\mathcal{L}(E_k,\ E_k)$ such that

$$\sum_{j=1}^{\infty} \| \rho_{j,k} \|_{\mathcal{L}(E_k, E_k)} < \infty. \tag{7.15}$$

It is easy to see that the linear operators $(\rho_{j,k},\ j \geq 1)$ from E_k to E_k are defined recursively by

$$\begin{cases} \rho_{1,k} - a_{1,k} = 0 \\ \rho_{j,k} - a_{j,k} + \displaystyle\sum_{p=1}^{j-1} \rho_{p,k} a_{j-p} = 0, \quad j \geq 2. \end{cases} \tag{7.16}$$

Proof

Apply Theorem 7.1 with $H = E_k$. ■

We now give a general invertibility result due to F. Merlevède.

THEOREM 7.2 *Suppose that X is a standard LPH that satisfies the conditions in Lemma 7.2, and such that*

$$1 - \sum_{j=1}^{\infty} z^j \| a_j \|_{\mathcal{L}} \neq 0 \ \textit{for}\ |z| \leq 1. \tag{7.17}$$

Then X is invertible:

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j(X_{n-j}), \quad n \in \mathbb{Z}, \tag{7.18}$$

where $(\rho_j) \subset \mathcal{L}$, $\sum_{j=1}^{\infty} \| \rho_j \|_{\mathcal{L}} < \infty$, and the series converges in $L^2_H(\Omega, \mathcal{A}, P)$ and with probability 1.

Proof See Section 5. ■

Example 7.1
Consider an $ARH(1)$ process (see Chapter 3) written in its linear form

$$X_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z},$$

where ρ is a symmetric compact operator with spectral decomposition

$$\rho = \sum_{j=1}^{\infty} \alpha_j g_j \otimes g_j,$$

where $|\alpha_1| < 1$ and (g_j) is as defined in (7.10).

Here $\det A_k^g(z) = \prod_{\ell=1}^{k} \left[\sum_{j=1}^{\infty} (z\alpha_\ell)^j \right]$ does not vanish if $|z| \leq 1$. Then, if $\gamma_j > 0$ for each j, one may apply Lemma 7.2 and $(Y_{n,k})$ is invertible for every k. Of course Theorem 7.2 also applies : if $|\alpha_1| < \frac{1}{2}$ we have

$$1 - \sum_{j=1}^{\infty} z^j \parallel \rho^j \parallel_{\mathcal{L}} = \frac{1 - 2|\alpha_1|z}{1 - |\alpha_1|z} \neq 0 \text{ for } |z| < \frac{1}{2|\alpha_1|}.$$

However, this is less interesting, since a direct computation gives

$$X_n = \varepsilon_n + \rho(X_{n-1}), \quad n \in \mathbb{Z}.$$

Example 7.2
An ***ARMAH*(p, q) process** is an H-valued strictly stationary process (X_n) such that

$$\sum_{j=0}^{p} c_j(X_{n-j}) = \sum_{j=0}^{q} d_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}, \tag{7.19}$$

where $c_0 = d_0 = I$, $(c_j) \subset \mathcal{L}$, $(d_j) \subset \mathcal{L}$, $E \parallel c_p(X_0) \parallel^2 > 0$, $E \parallel d_q(\varepsilon_0) \parallel^2 > 0$, and (ε_n) is a strong Hilbertian white noise.

Putting

$$Z_n = \sum_{j=0}^{p} c_j(X_{n-j}), \quad n \in \mathbb{Z}, \tag{7.20}$$

one obtains an LPH process that is in fact a Hilbertian moving average of order q.

Now Lemma 7.2 can be applied to (Z_n), provided that $\gamma_j > 0$ for each j and

$$\det \left(\sum_{j=0}^{q} z^j < d_j(g_\ell), \, g_{\ell'} > \right)_{1 \leq \ell, \ell' \leq k} \neq 0 \text{ for } |z| \leq 1. \tag{7.21}$$

Theorem 7.2 works for (Z_n) provided that (7.21) holds for all k and

$$1 - \sum_{j=1}^{q} z^j \parallel d_j \parallel_{\mathcal{L}} \neq 0 \text{ if } |z| \leq 1. \tag{7.22}$$

Then we have

$$Z_n = \varepsilon_n + \sum_{j=1}^{\infty} f_j(Z_{n-j}), \quad n \in \mathbb{Z},$$

where $\sum_{j=1}^{\infty} \| f_j \|_{\mathcal{L}} < \infty$, and using (7.20) one obtains

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j(X_{n-j}), \quad n \in \mathbb{Z},$$

with $\sum_{j=1}^{\infty} \| \rho_j \|_{\mathcal{L}} < \infty$.

7.3 Markovian representations of LPH: applications

In order to construct Markovian representations of a linear process we consider the cartesian product H^∞ of a countable number of copies of H and a sequence $w = (w_i, \ i \geq 1)$ of strictly positive weights such that $\sum_{i=1}^{\infty} w_i < \infty$.

Let H_w be the subspace of H^∞ consisting of those $x = (x_i, \ i \geq 1)$ such that

$$\sum_{i=1}^{\infty} w_i \| x_i \|^2 < \infty.$$

Set

$$< x, y >_w = \sum_{i=1}^{\infty} w_i < x_i, y_i > \ ; \quad x, y \in H_w.$$

Then H_w equipped with $< \cdot, \cdot >_w$ becomes a separable Hilbert space. Below; $\mathcal{L}_w$ will denote the class of bounded linear operators from H_w to H_w.

Now let X be an invertible LPH with $E \| X_0 \|^2 < \infty$. For each n in $\mathbb{Z}$ set

$$Y_n = (X_n, X_{n-1}, X_{n-2}, \ldots)$$

and

$$e_n = (\varepsilon_n, 0, 0, \ldots)$$

then Y_n defines an H_w-valued random variable. Actually

$$E\left(\sum_{i=1}^{\infty} w_i \| X_{n-i+1} \|^2\right) = E \| X_0 \|^2 \sum_{i=1}^{\infty} w_i < \infty.$$

Thus $Y_n \in H_w$ (a.s.) and measurability is clear. Similarly, e_n is an H_w-random variable.

Next define a matrix operator over H_w by setting

$$R = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & & \\ I_H & O_H & \cdots & & \\ O_H & I_H & & & \\ & O_H & I_H & & \\ & & \ddots & \ddots & \end{bmatrix},$$

where O_H (resp. I_H) denotes the null operator (resp. identity) of H.

In fact, $R = (R_{\ell,\ell'})$, with

$$\begin{aligned} R_{\ell'+1,\ell'} &= I_H, \ \ell' \geq 1 \\ R_{\ell,\ell'} &= O_H, \ \ell \notin \{1, \ \ell'+1\}. \end{aligned}$$

The following lemma furnishes a condition for R to be bounded.

LEMMA 7.3 *Suppose that*

$$w_{i+1} \leq c w_i, \quad i \geq 1, \ \textit{where } c \textit{ is a constant} \tag{7.23}$$

and

$$\sum_{i=1}^{\infty} w_i^{-1} \parallel \rho_i \parallel_{\mathcal{L}}^2 < \infty. \tag{7.24}$$

Then R is in $\mathcal{L}_w$.

Proof

Observe that

$$\parallel R(x) \parallel_w^2 = w_1 \sum_{i=1}^{\infty} \parallel \rho_i(x_i) \parallel^2 + \sum_{i=2}^{\infty} w_i \parallel x_{i-1} \parallel^2 .$$

Using the Cauchy-Schwarz inequality we get

$$\parallel R(x) \parallel_w^2 \leq w_1 \left(\sum_{i=1}^{\infty} w_i^{-1} \parallel \rho_i \parallel_{\mathcal{L}}^2 \right) \left(\sum_{i=1}^{\infty} w_i \parallel x_i \parallel^2 \right) + \sum_{i=1}^{\infty} w_{i+1} \parallel x_i \parallel^2$$

and from (7.23) it follows that

$$\parallel R(x) \parallel_w^2 \leq \left(c + w_1 \sum_{i=1}^{\infty} w_i^{-1} \parallel \rho_i \parallel_{\mathcal{L}}^2 \right) \parallel x \parallel_w^2 .$$

Hence $R \in \mathcal{L}_w$. ■

Concerning validity of (7.24) notice that the proof of Theorem 7.2 shows that $\rho_0 = I$, $\rho_1 = a_1$, and

$$\rho_j = a_j + \sum_{j=1}^{p-1} \rho_p a_{j-p}, \quad j \geq 2. \tag{7.25}$$

From this it is easy to see that if

$$1 - \sum_{k=1}^{\infty} z^k w_k^{1/2} \parallel a_k \parallel_{\mathcal{L}} \neq 0 \ \text{ for } \ |z| \leq 1, \tag{7.26}$$

then $\displaystyle\sum_{k=1}^{\infty} w_k^{-1/2} \parallel \rho_k \parallel_{\mathcal{L}} < \infty$ and (7.24) holds.

On the other hand, if $\parallel \rho_{i+1} \parallel_{\mathcal{L}} \leq c \parallel \rho_i \parallel_{\mathcal{L}}$ $(c > 0)$ and $\displaystyle\sum_i \parallel \rho_i \parallel_{\mathcal{L}} < \infty$, then the choice $w_i = \parallel \rho_i \parallel_{\mathcal{L}}$ entails (7.24).

Finally, if $\displaystyle\sum_{k=1}^{\infty} w_k^{-1/2} \parallel a_k \parallel_{\mathcal{L}} < 1$ then (7.26) holds. It is possible to force this condition by replacing w_k by dw_k, where d is a suitable constant, provided $\displaystyle\sum_{k=1}^{\infty} w_k^{-1/2} \parallel a_k \parallel_{\mathcal{L}} < \infty$.

Now we have a representation theorem due to F. Merlevède.

THEOREM 7.3 *If X is a standard invertible LPH and if*

$$\sum_{i=1}^{\infty} w_i^{-1} \parallel \rho_i \parallel_{\mathcal{L}}^2 < \infty,$$

then

$$Y_n = R(Y_{n-1}) + e_n, \quad n \in \mathbb{Z}, \tag{7.27}$$

where $E \in \mathcal{L}_w$ and (e_n) is a strong H-white noise.

Proof: clear. ■

From (7.27) follows

COROLLARY 7.1 *Suppose that (7.26) holds and $w_j = s^j$, $j \geq 1$ $(0 < s < 1)$. Then*

$$Y_n = \sum_{j=0}^{\infty} R^j(e_{n-j}), \quad n \in \mathbb{Z}, \tag{7.28}$$

in $L^2_{H_w}(\Omega, \mathcal{A}, P)$ and a.s.
Moreover,

$$\parallel R^j \parallel_{\mathcal{L}_w} \leq a\, b^j, \quad j \geq 1, \tag{7.29}$$

where $a > 0$ and $0 < b < 1$.

Proof: See Section 5. ■

From here on we will say that X is a **standard invertible *LPH*** if (7.27) and (7.29) hold.

7.4 Limit theorems for *LPB* and *LPH*

The autoregressive representation of *LPH* allows us to use the limit theorems in Chapter 3 to derive similar results for *LPH*.

LEMMA 7.4 *A standard invertible LPH satisfies*

$$|E < X_0,\ X_h > | \leq ab^h E \parallel X_0 \parallel^2, \ \ h > 0, \tag{7.30}$$

where $a > 0$ *and* $0 < b < 1$ *depends only on* P_X *and* w.

Proof

Lemma 3.2 applied to (Y_n) gives

$$|E(< Y_0, Y_h >_w)| \leq \parallel R^h \parallel_{\mathcal{L}_w} E \parallel Y_0; \parallel_w^2$$

thus

$$\left| E\left(\sum_{i=1}^{\infty} w_i < X_{-i}, X_{h-i} >\right)\right| \leq \parallel R^h \parallel_{\mathcal{L}_w} E\left(\sum_{i=1}^{\infty} w_i \parallel X_{-i} \parallel^2\right).$$

By stationarity one obtains

$$\left(\sum_{i=1}^{\infty} w_i\right) |E < X_0, X_h > | \leq \parallel R^h \parallel_{\mathcal{L}_w} \left(\sum_{i=1}^{\infty} w_i\right) E \parallel X_0 \parallel^2$$

and (7.29) implies (7.30). ■

Laws of large numbers follow from Lemma 7.4.

THEOREM 7.4 *Let X be a standard invertible LPH. Then*

$$nE \parallel \overline{X}_n \parallel^2 \longrightarrow \sum_{h=-\infty}^{+\infty} E < X_0, X_h > \tag{7.31}$$

and

$$\frac{n^{1/4}}{(\text{Log} n)^{\beta}} \parallel \overline{X}_n \parallel \longrightarrow 0 \ \ a.s., \ \ \beta > \frac{1}{2}. \tag{7.32}$$

Proof

Apply Theorems 3.7 and 3.8 to (Y_n) and use stationarity and the definition of $< \cdot, \cdot >_w$. ■

Concerning exponential rates we have the following.

THEOREM 7.5 *Suppose that* $\sum_{i=1}^{\infty} w_i^{1/2} < \infty$ *and consider* X, *a standard invertible LPH such that* $E(\exp \gamma \parallel X_0 \parallel) < \infty$ *for some* $\gamma > 0$. *Then*

$$P(\parallel \overline{X}_n \parallel \geq \eta) \leq 4 \exp\left(-\frac{n\eta^2}{\alpha + \beta\eta}\right), \quad \eta > 0, \tag{7.33}$$

where $\alpha > 0$ *and* $\beta > 0$ *depend only on* P_X *and* w.

Proof

Note first that

$$\parallel Y_0 \parallel_w = \left(\sum_{i=1}^{\infty} w_i \parallel X_{-i} \parallel^2\right)^{1/2} \leq \sum_{i=1}^{\infty} w_i^{1/2} \parallel X_{-i} \parallel .$$

Now convexity of the exponential gives

$$\exp\left(\gamma \sum_1^p w_i^{1/2} \parallel X_{-i} \parallel / \sum_1^p w_i^{1/2}\right) \leq \frac{\sum_1^p w_i^{1/2} \exp(\gamma \parallel X_{-i} \parallel)}{\sum_1^p w_i^{1/2}}.$$

Therefore

$$E\left(\exp \gamma \left(\sum_1^{\infty} w_i^{1/2}\right)^{-1} \sum_1^p w_i^{1/2} \parallel X_{-i} \parallel\right) \leq E(\exp \gamma \parallel X_0 \parallel) < \infty,$$

which implies

$$E\left(\exp \gamma \left(\sum_1^{\infty} w_i^{1/2}\right)^{-1} \parallel Y_0 \parallel_w\right) < \infty.$$

We are now in a position to apply Theorem 3.9 to (Y_n):

$$P(\parallel \overline{Y}_n \parallel_w \geq \eta) \leq 4 \exp\left(-\frac{n\eta^2}{\alpha_0 + \beta_0\eta}\right), \quad \eta > 0,$$

that is,

$$P\left(\sum_{i=1}^{\infty} w_i \parallel (\overline{Y}_n)_i \parallel^2 \geq \eta^2\right) \leq 4\exp\left(-\frac{n\eta^2}{\alpha_0 + \beta_0 \eta}\right).$$

But $\overline{X}_n = (\overline{Y}_n)_1$ and

$$\parallel \overline{X}_n \parallel \geq \eta \Rightarrow \sum_{i=1}^{\infty} w_i \parallel (\overline{Y}_n)_i \parallel^2 \geq \eta^2 w_1;$$

thus

$$P(\parallel \overline{X}_n \parallel \geq \eta) \leq \exp\left(-\frac{n\eta^2 w_1}{\alpha_0 + \beta_0 \eta w_1^{1/2}}\right),$$

which gives (7.33). ■

COROLLARY 7.2

$$\parallel \overline{X}_n \parallel = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \tag{7.34}$$

Proof clear. ■

We now give the *CLT*.

THEOREM 7.6 *If X is a standard invertible LPH then, for each $k \geq 1$,*

$$\sqrt{n}\sum_{i=1}^{n}(X_i,\ X_{i+1}, \ldots, X_{i+k-1}) \xrightarrow{\mathcal{D}} N_k, \tag{7.35}$$

where N_k is a zero-mean Gaussian H^k-valued random variable.

Proof

Applying Theorem 3.10 to (Y_n) we obtain

$$\sqrt{n}\ \overline{Y}_n \xrightarrow{\mathcal{D}} N \sim \mathcal{N}(0, \Gamma),$$

where $\Gamma = (I - R)^{-1} C_{e_0} (I - R^*)^{-1}$.

Now $p_k : (x_i,\ i \geq 1) \mapsto (x_1, \ldots, x_k)$ is a continuous linear operator from H_w to H^k; thus

$$\sqrt{n} p_k(\overline{Y}_n) \xrightarrow{\mathcal{D}} p_k(N),$$

and hence (7.35) with $C_{N_k} = p_k \Gamma p_k^*$. ■

Let us now give some indications concerning **estimation of covariance operators**. Autoregressive representation of (Y_n) allows us to derive asymptotic properties of associated empirical operators from results in Chapter 4. But since $Y_1, \dots, Y_n$ cannot be completely observed, these results are not directly usable.

For convenience we suppose that data are $X_{-k+2}, \dots, X_0, \dots, X_n$ and set

$$Z_{ik} = (X_i,\ X_{i-1}, \dots, X_{i-k+1}),\ \ 1 \leq i \leq n.$$

Then $Z_{1k}, \dots, Z_{nk}$ are observed and may be written under the form $p^k(Y_1), \dots, p^k(Y_n)$, where p^k is the orthogonal projector of the subspace of H_w defined as $H^k \times \{0\} \times \{0\} \times \dots$.

The associated process $Z^{(k)} = (Z_{ik},\ i \in \mathbb{Z})$ has covariance operator

$$C_{Z^{(k)}} = p^k C_Y p^k \tag{7.36}$$

and empirical covariance operator given by

$$C_{n,Z^{(k)}} = p^k C_{n,Y} p^k, \tag{7.37}$$

where C_Y and $C_{n,Y}$ denote the corresponding operators associated with Y.

Thus we have

$$C_{n,Z^{(k)}} - C_{Z^{(k)}} = p^k [C_{n,Y} - C_Y] p^k. \tag{7.38}$$

Clearly this allows to derive asymptotic results concerning $Z^{(k)}$ by using results in Chapter 4 and boundedness of p^k. In particular, observe that

$$\| C_{n,Z^{(k)}} - C_{Z^{(k)}} \|_{\mathcal{S}_{H_w}} \leq \| C_{n,Y} - C_Y \|_{\mathcal{S}_{H_w}}, \tag{7.39}$$

where $\| \cdot \|_{\mathcal{S}_{H_w}}$ denotes Hilbert-Schmidt norm.

Finally, the special case $k = 1$ provides results for X.

Similar properties hold for cross-covariance operators. Details are straightforward and therefore omitted.

General limit theorems for linear processes in Banach spaces

Limit theorems for LP can be obtained in a more general setting than above. In particular, one may remove the invertibility condition and consider the more general case of an LPB.

We now indicate without proof two results in this direction. We first state a strong law of large numbers.

THEOREM 7.7 *Let X be an LPB, where B has type p. Suppose that (ε_n) is a strong white noise, $E \| X_0 \|^p < \infty$, and $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}}^p < \infty$. Then*

$$\left\| \frac{S_n}{n} - \mu \right\| \to 0 \quad a.s.$$

The next assertion is a *CLT*.

THEOREM 7.8 *Let X be an LPB, where B has type 2. If (ε_n) is a strong white noise, $E \parallel X_0 \parallel^2 < \infty$, and $\sum_{j=0}^{\infty} \parallel a_j \parallel_{\mathcal{L}} < \infty$, then*

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{\mathcal{D}} N,$$

where N is a zero-mean Gaussian B-random variable.

COROLLARY 7.3 *If H is a Hilbert space and $\mu = 0$, then N has covariance operator*

$$\Gamma_N = A\, C_{\varepsilon_0} A^*,$$

where

$$A = \sum_{j=0}^{\infty} a_j.$$

It can be shown that the condition $\sum_{j=0}^{\infty} \parallel a_j \parallel_{\mathcal{L}} < \infty$ cannot be replaced by $\sum_{j=0}^{\infty} \parallel a_j \parallel_{\mathcal{L}}^2 < \infty$ (see Merlevède, Peligrad, and Utev (1997)).

On the other hand, Theorems 7.7 and 7.8 extend to noncausal linear processes, that is, processes of the form

$$X_n = \sum_{j=-\infty}^{+\infty} a_j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}.$$

7.5 * Derivation of invertibility

This section is devoted to the proofs of Theorems 7.2 and Corollary 7.1.

Proof of Theorem 7.2

We first state the following technical lemma:

LEMMA 7.5 *Under (7.17) we have*

$$\sum_{j \geq 1} \|\rho_{j,k}\|_{\mathcal{L}} < \infty \textit{ for all } k \geq 1, \tag{7.40}$$

where the $\rho_{j,k}$ satisfy (7.16).

Proof of Lemma

Let us first introduce for $j \geq 1$ the positive real numbers c_j defined by

$$\begin{cases} c_1 & = \|a_1\|_{\mathcal{L}} \\ c_j & = \|a_j\|_{\mathcal{L}(H)} + \sum_{p=1}^{j-1} c_p \|a_{j-p}\|_{\mathcal{L}},\ j \geq 2. \end{cases} \tag{7.41}$$

Now from (7.16) we get, for $j \geq 1$ and $k \geq 1$, $\|\rho_{j,k}\|_{\mathcal{L}} \leq c_j$; therefore

$$\sum_{j\geq 1} \|\rho_{j,k}\|_{\mathcal{L}} \leq \sum_{j\geq 1} c_j,$$

but (7.17) implies $\sum_{j\geq 1} c_j < \infty$ and hence the desired result. ■

We now return to the proof of Theorem 7.2. Set $r := \sum_{j\geq 1} \|a_j\|_{\mathcal{L}} < \infty$. Since the conditions of Lemma 7.2 are satisfied for all $k \geq k_0$, we have

$$Y_{0,k} = P^k(\varepsilon_0) + \sum_{j\geq 1} \rho_{j,k}(Y_{-j,k}).$$

Now note that, for all $x \in H$, $\|(I - P^k)x\| \to 0, k \to \infty$,; hence we obtain

$$\|\varepsilon_0 - P^k(\varepsilon_0)\| \to 0 \tag{7.42}$$

and

$$\|X_0 - P^k(X_0)\| \to 0. \tag{7.43}$$

On the other hand,

$$\mathbb{E} \parallel Z_{0,k} \parallel^2 \leq r^2 \sum_{m\geq k+1} \gamma_m,$$

and since $\sum_{m\geq 1} \gamma_m < \infty$ we get

$$\mathbb{E}\|Z_{0,k}\|^2 \underset{k\to\infty}{\longrightarrow} 0\ . \tag{7.44}$$

Combining (7.43) and (7.44) we obtain

$$Y_{0,k} \to X_0 \text{ in } L^2(\Omega, \mathcal{A}, P). \tag{7.45}$$

Thus, the proof will be complete if we prove mean square convergence of $\sum_{j\geq 1} \rho_{j,k}(Y_{-j,k})$ to $\sum_{j\geq 1} \rho_j(X_{-j})$, as $k \to \infty$. To do this, we first recursively define the linear operators b_j from H to H by

$$\begin{cases} b_1 - a_1 = 0, \\ b_j - a_j + \sum_{p=1}^{j-1} b_p a_{j-p} = 0, \ j \geq 2. \end{cases} \tag{7.46}$$

Now observe that

$$\sum_{j\geq 1} b_j(X_{-j}) - \sum_{j\geq 1} \rho_{j,k}(Y_{-j,k}) = \sum_{j\geq 1}(b_j - \rho_{j,k}P^k)(X_{-j}) + \sum_{j\geq 1} \rho_{j,k}(Z_{-j,k}).$$

First, one easily gets

$$\begin{aligned} E\Big\| \sum_{j\geq 1} \rho_{j,k}(Z_{-j,k})\Big\|^2 &= E\Big\| \sum_{j\geq 1} \rho_{j,k}P^k\Big(\sum_{i\geq 1} a_j(I - P^k)(\varepsilon_{-j-i})\Big)\Big\|^2 \\ &\leq r^2\Big(\sum_{j\geq 1} \|\rho_{j,k}P^k\|_{\mathcal{L}}\Big)^2\Big(\sum_{m=k+1}^{\infty} \gamma_m\Big). \end{aligned}$$

By Lemma 7.5 and (7.17) we have $\sum_{j\geq 1} \|\rho_{j,k}P^k\|_{\mathcal{L}} < \infty$ for all k; thus

$$E \| \sum_{j\geq 1} \rho_{j,k}(Z_{-j,k}) \|^2 \to 0 \ , \ k \to \infty. \tag{7.47}$$

Now let us prove that $E \| \sum_{j\geq 1}(b_j - \rho_{j,k}P^k)(X_{-j}) \|^2 \to 0$.

By using (7.46) and (7.16) together with the fact that X is a linear process, we easily obtain the following bound:

$$\begin{aligned} \Big\| \sum_{j\geq 1}(b_j - \rho_{j,k}P^k)X_{-j}\Big\| &\leq \Big\| \sum_{j\geq 1}(-a_j + P^k a_j P^k)(\varepsilon_{-j})\Big\| \\ &\quad + \Big\| \sum_{j\geq 2}\sum_{i=0}^{j-1}(\rho_{j-i,k}P^k a_i)(I - P^k)(\varepsilon_{-j})\Big\|. \end{aligned}$$

Now, note that

$$\begin{aligned} E & \quad \Big\| \sum_{j\geq 1}(-a_j + P^k a_j P^k)(\varepsilon_{-j})\Big\|^2 \\ = & \quad E\Big\| \sum_{j\geq 1} \Big((I - P^k)a_j - P^k a_j(I - P^k)\Big)(\varepsilon_{-j})\Big\|^2 \\ \leq & \quad E\Big\| \sum_{j\geq 1} -(I - P^k)a_j(\varepsilon_{-j})\Big\|^2 + E\Big\| \sum_{j\geq 1} P^k a_j(I - P^k)(\varepsilon_{-j})\Big\|^2 \\ & \quad + 2E\Big\| \sum_{j\geq 1} -(I - P^k)a_j(\varepsilon_{-j})\Big\|\Big\| \sum_{j\geq 1} P^k a_j(I - P^k)(\varepsilon_{-j})\Big\|. \end{aligned}$$

Thus, from the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} E\Big\|\sum_{j\geq 1}(-a_j + P^k a_j P^k)(\varepsilon_{-j})\Big\|^2 &\leq r^2 \sum_{m=k+1}^{\infty} \gamma_m + \sum_{m=k+1}^{\infty} E < X_0, g_m >^2 \\ &+ 2r\Big(\sum_{m=k+1}^{\infty} \gamma_m\Big)^{1/2}\Big(\sum_{m=k+1}^{\infty} E < X_0, g_m >^2\Big)^{1/2}. \end{aligned}$$

But since each term converges to 0 as $k \to \infty$, we deduce that

$$E \parallel \sum_{j\geq 1}(-a_j + P^k a_j P^k)(\varepsilon_{-j}) \parallel^2 \to 0,\ k \to \infty. \tag{7.48}$$

On the other hand, we have

$$\begin{aligned} &E\Big\|\sum_{j\geq 2}\sum_{i=0}^{j-1}(\rho_{j-i,k}P^k a_i)(I - P^k)(\varepsilon_{-j})\Big\|^2 \\ \leq\ & E\Big(\sum_{j\geq 2}\sum_{i=0}^{j-1}\|\rho_{j-i,k}P^k\|_{\mathcal{L}}\|a_i\|_{\mathcal{L}}\|(I - P^k)(\varepsilon_{-j})\|\Big)^2 \\ \leq\ & r^2\Big(\sum_{j\geq 1}\|\rho_{j,k}P^k\|_{\mathcal{L}}\Big)^2 \sum_{m=k+1}^{\infty} \gamma_m, \end{aligned}$$

which converges to zero by involving arguments already used. This, combined with (7.48) and the Cauchy-Schwarz inequality, leads to

$$E \parallel \sum_{j\geq 1}(b_j(X_{-j}) - b_{j,k}(Y_{-j,k})) \parallel^2 \to 0,\ k \to \infty, \tag{7.49}$$

which yields the desired result, provided

$$\rho_j = b_j\ ,\ \forall j \geq 1.$$

We now complete the proof by noticing that almost sure convergence comes from Lemma 7.1. ■

Proof of Corollary 7.1.

We start the proof by showing the following technical lemma:

LEMMA 7.6 *Let* $\left(r^j_{l,k}\right)_{l\geq 1, k\geq 1}$ *be the terms of the matrix operator* R^j, $j \geq 0$. *For all* j *we have*

- *if* $k = 1$,

$$\begin{cases} r^j_{l,1} &= a_{j+1-l} \text{ for } 1 \leq l \leq j+1 \\ r^j_{l,1} &= 0 \text{ for } l > j+1, \end{cases} \tag{7.50}$$

• *if* $k \geq 2$,

$$\begin{cases} r^j_{l,k} & = a_{j+k-l} - \sum_{m=1}^{k-1} a_{j+k-l-m}\rho_m \text{ for } 1 \leq l \leq j \\ r^j_{j+k,k} & = I \\ r^j_{l,k} & = 0 \text{ otherwise.} \end{cases} \tag{7.51}$$

Proof of lemma

We use induction. For $j = 0$, one easily obtains that for all $l \geq 1$, $r^0_{l,l} = I$, and all $l \geq 1$ and all $k \geq 1$, such that $l \neq k$, $r^0_{l,k} = 0$. Thus $R^0 = I_{H_w}$.

Suppose now that (7.50) and (7.51) hold true for all $m \leq j$. Then we get

• if $k = 1$ and $1 \leq l \leq j + 2$,

$$\begin{aligned} r^{j+1}_{l,1} &= r^j_{l,1}\rho_1 + r^j_{l,2} \\ &= a_{j+1-l}\rho_1 + a_{j+2-l} - a_{j+1-l}\rho_1 \\ &= a_{j+2-l}, \end{aligned}$$

• if $k = 1$ and $l > j + 3$,

$$r^{j+1}_{l,1} = 0,$$

• if $k \geq 2$ and $1 \leq l \leq j$,

$$\begin{aligned} r^{j+1}_{l,k} &= r^j_{l,1}\rho_k + r^j_{l,k+1} \\ &= a_{j+1-l}\rho_k + a_{j+k+1-l} - \sum_{m=1}^{k} a_{j+k+1-l-m}\rho_m \\ &= a_{j+k+1-l} - \sum_{m=1}^{k-1} a_{j+k+1-l-m}\rho_m. \end{aligned}$$

Finally,

$$\begin{aligned} r^{j+1}_{j+1,k} &= r^j_{j,k} = a_{j+k-j} - \sum_{m=1}^{k-1} a_{j+k-j-m}\rho_m = a_k - \sum_{m=1}^{k-1} a_{k-m}\rho_m. \\ r^{j+1}_{j+1+k,k} &= r^j_{j+1+k,k}\rho_k = 0. \\ r^{j+1}_{j+1+k,k+1} &= I. \end{aligned}$$

Thus relations (7.50) and (7.51) are true at rank $j + 1$, and the proof is complete. ■

We now return to the proof of Corollary 7.1. From Lemma 7.1, in order to show that $Y_n = \sum_{j=0}^{\infty} R^j(e_{n-j})$ converges in $L^2_{H_w}(\Omega, \mathcal{A}, P)$ and almost surely, it suffices to prove that $\sum_{j=0}^{\infty} \| R^j \|^2_{\mathcal{L}_w} < \infty$.

Let $x = (x_i)_{i\geq 1} \in H_w$. From Lemma 7.6 and due to the special form of (w_j), it follows that

$$\begin{aligned}\| R^j x \|_w^2 &= \sum_{l=1}^{j} s^l \sum_{k=1}^{\infty} \| r_{l,k}^j x_k \|^2 + \sum_{l=j+1}^{\infty} s^l \| x_{l-j} \|^2 \\ &= \sum_{l=1}^{j} s^l \sum_{k=1}^{\infty} \| r_{l,k}^j x_k \|^2 + \sum_{k=1}^{\infty} s^{k+j} \| x_k \|^2 .\end{aligned} \tag{7.52}$$

But $\| x \|_{H_w}^2 = \sum_{k=1}^{\infty} s^k \| x_k \|^2$; thus $\sup_{\substack{x\in H_w \\ \|x\|_{H_w}\leq 1}} \sum_{k=1}^{\infty} s^{k+j} \| x_k \|^2 \leq s^j$.

Moreover, the Cauchy-Schwarz inequality yields

$$\begin{aligned}\sum_{l=1}^{j} s^l \| \sum_{k=1}^{\infty} r_{l,k}^j x_k \|^2 &\leq \sum_{l=1}^{j} s^l \left(\sum_{k=1}^{\infty} \| r_{l,k}^j \|_{\mathcal{L}} \| x_k \| \right)^2 \\ &= s^j \sum_{l=1}^{j} \left(\sum_{k=1}^{\infty} s^{-(j+k-l)/2} \| r_{l,k}^j \|_{\mathcal{L}} \, s^{k/2} \| x_k \| \right)^2 \\ &\leq s^j \sum_{l=1}^{j} \sum_{k=1}^{\infty} s^{-(j+k-l)} \| r_{l,k}^j \|_{\mathcal{L}}^2 \sum_{k'=1}^{\infty} s^{k'} \| x_{k'} \|^2;\end{aligned}$$

hence

$$\sup_{\substack{x\in H_w \\ \|x\|_{H_w}\leq 1}} \sum_{l=1}^{j} s^l \| \sum_{k=1}^{\infty} r_{l,k}^j x_k \|^2 \leq s^j \sum_{l=1}^{j} \sum_{k=1}^{\infty} s^{-(j+k-l)} \| r_{l,k}^j \|_{\mathcal{L}}^2 .$$

It follows that

$$\| R^j \|_{L(H_w)}^2 = \sup_{\substack{x\in H_w \\ \|x\|_{H_w}\leq 1}} \| R^j x \|_{H_\beta}^2 \leq s^j \sum_{l=1}^{j} \sum_{k=1}^{\infty} s^{-(j+k-l)} \| r_{l,k}^j \|_{\mathcal{L}}^2 + s^j .$$

Now Lemma 7.6 leads to

$$\| r_{l,1}^j \|_{\mathcal{L}} = \| a_{j+1-l} \|_{\mathcal{L}}$$
$$\| r_{l,k}^j \|_{\mathcal{L}} \leq \| a_{j+k-l} \|_{\mathcal{L}} + \sum_{m=1}^{k-1} \| a_{j+k-l-m} \|_{\mathcal{L}} \| \rho_m \|_{\mathcal{L}}, \ \forall k \geq 2,$$

for all $l \in [1, j]$.

However, in the proof of Lemma 7.5, we have defined positive numbers $(c_j, j \geq 1)$, which satisfy the induction relation (7.41). From (7.46) we get $\|\rho_j\|_{\mathcal{L}} \leq c_j$ for all $j \geq 1$. Therefore

$$\| r_{l,k}^j \|_{\mathcal{L}} \leq c_{j+k-l},$$

for all $j \geq 1$, $k \geq 1$, and $1 \leq l \leq j$;
hence

$$\begin{aligned}
\sum_{j\geq 1} s^j \sum_{l=1}^{j} \sum_{k=1}^{\infty} s^{-(j+k-l)} \parallel r_{l,k}^j \parallel_{\mathcal{L}}^2 &\leq \sum_{j\geq 1} s^j \sum_{l=1}^{j} \sum_{k=1}^{\infty} s^{-(j+k-l)} c_{j+k-l}^2 \\
&\leq \sum_{j\geq 1} \sum_{l=1}^{j} s^j \sum_{p=1}^{\infty} s^{-p} c_p^2 \\
&\leq \left(\sum_{j\geq 1} j s^j \right) \left(\sum_{k=1}^{\infty} s^{-(p/2)} c_p \right)^2 .
\end{aligned}$$

Obviously $\sum_{j\geq 1} j \exp(-j \ln \frac{1}{s}) < \infty$. Moreover, since the c_j's satisfy (7.41), standard arguments show that (7.25) yields $\sum_{k=1}^{\infty} s^{-(p/2)} c_p < \infty$.

Using (7.52) we finally obtain

$$\sum_{j\geq 1} \parallel R^j \parallel_w^2 < \infty,$$

and Lemma 3.1 implies (7.29) and consequently (7.28). ■

NOTES

7.1 Lemma 7.1 is essentially due to Denisevskii and Dorogovtsev (1988).

7.2 Theorem 7.1 appears in Nsiri and Roy (1993). Lemma 7.2 and Theorem 7.2 come from Merlevède (1995, 1996,a).

7.3 All results are due to Merlevède (1995, 1996, a).

7.4 Lemma 7.4 and Theorems 7.4 and 7.5 are new. Theorem 7.6 is an easy consequence of results established by Merlevède (1996, b). Denisevskii and Dorogovtsev (1988) have obtained Theorem 7.7. Merlevède, Peligrad, and Utev (1997) have proved Theorem 7.8, Corollary 7.3, and sharpness of the condition $\sum_{j=0}^{\infty} \| a_j \|_{\mathcal{L}} < \infty$.

8

Estimation of Autocorrelation Operator and Prediction

We have seen that regular enough $ARH(p)$ and LPH have Markov representations that let us interpret them as $ARH'(1)$ processes, where H' is a suitable Hilbert space (see Chapters 5 and 7). Similar representations hold for empirical autocovariance operators associated with such processes (Chapter 4).

Thus the $ARH(1)$ model studied in Chapter 3 appears as basic. Recall that it is written as

$$X_n = \rho(X_{n-1}) + \varepsilon_n,\ n \in \mathbb{Z}, \tag{8.1}$$

where $(\varepsilon_n,\ n \in \mathbb{Z})$ is a Hilbertian white noise and ρ is a linear bounded operator over H.

Identification of this model takes place via estimation of ρ. Moreover, since $\rho(X_n)$ is the best predictor of X_{n+1} in some sense, an estimator $\overline{\rho_n}$ of ρ provides the statistical predictor $\overline{\rho_n}(X_n)$.

The current chapter is devoted to this topic. Section 1 describes the **finite-dimensional case**. It is shown that the empirical autocorrelation operator has good asymptotic behavior. Results presented here are more precise than those usually given in literature on the subject. As a by-product, consistent estimators of C_ε and $\sigma^2 = E\,\|\varepsilon_0\|^2$ are obtained.

The infinite-dimensional case is much more intricate. This is due to unboundedness of C^{-1} which induces erratic behavior of C_n^{-1}. A remedy is projection of data over a suitable finite-dimensional subspace of H. A natural choice of such a space is $H_{k_n} = \mathrm{sp}\,\{v_1, \ldots, v_{k_n}\}$, where (v_j) is a sequence of eigenvectors of C, the covariance operator of X_n. If (v_j) is unknown, H_{k_n} is replaced by $\widetilde{H}_{k_n} = \mathrm{sp}\,\{v_{1n}, \ldots, v_{nk_n}\}$ $((v_{jn})$ denotes a sequence of empirical eigenvectors).

The special case "(v_j) known" appears in Section 2 while "(v_j) unknown" is considered in Section 3. This method for estimating ρ is tractable and provides good results, as we will see in Chapter 9. Speeds of convergence depend on the choice of (k_n).

Section 4 is devoted to estimation of ρ in $C[0, 1]$. Section 5 deals with application to forecasting and Section 6 contains some special proofs.

8.1 Estimation of ρ if H is finite dimensional

Let $X = (X_n,\ n \in \mathbb{Z})$ be a **standard** $ARH(1)$ process, i.e,. X satisfies (8.1) and there exists $j_0 \geq 1$ such that $\| \rho^{j_0} \|_{\mathcal{L}} < 1$.

In this Section we suppose that H is **finite dimensional** and set $\dim H = k$. Thus H is isomorphic to $\mathbb{R}^k$ equipped with its natural scalar product

$$< x, y >= \sum_{j=1}^{k} x_j y_j; \quad x = (x_1, \ldots, x_k),\ y = (y_1, \ldots, y_k) \in \mathbb{R}^k.$$

We intend to estimate ρ from the data $X_1, \ldots, X_n$. Using (3.16) and the notation $D = C_{X_n,\, X_{n+1}}$ we arrive at

$$D = \rho C. \tag{8.2}$$

Then, if C is invertible, we have

$$\rho = DC^{-1} \tag{8.3}$$

and a genuine method for estimating ρ is to use

$$C_n = \frac{1}{n} \sum_{i=1}^{n} < X_i, \cdot > X_i \tag{8.4}$$

and

$$D_n = \frac{1}{n-1} \sum_{i=1}^{n-1} < X_i, \cdot > X_{i+1},\ n \geq 2, \tag{8.5}$$

for obtaining the estimator

$$\overline{\rho}_n = D_n C_n^{-1} \tag{8.6}$$

provided invertibility of C_n.

This estimator has good asymptotic properties with respect to pointwise consistency and convergence in distribution. In order to study asymptotic error in $L^2_{\mathcal{L}}$ we will consider a slight modification of $\overline{\rho}_n$.

In the remainder of this section the general assumption we will use is the following.

Assumption A

- X is a standard $ARH(1)$ with $\dim H < \infty$,
- $E \parallel X_0 \parallel^4 < \infty$,
- (ε_n) is a strong white noise,
- C is invertible,
- C_n is almost surely invertible for $n \geq k$.

We first deal with pointwise convergence.

THEOREM 8.1 *If A holds, then*

$$\frac{n^{1/4}}{(\text{Log } n)^\beta} \parallel \overline{\rho}_n - \rho \parallel_{\mathcal{L}} \to 0 \quad a.s.\,, \ \beta > \frac{1}{2}. \tag{8.7}$$

Proof

For $n \geq k$, write

$$\overline{\rho}_n - \rho = (D_n - D)C_n^{-1} + D\left(C_n^{-1} - C^{-1}\right),\ (a.s.) \tag{8.8}$$

and note that

$$C_n^{-1} - C^{-1} = C_n^{-1}(C - C_n)C^{-1}. \tag{8.9}$$

Therefore

$$\parallel \overline{\rho}_n - \rho \parallel_{\mathcal{L}} \leq \parallel C_n^{-1} \parallel_{\mathcal{L}} \left[\parallel D_n - D \parallel_{\mathcal{L}} + \parallel D \parallel_{\mathcal{L}} \parallel C^{-1} \parallel_{\mathcal{L}} \parallel C_n - C \parallel_{\mathcal{L}}\right]. \tag{8.10}$$

Now from (4.40) in Theorem 4.4 it follows that

$$\parallel C_n^{-1} \parallel_{\mathcal{L}} = \lambda_{k,n}^{-1} \to \lambda_k^{-1} \ \ a.s.$$

and from (4.18) in Theorem 4.1 and (4.73) in Theorem 4.8 one obtains (8.7). ■

The next statement provides a large deviation inequality.

THEOREM 8.2 *If A is satisfied and $\parallel X_0 \parallel$ is bounded then for all $\eta > 0$,*

$$P(\parallel \overline{\rho}_n - \rho \parallel_{\mathcal{L}} \geq \eta) \leq 16 \exp\left(-\frac{n\lambda_k^2\eta'^2}{a + b\lambda_k\eta'}\right), \tag{8.11}$$

where $\eta' = \min\left(\eta, \dfrac{\eta}{\parallel D \parallel_{\mathcal{L}} \parallel C^{-1} \parallel_{\mathcal{L}}}, 2\right)$, $a > 0$, $b > 0$ *depend only on the distribution of the process* X.

Proof

Let $\eta > 0$ be arbitrary. We have

$$\begin{aligned} P(\| \overline{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta) &= P\left(\| \overline{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta,\ \lambda_{k,n} \geq \frac{\lambda_k}{2}\right) \\ &\quad + P\left(\| \overline{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta,\ \lambda_{k,n} < \frac{\lambda_k}{2}\right). \end{aligned}$$

Thus (8.10) yields

$$\begin{aligned} P(\| \overline{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta) \leq\ & P\left(\frac{2}{\lambda_k}(\| D_n - D \|_{\mathcal{L}} + \| D \|_{\mathcal{L}} \| C^{-1} \|_{\mathcal{L}} \| C_n - C \|_{\mathcal{L}}) \geq \eta\right) \\ & + P\left(|\lambda_{k,n} - \lambda_k| \geq \frac{\lambda_k}{2}\right), \end{aligned}$$

which, from (4.43), implies

$$\begin{aligned} P(\| \overline{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta) \leq\ & P\left(\| D_n - D \|_{\mathcal{L}} \geq \frac{\lambda_k \eta}{4}\right) \\ & + P\left(\| C_n - C \|_{\mathcal{L}} \geq \frac{\lambda_k \eta}{4 \| D \|_{\mathcal{L}} \| C^{-1} \|_{\mathcal{L}}}\right) \\ & + P\left(\| C_n - C \|_{\mathcal{L}} \geq \frac{\lambda_k}{2}\right), \end{aligned}$$

and the desired result follows from (4.22) in Theorem 4.2 and (4.74) in Theorem 4.8. ∎

COROLLARY 8.1

$$\| \overline{\rho}_n - \rho \|_{\mathcal{L}} = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \quad a.s. \tag{8.12}$$

Proof

Choose $\eta = c\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}$, apply (8.11), and use the Borel-Cantelli Lemma with a suitable choice of c. ∎

Asymptotic normality of $\overline{\rho}_n$ appears in the next statement. Note that here $\mathcal{L} = \mathcal{S} = \mathcal{N}$ and all norms over $\mathcal{L}$ are equivalent.

THEOREM 8.3 *If A holds, then*

$$\sqrt{n}(\overline{\rho}_n - \rho) \xrightarrow{\mathcal{D}} N', \tag{8.13}$$

where N' is a zero-mean Gaussian random operator.

Proof

Theorem 4.9 gives

$$\sqrt{n}(C_n - C,\ D_n - D) \xrightarrow{\mathcal{D}} N.$$

On the other hand, (8.8) and (8.9) yield

$$\begin{aligned} U_n := \sqrt{n}(\overline{\rho}_n - \rho) = \quad & DC_n^{-1}[\sqrt{n}(C - C_n)]C^{-1} \\ & + [\sqrt{n}(D_n - D)]C_n^{-1}. \end{aligned}$$

Now set

$$V_n = DC^{-1}[\sqrt{n}(C - C_n)]C^{-1} + [\sqrt{n}(D_n - D)]C^{-1}$$

and $\Lambda_n = \| \sqrt{n}(C - C_n) \|_{\mathcal{L}} + \| \sqrt{n}(D_n - D) \|_{\mathcal{L}}$.
Then

$$\| V_n - U_n \|_{\mathcal{L}} \leq \| C_n^{-1} - C^{-1} \|_{\mathcal{L}} \Lambda_n.$$

By continuity and (2.11), Λ_n converges in distribution and since $\| C_n^{-1} - C^{-1} \|_{\mathcal{L}} \xrightarrow[p]{} 0$, we have $\| U_n - V_n \|_{\mathcal{L}} \xrightarrow[p]{} 0$.

Using (2.11) again we see that

$$V_n \xrightarrow{\mathcal{D}} DC^{-1}N_1C^{-1} + (\rho N_1 + N_2)C^{-1}, \tag{8.14}$$

where (N_1, N_2) is defined in Theorem 4.9.

Finally, (2.9) gives the same limit in distribution for U_n and the proof is complete. ■

The covariance operator of N can be specified by using (8.14) and Theorem 4.9.

We now define a new estimator of ρ that works even if C_n is non-invertible. It is also interesting if C_n admits some very small eigenvalues that induce a very large norm for C_n^{-1} and make its computation difficult.

For this purpose we write C_n under the spectral form

$$C_n = \sum_{j=1}^{k} \lambda_{j,n} < v_{jn}, \cdot > v_{jn}, \quad n \geq 1, \tag{8.15}$$

where (v_{jn}) is an orthonormal basis of H such that

$$C_n(v_{jn}) = \lambda_{j,n} v_{jn}, \quad 1 \leq j \leq k, \tag{8.16}$$

with, as usual,

$$\lambda_{1,n} \geq \lambda_{2,n} \geq \ldots \geq \lambda_{k,n} \geq 0.$$

Now let $a = (a_n,\ n \geq 1)$ be a sequence of real numbers such that $0 < a_n \leq 1$.

To (a_n) we associate a new estimator of C, defined as

$$C_{n,a} = \sum_{j=1}^{k} \max(\lambda_{jn}, a_n) \ < v_{jn}, \cdot > v_{jn},\ n \geq 1. \tag{8.17}$$

This estimator is clearly invertible and

$$\| C_{n,a}^{-1} \|_{\mathcal{L}} \leq a_n^{-1}. \tag{8.18}$$

Our estimator of ρ is now obtained by setting

$$\overline{\rho}_{n,a} = D_n C_{n,a}^{-1}. \tag{8.19}$$

Note that $\overline{\rho}_{n,a}$ coincides with $\overline{\rho}_n$ as soon as $\lambda_{k,n} \geq a_n$.

Theorems 8.1, 8.2, and 8.3 remain valid for $\overline{\rho}_{n,a}$, with slight modifications concerning coefficients. We now consider the $L^2_{\mathcal{L}}$ error.

THEOREM 8.4 *If A holds and if $(a_n) \to 0$, there exists an integer $n(a, \lambda_k)$ such that*

$$E \| \overline{\rho}_{n,a} - \rho \|_{\mathcal{L}}^2 \leq \frac{c}{n a_n^2}\ ,\ \ n \geq n(a, \lambda_k), \tag{8.20}$$

where c is a positive constant.

Proof

Let us start from the identity

$$\begin{aligned} \overline{\rho}_{n,a} - \rho &= (D_n - D) C_{n,a}^{-1} + D \left(C_{n,a}^{-1} - C^{-1} \right) \\ &= (D_n - D) C_{n,a}^{-1} - D C_{n,a}^{-1} (C_{n,a} - C) C^{-1}, \end{aligned} \tag{8.21}$$

which yields

$$\begin{aligned} E \| \overline{\rho}_{n,a} - \rho \|_{\mathcal{L}}^2 \leq \ & 2 a_n^{-2} E \| D_n - D \|_{\mathcal{L}}^2 \\ & + 2 a_n^{-2} \| D \|_{\mathcal{L}}^2 \| C^{-1} \|_{\mathcal{L}}^2 \, E \| C_{n,a} - C \|_{\mathcal{L}}^2 \ . \end{aligned} \tag{8.22}$$

The main task is to evaluate $E \| C_{n,a} - C \|_{\mathcal{L}}^2$. Noting that there exists an integer $n(a, \lambda_k)$ such that

$$a_n \leq \frac{\lambda_k}{2}\ ,\ \ n \geq n(a, \lambda_k), \tag{8.23}$$

one may write

$$\begin{aligned} P(C_{n,a} \neq C_n) \ &= P(\lambda_{kn} < a_n) = P(\lambda_{kn} - \lambda_k < a_n - \lambda_k) \\ &\leq P\left(|\lambda_{kn} - \lambda_k| \geq \frac{\lambda_k}{2} \right),\ \ n \geq n(a, \lambda_k). \end{aligned}$$

Hence (4.43) entails

$$P(C_{n,a} \neq C_n) \leq P\left(\| C_n - C \|_{\mathcal{L}} \geq \frac{\lambda_k}{2}\right), \ n \geq n(a, \lambda_k). \tag{8.24}$$

Now, from (4.17), there exists a constant γ such that

$$E \| C_n - C \|_{\mathcal{L}}^2 < \frac{\gamma}{n}; \tag{8.25}$$

thus the Bienaymé-Tchebychev inequality gives

$$P(C_{n,a} \neq C_n) < \frac{4\gamma}{\lambda_k^2} \frac{1}{n}, \ n \geq n(a, \lambda_k). \tag{8.26}$$

Next consider the decomposition

$$\begin{aligned} E \| C_{n,a} - C \|^2 = & \ E\left(\| C_{n,a} - C \|_{\mathcal{L}}^2 \mathbf{I}_{\{C_{n,a} \neq C_n\}}\right) \\ & + E\left(\| C_{n,a} - C \|_{\mathcal{L}}^2 \mathbf{I}_{\{C_{n,a} = C_n\}}\right), \end{aligned}$$

and note that $C_{n,a} \neq C_n$ implies $a_n > \lambda_{k_{n,n}}$. Then

$$\begin{aligned} E \| C_{n,a} - C \|^2 \leq & \ 2E\left(\| C_{n,a} - C_n \|_{\mathcal{L}}^2 \mathbf{I}_{\{C_{n,a} \neq C_n\}}\right) \\ & + 2E\left(\| C - C_n \|_{\mathcal{L}}^2 \mathbf{I}_{\{C_{n,a} \neq C_n\}}\right) \\ & + E\left(\| C_{n,a} - C \|_{\mathcal{L}}^2 \mathbf{I}_{\{C_{n,a} = C_n\}}\right) \\ & \leq 2a_n^2 P(C_{n,a} \neq C_n) + 3E \| C_n - C \|_{\mathcal{L}}^2; \end{aligned}$$

thus

$$E \| C_{n,a} - C \|_{\mathcal{L}}^2 = O\left(\frac{1}{n}\right). \tag{8.27}$$

On the other hand, (4.69) induces

$$E \| D_n - D \|_{\mathcal{L}}^2 = O\left(\frac{1}{n}\right), \tag{8.28}$$

hence (8.20) by combining (8.27) and (8.28). ■

If something is known about λ_k, it is possible to reach a $\frac{1}{n}$-rate, as shows the following corollary.

COROLLARY 8.2 *If A holds and $\lambda_k \geq m > 0$, where m is known, then the choice $a_n = \min\left(1, \frac{m}{2}\right)$, $n \geq 1$ leads to*

$$E \| \overline{\rho}_{n,a} - \rho \|_{\mathcal{L}}^2 = O\left(\frac{1}{n}\right). \tag{8.29}$$

Proof

Here (8.23) is valid for every n. Thus (8.29) follows from (8.20). ■

Example 8.1
If $k = 1$ we have

$$\overline{\rho}_n = \frac{n}{n-1} \frac{\sum_{i=1}^{n-1} X_i X_{i+1}}{\sum_{i=1}^{n} X_i^2}$$

and

$$\overline{\rho}_{n,a} = \frac{n}{n-1} \frac{\sum_{i=1}^{n-1} X_i X_{i+1}}{\max\left(na_n, \sum_{i=1}^{n} X_i^2\right)}$$

and the above results hold provided X is an $AR(1)$ associated with a strong white noise and is such that $EX_0^4 < \infty$ (resp. X_0 is bounded). Here asymptotic variance obtained in Theorem 8.3 is $1 - \rho^2$.

Estimation of C_ε

As a by-product, it is interesting to consider estimation of C_ε and $\sigma^2 = E \parallel \varepsilon_0 \parallel^2$, since these parameters characterize prediction error (see Section 5).

Recall that (3.13) gives

$$C_\varepsilon = C - \rho C \rho^*.$$

Then a plug-in estimator of C_ε is

$$C_{\varepsilon,n} = C_n - \overline{\rho}_n C_n \overline{\rho}_n^*. \tag{8.30}$$

Now it is easy to check that

$$\frac{n^{1/4}}{(\text{Log } n)^\beta} \parallel C_{\varepsilon,n} - C_\varepsilon \parallel_{\mathcal{L}} \to 0 \quad a.s., \quad \beta > \frac{1}{2}, \tag{8.31}$$

and if $\parallel X_0 \parallel$ is bounded then

$$\parallel C_{\varepsilon,n} - C_\varepsilon \parallel_{\mathcal{L}} = O\left(\left(\frac{\text{Log } n}{n}\right)^{1/2}\right) \quad a.s. \tag{8.32}$$

Finally, Theorems 4.9 and 8.3 allow us to obtain asymptotic normality of $C_{\varepsilon,n}$. Details are left to the reader.

Concerning σ^2, it suffices to recall that

$$\sigma^2 = \parallel C_\varepsilon \parallel_{\mathcal{N}}$$

and that here $\| \cdot \|_{\mathcal{N}}$ and $\| \cdot \|_{\mathcal{L}}$ are equivalent since H is finite dimensional.

Thus a natural estimator of σ^2 is

$$\sigma_n^2 = \| C_{\varepsilon,n} \|_{\mathcal{N}} \tag{8.33}$$

and it possesses asymptotic properties that are easily deduced from those of $C_{\varepsilon,n}$.

Computation of σ_n^2 can be performed by using the identity

$$\sigma_n^2 = \sum_{j=1}^{k} \gamma_{j,n}, \tag{8.34}$$

where $\gamma_{j,n}$ is the j^{th} eigenvalue of $C_{\varepsilon,n}$.

8.2 Estimation of ρ in a special case

If $\dim H = +\infty$, estimation of ρ is a rather difficult problem since $D = \rho C$ does not imply $\rho = DC^{-1}$ over the whole space.

Actually if $\lambda_j > 0$ for every j, the range of C is defined as

$$C(H) = \left\{ y : y \in H, \ \sum_{j=1}^{\infty} \frac{< y, v_j >^2}{\lambda_j^2} < \infty \right\}. \tag{8.35}$$

Thus $C(H)$ is strictly included in H.

What is worse, C_n^{-1} can only be defined over the n-dimensional space $\text{sp}(X_1, \ldots, X_n)$, and so $D_n C_n^{-1}$ is not a satisfactory estimator.

A possible method is then to project data over a suitable finite dimensional subspace of H say H_{k_n}.

In this section we suppose that the eigenvectors (v_j) of C are **known**. In that case it is natural to select $H_{k_n} = \text{sp}\,(v_1, \ldots, v_{k_n})$ since projection of data over this space minimizes loss of information, as shown by Grenander (1981).

We first give an example of such a situation.

Example 8.2

Consider Example 3.8, with $\left(\dfrac{\sigma_j^2}{1 - \alpha_j^2} \right)$ decreasing. Here we have $v_1 = (1, 0, 0, \ldots)$, $v_2 = (0, 1, 0, \ldots), \ldots$ and the sequence (v_j) is thus known.

We now make the following assumptions.

A_1 X is a standard $ARH(1)$ associated with a strong H-white noise and such that $E \| X_0 \|^4 < \infty$.

B_1 $\lambda_j > 0$ for all $j \geq 1$,

C_1 $P(< X_0, v_j > = 0) = 0$ for all $j \geq 1$.

Note that if X_0 is Gaussian, then $< X_0, v_j > \sim \mathcal{N}(0, \lambda_j)$ and consequently B_1 implies C_1.

Now consider the alternative estimator $(\widehat{\lambda}_{jn})$ of (λ_j) defined in Chapter 4:

$$\widehat{\lambda}_{jn} = \frac{1}{n} \sum_{i=1}^{n} < X_i, v_j >^2, \; j \geq 1 \; ; \; n \geq 1. \tag{8.36}$$

C_1 yields $\widehat{\lambda}_{jn} \neq 0$ (a.s.) and $(\widehat{\lambda}_{jn})$ induces an estimator of C defined as

$$\widehat{C}_n = \sum_{j=1}^{\infty} \widehat{\lambda}_{jn} < v_j, \cdot > v_j, \tag{8.37}$$

with

$$\sum_{j=1}^{\infty} \widehat{\lambda}_{jn} = \frac{1}{n} \sum_{j=1}^{\infty} \sum_{i=1}^{n} < X_i, v_j >^2 = \frac{1}{n} \sum_{i=1}^{n} \| X_i \|^2 < \infty, \tag{8.38}$$

which proves that $\widehat{C}_n$ is nuclear.

Note also that $\widehat{\lambda}_{jn}$ is an unbiased estimator of λ_j but that $(\widehat{\lambda}_{jn}, \; j \geq 1)$ is not a decreasing sequence.

Now $\widehat{C}_n$ is invertible over H_{k_n} with inverse given by

$$\widehat{C}_n^{-1}(x) = \sum_{j=1}^{k_n} \widehat{\lambda}_{jn}^{-1} < x, v_j > v_j, \;\; x \in H_{k_n} \;\; (a.s.). \tag{8.39}$$

Thus we may define an estimator of ρ by setting:

$$\widehat{\rho}_n(x) = \left(\pi^{k_n} D_n \widehat{C}_n^{-1} \pi^{k_n} \right)(x), \; x \in H, \tag{8.40}$$

where π^{k_n} denotes the orthogonal projector of H_{k_n}.

Observe that $\widehat{\rho}_n$ is completely determined by $(< X_i, v_j >, \; 1 \leq j \leq k_n, \; 1 \leq i \leq n)$ or equivalently by $\pi^{k_n} X_1, \ldots, \pi^{k_n} X_n$. Actually, it may be written under the form

$$\widehat{\rho}_n(x) = \sum_{\ell=1}^{k_n} \widehat{\rho}_{n\ell}(x) v_\ell, \tag{8.41}$$

where

$$\widehat{\rho}_{n\ell}(x) = \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \widehat{\lambda}_{jn}^{-1} < x, v_j >< X_i, v_j >< X_{i+1}, v_\ell >, \tag{8.42}$$

$1 \leq \ell \leq k_n$, $n \geq 2$, $x \in H$.

Moreover, $\widehat{\rho}_n$ is a bounded linear operator with

$$\| \widehat{\rho}_n \|_{\mathcal{L}} \leq \| D_n \|_{\mathcal{L}} \max_{1 \leq j \leq k_n} \widehat{\lambda}_{j,n}^{-1}. \tag{8.43}$$

We first indicate some simple pointwise results. In the following we always suppose that $(k_n) \to \infty$.

LEMMA 8.1

1) For all $x \in \mathrm{sp}(v_j,\ j \geq 1)$,

$$\| \widehat{\rho}_n(x) - \rho(x) \| \to 0 \quad a.s. \tag{8.44}$$

2) If $\underline{\lim} \dfrac{n\lambda_{k_n}^4}{(\mathrm{Log}\ n)^\alpha} > 0$ *for some* $\alpha > 2$,

$$\| \widehat{\rho}_n(x) - \rho(x) \| \to 0 \quad a.s.\ , \quad x \in C(H). \tag{8.45}$$

3) If $\underline{\lim} \dfrac{n\lambda_{k_n}^8}{(\mathrm{Log}\ n)^\alpha} > 0$ *for some* $\alpha > 2$

$$\| \widehat{\rho}_n(x) - \rho(x) \| \to 0 \quad a.s.\ , \quad x \in H. \tag{8.46}$$

Proof

1) Let x be an element of $\mathrm{sp}(v_j,\ j \geq 1)$. There exists an integer k such that

$$x = \sum_{j=1}^{k} < x, v_j > v_j;$$

thus, for large enough n , $k_n \geq k$ and $\pi^{k_n} x = x$.
Then we have

$$\widehat{\rho}_n(x) = \sum_{j=1}^{k} \widehat{\lambda}_{jn}^{-1} < x, v_j > \pi^{k_n} D_n(v_j).$$

Using Corollary 4.5 and Theorem 4.8 we get

$$\widehat{\lambda}_{jn} \to \lambda_j \ \ a.s. \text{ and } \ \pi^{k_n} D_n(v_j) \to D(v_j)\ a.s.;$$

therefore

$$\widehat{\rho}_n(x) \underset{a.s.}{\to} \sum_{j=1}^{k} \lambda_j^{-1} < x, v_j > D(v_j)$$

and

$$\begin{aligned} \sum_{j=1}^{k} \lambda_j^{-1} < x, v_j > D(v_j) &= \sum_{j=1}^{k} \lambda_j^{-1} < x, v_j > \rho(C v_j) \\ &= \rho(x), \end{aligned}$$

which gives (8.44).

2) Let $x \in C(H)$, and write

$$\widehat{\rho}_n(x) = \pi^{k_n} D_n \left(\sum_{j=1}^{k_n} \widehat{\lambda}_{jn}^{-1} < x, v_j > v_j \right) := \pi^{k_n} D_n(\widehat{V}_n),$$

and

$$\left(\pi^{k_n} \rho \pi^{k_n}\right)(x) = \pi^{k_n} D \left(\sum_{j=1}^{k_n} \lambda_j^{-1} < x, v_j > v_j \right) := \pi^{k_n} D(V_n).$$

Thus

$$\| \widehat{\rho}_n(x) - \left(\pi^{k_n} \rho \pi^{k_n}\right)(x) \| \leq \| D_n \|_{\mathcal{L}} \| \widehat{V}_n - V_n \| + \| D_n - D \|_{\mathcal{L}} \| V_n \| . \tag{8.47}$$

Remark that since $x \in C(H)$ we have

$$\| V_n \| = \left(\sum_{j=1}^{k_n} \frac{< x, v_j >^2}{\lambda_j^2} \right)^{1/2} \leq \left(\sum_{j=1}^{\infty} \frac{< x, v_j >^2}{\lambda_j^2} \right)^{1/2} < \infty. \tag{8.48}$$

Concerning $\widehat{V}_n - V_n$ we have

$$\| \widehat{V}_n - V_n \| = \left(\sum_{j=1}^{k_n} \frac{< x, v_j >^2}{\lambda_j^2} \frac{(\widehat{\lambda}_{jn} - \lambda_j)^2}{\widehat{\lambda}_{jn}^2} \right)^{1/2},$$

and hence the bound

$$\| \widehat{V}_n - V_n \| \leq \frac{\| C_n - C \|_{\mathcal{L}}}{\inf\limits_{1 \leq j \leq k_n} \widehat{\lambda}_{jn}} \left(\sum_{j=1}^{\infty} \frac{< x, v_j >^2}{\lambda_j^2} \right)^{1/2} . \tag{8.49}$$

Now Theorem 4.1 gives

$$\frac{n^{1/4}}{(\text{Log } n)^{\alpha/4}} \sup_{j \geq 1} |\widehat{\lambda}_{jn} - \lambda_j| \leq \frac{n^{1/4}}{(\text{Log } n)^{\alpha/4}} \| C_n - C \|_{\mathcal{L}} \to 0 \;\; a.s..$$

Then, if $m := \underline{\lim} \dfrac{n^{1/4} \lambda_{k_n}}{(\text{Log } n)^{\alpha/4}}$, there exists an integer n_0 such that $n \geq n_0$ implies

$$\lambda_j \geq \frac{m}{2} n^{-1/4} (\text{Log } n)^{\alpha/4}, \;\; 1 \leq j \leq k_n.$$

On the other hand, for almost all ω in Ω and all η in $\left]0, \dfrac{m}{4}\right[$ there exists $n_1(\omega, \eta)$ such that

$$\frac{n^{1/4}}{(\text{Log } n)^{\alpha/4}} \sup |\widehat{\lambda}_{jn} - \lambda_j| \leq \frac{n^{1/4}}{(\text{Log } n)^{\alpha/4}} \| C_n - C \|_{\mathcal{L}} < \eta, \; n \geq n_1(\omega, \eta).$$

If $n \geq \max(n_0, n_1)$, we then have

$$\begin{aligned}\widehat{\lambda}_{jn} &> \lambda_j - \eta n^{-1/4}(\text{Log}\, n)^{\alpha/4} \\ &\geq \left(\frac{m}{2} - \eta\right) n^{-1/4}(\text{Log}\, n)^{\alpha/4}, \ \ 1 \leq j \leq k_n,\end{aligned}$$

and from (8.49) it follows that

$$\| \widehat{V}_n - V_n \| \leq \frac{\eta}{\frac{m}{2} - \eta} \left(\sum_{j=1}^{\infty} \frac{< x, v_j >^2}{\lambda_j^2} \right)^{1/2} < \frac{4\eta}{m} \left(\sum_{j=1}^{\infty} \frac{< x, v_j >^2}{\lambda_j^2} \right)^{1/2}.$$

Since η is arbitrary small, we have proved that

$$\| \widehat{V}_n - V_n \| \rightarrow 0 \quad a.s. \tag{8.50}$$

Now Theorem 4.8 entails

$$\| D_n - D \|_{\mathcal{L}} \rightarrow 0 \ \ a.s. \tag{8.51}$$

Finally, combining (8.47), (8.48), and (8.50), we obtain

$$\| \widehat{\rho}_n(x) - \left(\pi^{k_n} \rho \pi^{k_n}\right)(x) \| \rightarrow 0 \ \ a.s.$$

The desired result follows from

$$\| \left(\pi^{k_n} \rho \pi^{k_n}\right)(x) - \rho(x) \| \underset{k_n \rightarrow \infty}{\longrightarrow} 0. \tag{8.52}$$

3) To show (8.46) one may use the same method as above, with (8.49) replaced by

$$\| \widehat{V}_n - V_n \| \leq \frac{\| C_n - C \|_{\mathcal{L}}}{\lambda_{k_n} \inf\limits_{1 \leq j \leq k_n} \widehat{\lambda}_{jn}} \| x \| . \tag{8.53}$$

Details are omitted. ■

The next statement deals with uniform convergence.

THEOREM 8.5 *If A_1, B_1, and C_1 hold and ρ is a Hilbert-Schmidt operator, then*

$$\| \widehat{\rho}_n - \rho \|_{\mathcal{L}} \rightarrow 0 \ \ a.s., \tag{8.54}$$

provided

$$\underline{\lim} \frac{n \lambda_{k_n}^8}{(\text{Log}\, n)^{\alpha}} > 0 \ \textit{for some}\ \alpha > 2. \tag{8.55}$$

Proof

It is easy to see that convergence of $\| \widehat{\rho}_n(x) - \pi^{k_n}\rho\pi^{k_n}(x) \|$ is uniform over $\{x : \| x \| \leq 1\}$.

It remains to show that

$$\sup_{\|x\|\leq 1} \| \left(\pi^{k_n}\rho\pi^{k_n}\right)(x) - \rho(x) \| \to 0. \tag{8.56}$$

We first have, for all x such that $\| x \| \leq 1$,

$$\begin{aligned}
&\| \left(\pi^{k_n}\rho\pi^{k_n}\right)(x) - \left(\pi^{k_n}\rho\right)(x) \|^2 \\
&\leq \| \rho\left(\sum_{j>k_n} < x, v_j > v_j\right) \|^2 \leq \left(\sum_{j>k_n} | < x, v_j > | \, \| \rho(v_j) \|\right)^2 \\
&\leq \left(\sum_{j>k_n} < x, v_j >^2\right) \sum_{j>k_n} \| \rho(v_j) \|^2 \\
&\leq \sum_{j>k_n} \| \rho(v_j) \|^2 \to 0,
\end{aligned}$$

since $\sum_{j=1}^{\infty} \| \rho(v_j) \|^2 = \| \rho \|_{\mathcal{S}}^2$.

Similarly,

$$\sup_{\|x\|\leq 1} \| \left(\pi^{k_n}\rho\right)(x) - \rho(x) \|^2 = \sup_{\|x\|\leq 1} \| \rho\left(\sum_{j>k_n} < x, v_j > v_j\right) \|^2 \to 0$$

and the proof is complete. ■

Example 8.3

Suppose that $\lambda_j = ar^j$, $j \geq 1$, where $a > 0$ and $0 < r < 1$. Then, if $k_n \to \infty$ and $\dfrac{k_n}{\text{Log } n} \to 0$, condition (8.55) is satisfied.

Example 8.4

Suppose that $\lambda_j = aj^{-\delta}$, $j \geq 1$, where $a > 0$ and $\delta > 1$. Then $k_n = [n^v]$, where $v < \dfrac{1}{8\delta}$ insures (8.55).

We now give a large deviation inequality.

THEOREM 8.6 *If A_1, B_1, and C_1 hold, ρ is Hilbert-Schmidt and $\| X_0 \|$ is bounded, then for all $\eta > 0$ there exists an integer $n_0(\eta, \rho, C, (k_n))$ such that $n \geq n_0$ implies*

$$P(\| \widehat{\rho}_n - \rho \|_{\mathcal{L}} \geq \eta) \leq c_1 \exp\left(-c_2 n \lambda_{k_n}^2\right), \tag{8.57}$$

where $c_1 > 0$ and $c_2 > 0$ depend only on η and P_X.

Proof (sketch)

Since (8.56) holds we have, for large enough n

$$\| \pi^{k_n} \rho \pi^{k_n} - \rho \|_{\mathcal{L}} < \frac{\eta}{2}.$$

It remains to study $P\left(\| \widehat{\rho}_n - \pi^{k_n} \rho \pi^{k_n} \|_{\mathcal{L}} \geq \frac{\eta}{2}\right)$.

For this purpose we use (8.47) and (8.53). We have

$$\begin{aligned} P & \left(\| \widehat{\rho}_n - \pi^{k_n} \rho \pi^{k_n} \|_{\mathcal{L}} \geq \frac{\eta}{2}\right) \\ \leq & \ P\left(\| D_n - D \|_{\mathcal{L}} \geq \frac{\eta}{4} \lambda_{k_n}\right) \\ & + P\left(\| D_n \|_{\mathcal{L}} \frac{\| C_n - C \|_{\mathcal{L}}}{\min\limits_{1 \leq j \leq k_n} \widehat{\lambda}_{jn}} \geq \frac{\eta}{4} \lambda_{k_n}\right) \\ \leq & \ p_1 + p_2. \end{aligned}$$

A bound for p_1 is given by Theorem 4.8. Concerning p_2 let us set

$$F_n = \{\| D_n \|_{\mathcal{L}} < 2 \| D \|_{\mathcal{L}} + 1\},$$

$$G_n = \left\{\min_{1 \leq j \leq k_n} \widehat{\lambda}_{jn} \leq \frac{\lambda_{k_n}}{2}\right\},$$

$$H_n = \left\{\| D_n \|_{\mathcal{L}} \frac{\| C_n - C \|_{\mathcal{L}}}{\min\limits_{1 \leq j \leq k_n} \widehat{\lambda}_{jn}} \geq \frac{\eta}{4} \lambda_{k_n}\right\}$$

and consider the decomposition

$$H_n = H_n \cap F_n \cap G_n + H_n \cap F_n^c \cap G_n + H_n \cap F_n^c \cap G_n^c + H_n \cap F_n \cap G_n^c.$$

By using Theorems 4.2 and 4.8 it is easy to derive an exponential bound for $p_2 = P(H_n)$, and (8.57) follows. Details are omitted. ■

From (8.57) it follows that, if $\| X_0 \|$ is bounded, condition $n\lambda_{k_n}^2 \to \infty$ suffices for $\| \widehat{\rho}_n - \rho \|_{\mathcal{L}} \xrightarrow[p]{} 0$ and that

$$\frac{n\lambda_{k_n}^2}{\operatorname{Log} n} \to \infty \text{ gives } \| \widehat{\rho}_n - \rho \|_{\mathcal{L}} \to 0 \ a.s. \tag{8.58}$$

(compare with (8.55)).

If X is geometrically strongly mixing, one may relax the assumption of boundedness. Actually, if X_0 has an exponential moment, one may use Corollary 4.2 to obtain a bound for $P(\parallel \widehat{\rho}_n - \rho \parallel_{\mathcal{L}} \geq \eta)$. Details are left to the reader.

As in the finite-dimensional case, a slight modification of $\widehat{\rho}_n$ allows us to obtain a bound for $E \parallel \widehat{\rho}_{n,a} - \rho \parallel_{\mathcal{L}}^2$, where $\widehat{\rho}_{n,a}$ is the modified estimator. This topic is omitted.

8.3 The general situation

The assumption "(v_j) known" is of course somewhat unrealistic even if it is adapted to some interesting models (see Example 8.2).

In the general, case (v_j) needs to be replaced by (v_{jn}) and the space H_{k_n} by the random space $\widetilde{H}_{k_n} = \mathrm{sp}\,(v_{1n}, \ldots, v_{k_n n})$. The corresponding orthogonal projector will be denoted $\widetilde{\pi}^{k_n}$. Here (k_n) is a sequence of integers such that $k_n \leq n$, $n \geq 1$, and $(k_n) \to \infty$.

Concerning C and C_n we will use the following assumptions:

$$B_1' : \quad \lambda_1 > \lambda_2 > \ldots > \lambda_j > \ldots > 0,$$

$$C_1' : \quad \lambda_{k_n,n} > 0, \;\; n \geq 1 \;\; (a.s.).$$

Now define

$$\widetilde{C}_n = \widetilde{\pi}^{k_n} C_n = \sum_{j=1}^{k_n} \lambda_{jn} < v_{jn}, \cdot > v_{jn};$$

thus

$$\widetilde{C}_n^{-1}(x) = \sum_{j=1}^{k_n} \lambda_{jn}^{-1} < x, v_{jn} > v_{jn}, \;\; x \in \widetilde{H}_{k_n}.$$

Here and below "a.s." is omitted.

The induced estimator of ρ is

$$\rho_n = \widetilde{\pi}^{k_n} D_n \widetilde{C}_n^{-1} \widetilde{\pi}^{k_n}. \tag{8.59}$$

Note that $\widetilde{C}_n^{-1}\widetilde{\pi}^{k_n}$ should simply be replaced by $\widetilde{C}_n^{-1}$ since this operator naturally extends to H. Of course this extension is not inverse of $\widetilde{C}_n$ over H. Now we have

$$\rho_n(x) = \sum_{\ell=1}^{k_n} \rho_{n\ell}(x) v_{\ell n},$$

with

$$\rho_{n\ell}(x) = \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \lambda_{jn}^{-1} < x, v_{jn} >< X_i, v_{jn} >< X_{i+1}, v_{\ell,n} >,$$

$1 \leq \ell \leq k_n$, $n \geq 2$.

ρ_n is a bounded linear operator and

$$\| \rho_n \|_{\mathcal{L}} \leq \| D_n \|_{\mathcal{L}} \lambda_{k_n,n}^{-1}.$$

In order to study consistency of ρ_n we need two technical lemmas. Recall that

$$a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$$

and

$$a_j = 2\sqrt{2} \max \left[(\lambda_{j-1} - \lambda_j)^{-1}, \; (\lambda_j - \lambda_{j+1})^{-1} \right], \; j \geq 2$$

(see Lemma 4.3).

LEMMA 8.2 *If A_1, B_1' hold and ρ is a Hilbert-Schmidt operator, then the condition*

$$\sum_{j=1}^{k_n} a_j = O(n^{1/4} (\mathrm{Log}\, n)^{-\beta}), \tag{8.60}$$

for some $\beta > \frac{1}{2}$, yields

$$\sum_{j > k_n} \| \rho(v_{jn}) \|^2 \to 0 \quad a.s. \tag{8.61}$$

Proof

Since ρ is Hilbert-Schmidt we have

$$\| \rho \|_{\mathcal{S}}^2 = \sum_{j=1}^{\infty} \| \rho(v_j') \|^2 = \sum_{j=1}^{\infty} \| \rho(v_{jn}) \|^2 < \infty,$$

where $v_j' = sgn < v_{jn}, v_j > v_j$, $j \geq 1$, $n \geq 1$.

Thus

$$\begin{aligned} R_n : \; &= \sum_{j=1}^{k_n} \left(\| \rho(v_j') \|^2 - \| \rho(v_{jn}) \|^2 \right) \\ &= \sum_{j > k_n} \left(\| \rho(v_{jn}) \|^2 - \| \rho(v_j') \|^2 \right) \end{aligned} \tag{8.62}$$

and elementary computations give

$$|R_n| \leq 2 \| \rho \|_{\mathcal{L}}^2 \sum_1^{k_n} \| v_j' - v_{jn} \| .$$

From Lemma 4.3 it follows that

$$|R_n| \leq 2 \| \rho \|_{\mathcal{L}}^2 \left(\sum_{j=1}^{k_n} a_j \right) \| C_n - C \|_{\mathcal{L}}, \tag{8.63}$$

and by using (8.60) we get

$$|R_n| \leq c n^{1/4} (\operatorname{Log} n)^{-\beta} \parallel C_n - C \parallel_{\mathcal{L}},$$

where c is a constant. Then Theorem 4.1 yields

$$R_n \to 0 \quad a.s..$$

Taking in account (8.62) and noting that

$$\sum_{j>k_n} \parallel \rho(v_j') \parallel^2 = \sum_{j>k_n} \parallel \rho(v_j) \parallel^2 \to 0$$

one arrives at (8.61). ■

LEMMA 8.3 *If X is a standard $ARH(1)$ we have the bound*

$$\parallel D_n(v_{jn}) \parallel \leq 2\lambda_{jn}^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} \parallel X_i \parallel^2 \right)^{1/2}, \quad n \geq 2, \ \ j \geq 1. \tag{8.64}$$

Proof

For each integer ℓ we have

$$| < D_n(v_{jn}), \ v_{\ell n} > | = \left| \frac{1}{n-1} \sum_{i=1}^{n-1} < X_i, v_{jn} >< X_{i+1}, v_{\ell n} > \right|$$

$$\leq \frac{1}{n-1} \left(\sum_{i=1}^{n-1} < X_i, v_{jn} >^2 \right)^{1/2} \left(\sum_{i=1}^{n-1} < X_{i+1}, v_{\ell n} >^2 \right)^{1/2}$$

$$\leq \frac{n}{n-1} (< C_n(v_{jn}), \ v_{jn} >)^{1/2} (< C_n(v_{\ell n}), v_{\ell n} >)^{1/2}$$

$$\leq 2\lambda_{jn}^{1/2} \ \lambda_{\ell n}^{1/2}.$$

Therefore

$$\begin{aligned} \parallel D_n(v_{jn}) \parallel^2 &= \sum_{\ell=1}^{\infty} < D_n(v_{jn}), \ v_{\ell n} >^2 \\ &\leq 4\lambda_{jn} \sum_{\ell=1}^{\infty} \lambda_{\ell n}, \end{aligned}$$

but

$$\sum_{\ell=1}^{\infty} \lambda_{\ell n} = \parallel C_n \parallel_{\mathcal{N}} = \frac{1}{n} \sum_{i=1}^{n} \parallel X_i \parallel^2;$$

hence (8.64) holds. ■

Observe that, if $j > n$, (8.64) becomes $D_n(v_{jn}) = 0$.

We now may state two theorems concerning the asymptotic behavior of ρ_n. Their proofs use Lemmas 8.2 and 8.3 together with results concerning $\| C_n - C \|_{\mathcal{L}}$ and $\| D_n - D \|_{\mathcal{L}}$. Since they are technical we postpone them until Section 6.

We first have

THEOREM 8.7 *Suppose that A_1, B_1', and C_1' hold and ρ is Hilbert-Schmidt. Then, if for some $\beta > 1$*

$$\lambda_{k_n}^{-1} \sum_1^{k_n} a_j = O(n^{1/4}(\text{Log } n)^{-\beta}), \tag{8.65}$$

we obtain

$$\| \rho_n - \rho \|_{\mathcal{L}} \longrightarrow 0 \ \ a.s. \tag{8.66}$$

In the bounded case the result is more precise.

THEOREM 8.8 *If the assumptions in Theorem 8.7 hold and if in addition $\| X_0 \|$ is bounded, then*

$$P(\| \rho_n - \rho \|_{\mathcal{L}} \geq \eta) \leq c_1(\eta) \exp \left(-c_2(\eta) n \lambda_{k_n}^2 \left(\sum_1^{k_n} a_j \right)^{-2} \right), \tag{8.67}$$

$\eta > 0$, $n \geq n_\eta$, where $c_1(\eta)$ and $c_2(\eta)$ are positive constants.

Thus $\dfrac{n\lambda_{k_n}^2}{\text{Log } n \left(\sum_1^{k_n} a_j \right)^2} \to \infty$ *implies* $\| \rho_n - \rho \|_{\mathcal{L}} \to 0$ *a.s.*

Example 8.5

If $\lambda_j = ar^j$, $j \geq 1$ $(a > 0,\ 0 < r < 1)$, then $k_n = o(\text{Log } n)$ yields (8.65). If $k_n \simeq \text{Log Log } n$, the bound in (8.67) takes the form

$$P(\| \rho_n - \rho \|_{\mathcal{L}} \geq \eta) \leq c_1(\eta) \exp \left(-c_2'(\eta) \frac{n}{(\text{Log } n)^4} \right)$$

Example 8.6

If $\lambda_j = aj^{-\gamma}$ $(a > 0,\ \gamma > 1)$ and $k_n \simeq \text{Log } n$ then (8.67) becomes

$$P(\| \rho_n - \rho \|_{\mathcal{L}} \geq \eta) \leq c_1(\eta) \exp \left(-c_2''(\eta) \frac{n}{(\text{Log } n)^{4(1+\gamma)}} \right).$$

As in Section 2, assumption of boundedness for $\| X_0 \|$ can be replaced by existence of some exponential moment provided X is geometrically strongly mixing.

Limit in distribution

The problem of weak convergence for ρ_n is rather intricate. We indicate a partial result due to Mas (1999).

THEOREM 8.9 *Suppose that*

- A_1, B_1', *and* C_1' *hold,*
- C_n^{-1} *exists over* $\widetilde{H}_{k_n}$,
- $E \parallel C^{-1}(\varepsilon_0) \parallel^2 < \infty$,
- $n\lambda_{k_n}^4 \to \infty$, $n^{-1}\sum_{j=1}^{k_n} a_j \lambda_j^{-2} \to \infty$,
- $\lambda_j \lambda_{j_n}^{-1}$ *is bounded in probability for each* j.

Then

$$\sqrt{n}\left(\rho_n - \widetilde{\pi}^{k_n}\rho\right) \xrightarrow{\mathcal{D}} N,$$

where the limit is taken in $\mathcal{S}$ *and* N *is a Gaussian random* $\mathcal{S}$*-valued operator.*

8.4 Estimation of autocorrelation operator in $C[0,1]$

We now consider the case where $X = (X_n,\ n \in \mathbb{Z})$ is a standard $ARC(1)$ process (see Section 6.6).

As we will use the spaces $C_0 := C[0,1]$ and $L^2 := L^2([0,1],\ \mathcal{B}_{[0,1]}, \lambda)$, where λ is Lebesgue measure, we will specify notation by putting index C_0 or L^2 on the various norms and scalar products associated with these spaces. Norms in $C[0,1]^2$ are indexed by C_0^2. On the other hand, the natural injection $x \mapsto$ "equivalence class of x" gives sense to $\parallel x \parallel_{L^2}$ and $< x, y >_{L^2}$ for x and y in C_0.

Now suppose that ρ is defined by a continuous **kernel** $r(\cdot,\cdot)$ that satisfies $\parallel r \parallel_{C_0^2} < 1$. We then have

$$\rho(x)(t) = \int_0^1 r(s,t)x(s)ds, \quad x \in C_0, \tag{8.68}$$

and $\parallel \rho \parallel_{\mathcal{L}_{C_0}} < 1$. This is assumption G_1.

Clearly ρ induces a bounded linear operator ρ' over L^2 defined by an analogous formula and such that $\parallel \rho' \parallel_{\mathcal{L}_{L^2}} < 1$.

Assume too that X_0 is **bounded** and satisfies the Hölder condition

$$|X_0(t) - X_0(s)| \leq M_0 |t-s|^{\alpha}, \ 0 \leq s, t \leq 1, \tag{8.69}$$

where M_0 is a bounded real random variable and $0 < \alpha \leq 1$. This is assumption G_2.

Now to X we may associate a standard $ARL^2(1)$ process $X' = (X'_n,\ n \in \mathbb{Z})$, where

$$X'_n = \sum_{j=1}^{\infty} \left(\int_0^1 X_n(s) e_j(s) ds \right) e_j, \quad n \in \mathbb{Z},$$

with $(e_j,\ j \geq 1)$ an orthonormal basis of L^2.

Then

$$X'_n = \rho'(X'_{n-1}) + \varepsilon'_n, \quad n \in \mathbb{Z},$$

where

$$\varepsilon'_n = \sum_{j=1}^{\infty} \left(\int_0^1 \varepsilon_n(s) e_j(s) ds \right) e_j, \quad n \in \mathbb{Z},$$

and (ε_n) is the strong C_0-white noise associated with X. This construction will allow us to use some tools concerning estimation of ρ' to estimate ρ.

Firstly, from the Mercer lemma 1.3 and the above considerations it follows that

$$c(s,t) := E(X_s X_t) = \sum_{j=1}^{\infty} \lambda_j v_j(s) v_j(t), \quad 0 \leq s,\ t \leq 1,$$

where $(\lambda_j, v_j,\ j \geq 1)$ are eigenelements of the covariance operator C of X'. Note that v_j is continuous and therefore bounded.

Secondly, we may define empirical operators associated with X by setting

$$C_n = \frac{1}{n} \sum_{i=1}^{n} X'_i \otimes X'_i$$

and

$$D_n = \frac{1}{n-1} \sum_{i=1}^{n-1} X'_i \otimes X'_{i+1}.$$

Observe that C_n has the continuous kernel

$$c_n(s,t) = \frac{1}{n} \sum_{i=1}^{n} X_i(s) X_i(t), \quad 1 \leq s, t \leq n,$$

which allows it to operate in C_0 as well as in L^2. As usual we will denote by (λ_{jn}, v_{jn}), $j \geq 1$, its eigenelements.

We are now in a position to define the estimator ρ_n of ρ. We suppose that B'_1 and C'_1 hold and put

$$\rho_n(x) = \widetilde{\pi}^{k_n} D_n \widetilde{C}_n^{-1} \widetilde{\pi}^{k_n}(x), \quad x \in C_0, \tag{8.70}$$

where $(k_n) \to \infty$, $\widetilde{C}_n = \widetilde{\pi}^{k_n} C_n$, and

$$\widetilde{\pi}^{k_n}(x) = \sum_{j=1}^{k_n} < x, v_{jn} >_{L^2} v_{jn}, \quad x \in C_0.$$

Compare with (8.59).

The following simple lemma is useful.

LEMMA 8.4

$$\| v_{jn} - v'_j \|_{C_0} \leq \lambda_{jn}^{-1} \| c_n - c \|_{C_0^2} + \| c \|_{C_0^2} \lambda_{jn}^{-1} \| v_{jn} - v'_j \|_{L^2} + \lambda_{jn}^{-1} |\lambda_{jn} - \lambda_j| \; \| v'_j \|_{C_0}, \tag{8.71}$$

where $v'_j = sgn \; < v_{jn}, \; v_j >_{L^2} v_j$.

Proof

Write

$$\begin{aligned}\lambda_{jn}[v_{jn}(t) - v'_j(t)] = & \int_0^1 [c_n(s,t) - c(s,t)] v_{jn}(s) ds \\ & + \int_0^1 c(s,t)[v_{jn}(s) - v'_j(s)] ds \\ & + (\lambda_j - \lambda_{jn}) v'_j(t).\end{aligned}$$

Then (8.71) easily follows. ∎

Let us now introduce the condition G_3:

(a) $v = \sup_{j \geq 1} \| v_j \|_{C_0} < \infty$

and

(b) $$\sup_{\substack{x \in C_0 \\ \| x \|_{C_0} \leq 1}} \| \rho(x) - \sum_{j=1}^{k} < \rho(x), v_j >_{L^2} v_j \|_{C_0} \underset{k \to \infty}{\longrightarrow} 0.$$

Notice that if (a) holds and $\sum_{j \geq k} \| \rho'^*(v_j) \|_{C_0} \underset{k \to \infty}{\longrightarrow} 0$, then (b) holds too.

We may now give a statement that is somewhat similar to Lemma 8.2.

LEMMA 8.5 *If G_3 holds and* $\sum_{j=1}^{k_n} \| v_{jn} - v'_j \|_{C_0} \underset{k\to\infty}{\longrightarrow} 0$ *a.s., then*

$$\sup_{\substack{x \in C_0 \\ \| x \|_{C_0} \le 1}} \| \rho(x) - \sum_{j=1}^{k_n} < \rho(x), v_{jn} >_{L^2} v_{jn} \|_{C_0} \underset{k\to\infty}{\longrightarrow} 0 \ a.s. \quad (8.72)$$

Proof

Consider the decomposition

$$\sum_1^{k_n} < \rho(x), v_{jn} >_{L^2} v_{jn} - \sum_1^{k_n} < \rho(x), v_j >_{L^2} v_j$$
$$= \sum_1^{k_n} < \rho(x), v_{jn} >_{L^2} (v_{jn} - v_j) + \sum_1^{k_n} < \rho(x), v_{jn} - v_j >_{L^2} v_j$$
$$=: \delta_n(x)$$

and note that

$$\sup_{\substack{x \in C_0 \\ \| x \|_{C_0} \le 1}} \| \delta_n(x) \|_{C_0} \le (1+v) \| \rho \|_{\mathcal{L}_{L^2}} \sum_1^{k_n} \| v_{jn} - v'_j \|_{C_0} . \quad (8.73)$$

Then (8.72) follows. ■

We finally state the main result of this section.

THEOREM 8.10 *Let X be a standard $ARC(1)$. If B'_1, C'_1, G_1, G_2, and G_3 hold true, then*

$$P\left(\| \rho_n - \rho \|_{\mathcal{L}_{C_0}} \ge \eta\right) \le d_1(\eta) \exp\left(-d_2(\eta) n \lambda_{k_n}^2 \left(\sum_1^{k_n} a_j \right)^{-2} \right), \quad (8.74)$$

where $d_1(\eta)$ and $d_2(\eta)$ are positive constants.

Thus, if $\dfrac{n\lambda_{k_n}^2}{\mathrm{Log}n(\sum_1^{k_n} a_j)^2} \to 0$, *we have*

$$\| \rho_n - \rho \|_{\mathcal{L}_{C_0}} \to 0 \ a.s. \quad (8.75)$$

Proof is postponed until Section 8.6.

8.5 Statistical prediction

Generalities

Statistical prediction is one of the main purposes of this book.

The general framework is the following. Consider an H-valued stochastic process $X = (X_n,\ n \in \mathbb{Z})$ with distribution P_X belonging to some family $\mathcal{D}$ of probability measures and such that $E_{P_X} \| X_n \|^2 < \infty,\ n \in \mathbb{Z},\ P_X \in \mathcal{D}$, where E_{P_X} denotes expectation with respect to P_X.

P_X is supposed to be unknown but $X_1, \dots, X_n$ are observed. The question is how to construct a good approximation of the nonobserved variable X_{n+h}, where $h \geq 1$. For convenience we take $h = 1$. Results obtained for $h = 1$ easily extend to the general case.

A **predictor** of X_{n+1} has the form

$$\widehat{X}_{n+1} = \varphi_n(X_1, \dots, X_n),$$

where $\varphi_n : H^n \to H$ is a measurable mapping.

If $\widehat{X}_{n+1}$ is in L_H^2, a classical criterion for measuring its accuracy is to use the second order operator of $X_{n+1} - \widehat{X}_{n+1}$ defined as

$$\Gamma_{X_{n+1}-\widehat{X}_{n+1}}(x) = E_{P_X}\left[< X_{n+1} - \widehat{X}_{n+1}, x > (X_{n+1} - \widehat{X}_{n+1})\right],\ x \in H.$$

If $X_{n+1} - \widehat{X}_{n+1}$ is zero-mean this operator coincides with the covariance operator of $X_{n+1} - \widehat{X}_{n+1}$.

Now let $\mathcal{P}_2$ be the class of predictors that belong to L_H^2 for all P_X in $\mathcal{D}$. A preference relation over $\mathcal{P}_2$ is defined as

$$\begin{aligned} \varphi_n \underset{1}{\prec} \psi_n \Leftrightarrow \quad & E_{P_X}(< X_{n+1} - \varphi_n(X_1, \dots, X_n), x >^2) \\ & \leq E_{P_X}\left(< X_{n+1} - \psi_n(X_1, \dots, X_n),\ x >^2\right),\ x \in H, P_X \in \mathcal{D}. \end{aligned}$$

A less precise but more tractable preference relation would be

$$\begin{aligned} \varphi_n \underset{2}{\prec} \psi_n \Leftrightarrow \quad & E_{P_X} \| X_{n+1} - \varphi_n(X_1, \dots, X_n) \|^2 \\ & \leq E_{P_X} \| X_{n+1} - \psi_n(X_1, \dots, X_n) \|^2,\ P_X \in \mathcal{D}. \end{aligned}$$

It is well known that the best predictor with respect to $\underset{1}{\prec}$ or $\underset{2}{\prec}$ is the conditional expectation

$$X_{n+1}^* = E_{P_X}^{\mathcal{T}_n}(X_{n+1}) := \varphi_{0,n}^{(P_X)}(X_1, \dots, X_n),\ P_X \in \mathcal{D},$$

where $\mathcal{T}_n = \sigma(X_1, \dots, X_n)$.

Unflortunately, this is not a **statistical predictor**, since $\varphi_{0,n}^{(P_X)}$ is in general unknown; thus it is impossible to compute $\varphi_{0,n}^{(P_X)}(X_1, \ldots, X_n)$!

Now note that

$$\Gamma_{X_{n+1}-\widehat{X}_{n+1}} = \Gamma_{X_{n+1}-X_{n+1}^*} + \Gamma_{X_{n+1}^*-\widehat{X}_{n+1}}$$

and

$$E_{P_X} \parallel X_{n+1}-\widehat{X}_{n+1} \parallel^2 = E_{P_X} \parallel X_{n+1}-X_{n+1}^* \parallel^2 + E_{P_X} \parallel X_{n+1}^*-\widehat{X}_{n+1} \parallel^2 .$$

Thus the total prediction error is the sum of a **structural prediction error** and a **statistical prediction error**. The statistician can control only the statistical prediction error term. The final problem is to approximate the best predictor X_{n+1}^*.

In this context a sequence $(\varphi_n(X_1, \ldots, X_n),\ n \geq 1)$ of statistical predictors is said to be **consistent** with respect to some stochastic mode of convergence m if

$$X_{n+1}^* - \varphi_n(X_1, \ldots, X_n) \underset{n \to \infty}{\overset{m}{\longrightarrow}} 0.$$

Prediction of *ARH*(1)

If X is a standard $ARH(1)$ we have the basic relation

$$X_{n+1} = \rho(X_n) + \varepsilon_{n+1}, \quad n \in \mathbb{Z},$$

and $\rho(X_n)$ is the best linear prediction of X_{n+1} (see (3.5) and Section 1.6).

It is easy to see that

$$\rho(X_n) = E^{\mathcal{T}_n}(X_{n+1})$$

if and only if ε is a martingale difference with respect to $(\mathcal{T}_n)$.

Thus a natural statistical predictor of X_{n+1} will have the form $\rho_n(X_n)$, where ρ_n is an estimator of ρ. We will say that $\rho_n(X_n)$ is a **functional autoregressive predictor (FAP)**.

Now the asymptotic results derived in the previous sections yield consistency of $\rho_n(X_n)$.

In the general case we have the following.

COROLLARY 8.3

1) If the assumptions in Theorem 8.7 hold, then

$$\parallel \rho_n(X_n) - \rho(X_n) \parallel \underset{p}{\longrightarrow} 0. \tag{8.76}$$

2) If the assumptions in Theorem 8.8 hold, then

$$P(\parallel \rho_n(X_n) - \rho(X_n) \parallel \geq \eta) \leq c_1'(\eta) \exp\left(-c_2'(\eta) n \lambda_{k_n}^2 \left(\sum_1^{k_n} a_j\right)^{-2}\right)$$

and, if $\dfrac{n\lambda_{k_n}^2}{\operatorname{Log} n \left(\sum_1^{k_n} a_j\right)^2} \longrightarrow \infty,$

$$\| \rho_n(X_n) - \rho(X_n) \| \underset{a.s.}{\longrightarrow} 0. \tag{8.77}$$

Proof

1) Observe that, for all $\eta > 0$ and $A > 0$,

$$\begin{aligned} &P(\| \rho_n(X_n) - \rho(X_n) \| \geq \eta) \leq P(\| X_n \| \| \rho_n - \rho \|_{\mathcal{L}} \geq \eta) \\ &\leq P\left(\| \rho_n - \rho \|_{\mathcal{L}} \geq \frac{\eta}{A}\right) + P(\| X_n \| \geq A) \\ &\leq P\left(\| \rho_n - \rho \|_{\mathcal{L}} \geq \frac{\eta}{A}\right) + \frac{E \| X_0 \|^4}{A^4}. \end{aligned}$$

Now let $\zeta > 0$ be arbitrary. Choosing $A = \left(\dfrac{2E \| X_0 \|^4}{\zeta}\right)^{1/4}$ one obtains

$$P(\| \rho_n(X_n) - \rho(X_n) \| \geq \eta) \leq P\left(\| \rho_n - \rho \|_{\mathcal{L}} \geq \frac{\eta}{A}\right) + \frac{\zeta}{2}$$

and, since $\| \rho_n - \rho \|_{\mathcal{L}} \rightarrow 0$ a.s., the bound is less than ζ for large enough n. Hence (8.76).

2) (8.77) comes from

$$\| \rho_n(X_n) - \rho(X_n) \| \leq \| X_0 \|_\infty \| \rho_n - \rho \|_{\mathcal{L}} \rightarrow 0 \ \ a.s. \qquad \blacksquare$$

Sharper results appear if H is finite dimensional. For example, under the conditions in Corollary 8.2 and if $\| X_0 \|$ is bounded, we have

$$E \| \rho_{n,a}(X_n) - \rho(X_n) \|^2 = O\left(\frac{1}{n}\right) \tag{8.78}$$

and, under the conditions in Theorem 8.2,

$$\| \rho_n(X_n) - \rho(X_n) \| = O\left(\left(\frac{\operatorname{Log} n}{n}\right)^{1/2}\right) \ \ a.s. \tag{8.79}$$

We will discuss implementation of ρ_n in Chapter 9.

Prediction in C[0,1]

In $C[0,1]$ the statistical predictor is again $\rho_n(X_n)$, where ρ_n is given by

(8.70).

Under the conditions in Theorem 8.10 we clearly have

$$\| \rho_n(X_n) - \rho(X_n) \|_{C_0} \to 0 \ \ a.s.$$

Prediction of an invertible *LPH*

Here the problem is different since an estimator of R based on $Y_1, \ldots, Y_n$ appears as purely theoretical because these variables are not completely observed (see Section 7.3).

In order to construct such an estimator it is necessary to use the observed variables $(Z_{i,p},\ 1 \leq i \leq n)$ defined in Section 7.4. The associated estimator, say $R_{n,p}$, is similar to ρ_n and defines an operator over $H^p \times \{0\} \times \{0\} \times \ldots$ that can be obviously extented to H_w.

Finally, consistency of $R_{n,p}$ may be obtained through a suitable choice of $p = p_n$ combined with a satisfactory choice of k_n. Such a study has been performed by Merlevède (1996). It leads to a consistent predictor R_{n,p_n} (Z_{n,p_n}), but computations are tedious and therefore omitted.

8.6 * Derivation of strong consistency

This section is devoted to the proofs of Theorems 8.7 and 8.8 in Section 3 and Theorem 8.10 in Section 4.

Proof of Theorem 8.7

Consider the decomposition

$$\begin{aligned}(\rho_n - \rho)(x) &= \left[\rho_n \widetilde{\pi}^{k_n}(x) - \rho\pi^{k_n}(x)\right] \\ &\quad + \left[\rho\pi^{k_n}(x) - \rho\widetilde{\pi}^{k_n}(x)\right] \\ &\quad + \left[\rho\widetilde{\pi}^{k_n}(x) - \rho(x)\right] \\ &:= a_n(x) + b_n(x) + c_n(x)\end{aligned} \tag{8.80}$$

and put

$$\alpha_n = \sup_{\|x\|\leq 1} \| a_n(x) \|, \quad \beta_n = \sup_{\|x\|\leq 1} \| b_n(x) \|, \quad \gamma_n = \sup_{\|x\|\leq 1} \| c_n(x) \| .$$

First we have

$$a_n(x) = \widetilde{\pi}^{k_n} D_n \left(\sum_1^{k_n} \lambda_{jn}^{-1} < x, v_{jn} > v_{jn} \right) - D \left(\sum_1^{k_n} \lambda_j^{-1} < x, v'_j > v'_j \right),$$

where $v_j' = sgn < v_{jn}, v_j > v_j$. This can be written,

$$a_n(x) = \sum_{i=1}^{4} a_{ni}(x) \tag{8.81}$$

where

$$\begin{aligned} a_{n1}(x) &= D_n \left(\sum_1^{k_n} \left(\lambda_{jn}^{-1} - \lambda_j^{-1}\right) < x, v_{jn} > v_{jn} \right), \\ a_{n2}(x) &= D_n \left(\sum_1^{k_n} \lambda_j^{-1} (< x, v_{jn} > - < x, v_j' >) v_{jn} \right), \\ a_{n3}(x) &= D_n \left(\sum_1^{k_n} \lambda_j^{-1} < x, v_j' > (v_{jn} - v_j') \right), \end{aligned}$$

and

$$a_{n4}(x) = (D_n - D) \left(\sum_1^{k_n} \lambda_j^{-1} < x, v_j' > v_j' \right).$$

We set

$$\alpha_{ni} = \sup_{\|x\| \le 1} \| a_{ni}(x) \|, \quad 1 \le i \le 4.$$

We now have

$$\| a_{n1}(x) \| \le \sum_1^{k_n} \frac{|\lambda_{jn} - \lambda_j|}{\lambda_{jn}\lambda_j} | < x, v_{jn} > | \, \| D_n(v_{jn}) \| .$$

Using (4.43), Lemma 8.3, and the Cauchy-Schwarz inequality we obtain the bound

$$\alpha_{n1} \le 2 \left(\frac{1}{n} \sum_1^n \| X_i \|^2 \right)^{1/2} \lambda_{k_n}^{-1} \lambda_{k_n,n}^{-1/2} k_n^{1/2} \| C_n - C \|_{\mathcal{L}} . \tag{8.82}$$

From (8.65) and Theorem 4.1 we may deduce that $\| C_n - C \|_{\mathcal{L}} \le \frac{\lambda_{k_n}}{2}$ for large enough n (a.s.). This implies

$$\lambda_{k_n,n} \ge \lambda_{k_n} - \| C_n - C \|_{\mathcal{L}} \ge \frac{\lambda_{k_n}}{2};$$

thus

$$\alpha_{n1} \le 2\sqrt{2} \left(\frac{1}{n} \sum_1^n \| X_i \|^2 \right)^{1/2} \lambda_{k_n}^{-3/2} k_n^{1/2} \| C_n - C \|_{\mathcal{L}} . \tag{8.83}$$

Using Theorem 4.1 again one gets

$$\lambda_{k_n}^{-3/2} k_n^{1/2} \parallel C_n - C \parallel_{\mathcal{L}} \to 0 \ \ a.s.,$$

and since Corollary 6.2 gives

$$\frac{1}{n} \sum_1^n \parallel X_i \parallel^2 \ \underset{a.s.}{\longrightarrow} \ E \parallel X_0 \parallel^2$$

it follows that

$$\alpha_{n1} \to 0 \ \ a.s. \tag{8.84}$$

Concerning α_{n2}, Lemma 8.3 yields the bound

$$\alpha_{n2} \leq 2 \left(\frac{1}{n} \sum_1^n \parallel X_i \parallel^2 \right)^{1/2} \sum_1^{k_n} \lambda_j^{-1} \lambda_{jn}^{1/2} \parallel v_{jn} - v_j' \parallel$$

and Lemma 4.3 implies

$$\alpha_{n2} \leq 2 \left(\frac{1}{n} \sum_1^n \parallel X_i \parallel^2 \right)^{1/2} \left(\sum_1^{k_n} \lambda_j^{-1} \lambda_{jn}^{1/2} a_j \right) \parallel C_n - C \parallel_{\mathcal{L}} .$$

As above, it can be checked that $\lambda_{1n} \leq \dfrac{3\lambda_1}{2}$ for large enough n (a.s.). Then

$$\alpha_{n2} \leq \sqrt{6} \left(\frac{1}{n} \sum_1^n \parallel X_i \parallel^2 \right)^{1/2} \lambda_1^{1/2} \lambda_{k_n}^{-1} \left(\sum_1^{k_n} a_j \right) \parallel C_n - C \parallel_{\mathcal{L}} . \tag{8.85}$$

Using Corollary 6.2 again, together with Theorem 4.1, and taking in account (8.65) one obtains

$$\alpha_{n2} \to 0 \ \ a.s. \tag{8.86}$$

Turning to α_{n3}, we have

$$\alpha_{n3} \leq \parallel D_n \parallel_{\mathcal{L}} \left(\sum_2^{k_n} a_j \right) k_n^{1/2} \lambda_{k_n}^{-1} \parallel C_n - C \parallel_{\mathcal{L}} \tag{8.87}$$

and since $\parallel D_n \parallel_{\mathcal{L}} \ \underset{a.s.}{\longrightarrow} \ \parallel D \parallel_{\mathcal{L}}$ (Theorem 4.8) we easily get

$$\alpha_{n3} \to 0 \ \ a.s. \tag{8.88}$$

Finally,

$$\parallel a_{n4}(x) \parallel \leq \parallel D_n - D \parallel_{\mathcal{L}} \left(\sum_1^{k_n} \lambda_j^{-2} < x, v_j' >^2 \right)^{1/2} ;$$

thus

$$\alpha_{n4} \leq \| D_n - D \|_{\mathcal{L}} \quad \lambda_{k_n}^{-1}. \tag{8.89}$$

Then Theorem 4.8 and (8.65) entail

$$\alpha_{n4} \to 0 \ \ a.s. \tag{8.90}$$

From (8.84), (8.86), (8.88), and (8.90) it follows that

$$\alpha_n \to 0 \quad a.s. \tag{8.91}$$

We turn to β_n and γ_n. First note that

$$\beta_n^2 \leq 2 \left[\sum_{j>k_n} \| \rho(v_{jn}) \|^2 + \sum_{j>k_n} \| \rho(v_j) \|^2 \right], \tag{8.92}$$

which tends to zero by Lemma 8.2.

For γ_n we have

$$\gamma_n \leq \sum_{j>k_n} \| \rho(v_{jn}) \|^2 \to 0 \ \ a.s., \tag{8.93}$$

again by Lemma 8.2, and the proof of Theorem 8.7 is now complete. ■

Proof of Theorem 8.8

Write

$$\begin{aligned} (\rho_n - \rho)(x) \quad &= (\rho_n - \rho)\left(x - \widetilde{\pi}^{k_n}(x)\right) + (\rho_n - \rho)\left(\widetilde{\pi}^{k_n}(x)\right) \\ &:= A_n(x) + B_n(x). \end{aligned}$$

Since $\rho_n \left(x - \widetilde{\pi}^{k_n}(x)\right) = 0$, we have

$$\| (\rho_n - \rho)(x - \widetilde{\pi}^{k_n}(x)) \| = \| \rho \left(\sum_{j>k_n} < x, v_{jn} > v_{jn} \right) \|$$

$$\leq \sum_{j>k_n} | < x, v_{jn} > | \, \| \rho(v_{jn}) \| \leq \left(\sum_{j>k_n} \| \rho(v_{jn}) \|^2 \right)^{1/2},$$

for all x such that $\| x \| \leq 1$.

Now (8.62) and (8.63) yield

$$\sum_{j>k_n} \| \rho(v_{jn}) \|^2 \leq \sum_{j>k_n} \| \rho(v_j') \|^2 + 2 \| \rho \|_{\mathcal{L}}^2 \left(\sum_1^{k_n} a_j \right) \| C_n - C \|_{\mathcal{L}} \, . \tag{8.94}$$

Then, given $\eta > 0$, there exists an integer n_η such that $n \geq n_\eta$ implies

$$\sum_{j>k_n} \| \rho(v'_j) \|^2 = \sum_{j>k_n} \| \rho(v_j) \|^2 < \frac{\eta^2}{4},$$

which in turn implies

$$P\left(\left(\sum_{j>k_n} \| \rho(v_{jn}) \|^2\right)^{1/2} \geq \eta\right) \leq P\left(\| \rho \|^2_{\mathcal{L}} \left(\sum_1^{k_n} a_j\right) \| C_n - C \|_{\mathcal{L}} \geq \frac{\eta^2}{4}\right)$$

and, by Theorem 4.2 and for large enough n,

$$P\left(\sup_{\|x\|\leq 1} \| A_n(x) \| \geq \eta\right) \leq 4 \exp\left(-cn\left(\sum_1^{k_n} a_j\right)^{-2}\right),$$

where c is a positive scalar depending on η.

Now

$$B_n(x) = a_n(x) + b_n(x),$$

where a_n and b_n are defined in the proof of Theorem 8.7 and where $a_n = \sum_1^4 a_{ni}$.

Since there exists a constant K such that $\| X_0 \| \leq K$, (8.82) gives

$$\alpha_{n1} \leq 2K\lambda_{k_n}^{-1}\lambda_{k_n,n}^{-1/2}k_n^{1/2} \| C_n - C \|_{\mathcal{L}} .$$

Noting that

$$\begin{aligned} P(\alpha_{n_1} \geq \eta) \quad &= \quad P\left(\alpha_{n1} \geq \eta, \| C_n - C \|_{\mathcal{L}} \geq \frac{\lambda_{k_n}}{2}\right) \\ &\quad + P\left(\alpha_{n1} \geq \eta, \| C_n - C \|_{\mathcal{L}} < \frac{\lambda_{k_n}}{2}\right) \end{aligned}$$

and that $\| C_n - C \|_{\mathcal{L}} < \frac{\lambda_{k_n}}{2}$ entails $\lambda_{k_n,n} > \frac{\lambda_{k_n}}{2}$, we arrive at

$$\begin{aligned} P(\alpha_{n1} \geq \eta) \quad &\leq \quad P\left(2\sqrt{2}K\lambda_{k_n}^{-3/2}k_n^{1/2} \| C_n - C \|_{\mathcal{L}} \geq \eta\right) \\ &\quad + P\left(\| C_n - C \|_{\mathcal{L}} \geq \frac{\lambda_{k_n}}{2}\right). \end{aligned}$$

Using Theorem 4.2 again we obtain, for large enough n,

$$P(\alpha_{n1} \geq \eta) \leq 4 \exp\left(-c_1 n k_n^{-1} \lambda_{k_n}^3\right) + 4 \exp\left(-c_2 n \lambda_{k_n}^2\right).$$

Now, since $\lambda_j \downarrow 0$ and $\Sigma\lambda_j < \infty$, we have $k_n\lambda_{k_n} \le 1$ for large enough n; therefore

$$P(\alpha_{n1} \ge \eta) \le 8\exp\left(-c_3 n\lambda_{k_n}^4\right), \tag{8.95}$$

where $c_3 > 0$ is a constant.

Deviations of α_{n2} and α_{n3} are majorized similarly and Theorem 4.8 provides a bound concerning α_{n4}. We obtain

$$P(\alpha_{n2} \ge \eta) \le c\exp\left(-c_4 n\lambda_{k_n}^2\left(\sum_1^{k_n} a_j\right)^{-2}\right), \tag{8.96}$$

$$P(\alpha_{n3} \ge \eta) \le c\exp\left(-c_5 n\lambda_{k_n}^2\left(\sum_1^{k_n} a_j\right)^{-2}\right), \tag{8.97}$$

$$P(\alpha_{n4} \ge \eta) \le c\exp\left(-c_6 n\,\lambda_{k_n}^2\right), \tag{8.98}$$

where c, c_4, c_5, and c_6 are positive constants.

Finally, β_n is studied by using (8.92) to obtain

$$P(\beta_n \ge \eta) \le c\exp\left(-c_7 n\left(\sum_1^{k_n} a_j\right)^{-2}\right). \tag{8.99}$$

The desired result follows by collecting all the above bounds and by noticing that

$$\sum_1^{k_n} a_j \ge \lambda_{k_n}^{-1}. \tag{8.100}$$

■

Proof of Theorem 8.10

We again use decompositions (8.80) and (8.81) and set

$$\alpha_{ni}^{C_0} = \sup_{\substack{x\in C_0\\ \|x\|_{C_0}\le 1}} \|\, a_{ni}(x)\,\|_{C_0}\ ;\ \ i = 1,2,3,4.$$

Noting that $\|\, x\,\|_{L^2} \le \|\, x\,\|_{C_0}$, it is easy to see that the bounds (8.95), (8.96), (8.97), and (8.98) remain respectively valid for $\alpha_{ni}^{C_0}$, $i = 1,2,3,4$, so that

$$P\Big(\sup_{\substack{x\in C_0\\ \|x\|_{C_0}\le 1}} \|\, a_n(x)\,\|_{C_0} \ge \eta\Big) \le \gamma_1(\eta)\exp\Big(-\gamma_2(\eta)n\lambda_{k_n}^2\Big(\sum_1^{k_n} a_j\Big)^{-2}\Big),\ \eta > 0,$$

where $\gamma_1(\eta)$ and $\gamma_2(\eta)$ are positive constants.

Concerning b_n we have

$$b_n(x) = \sum_1^{k_n} < x, v'_j - v_{jn} >_{L^2} \rho(v'_j) + \sum_1^{k_n} < x, v_{jn} >_{L^2} \rho(v'_j - v_{jn}).$$

Thus

$$\begin{aligned} \| b_n(x) \|_{C_0} \leq \; & \sum_1^{k_n} \left| < x, v'_j - v_{jn} >_{L^2} \right| \| \rho(v'_j) \|_{C_0} \\ & + \sum_1^{k_n} | < x, v_{jn} >_{L^2} | \, \| \rho(v'_j - v_{jn}) \|_{C_0} \end{aligned}$$

and, since $\| x \|_{L^2} \leq \| x \|_{C_0}$ and $V = \sup_{j \geq 1} \| v'_j \|_{C_0} < \infty$, we obtain

$$\begin{aligned} \beta'_n := \sup_{\substack{x \in C_0 \\ \|x\|_{C_0} \leq 1}} \| b_n(x) \|_{C_0} \leq \; & V \| \rho \|_{\mathcal{L}_{C_0}} \sum_1^{k_n} \| v'_j - v_{jn} \|_{L^2} \\ & + \| \rho \|_{\mathcal{L}_{C_0}} \sum_1^{k_n} \| v'_j - v_{jn} \|_{C_0} . \end{aligned}$$

Now Lemma 8.4 and Theorem 6.15 allow us to bound deviations of β'_n. Some easy computations give

$$P(\beta'_n \geq \eta) \leq \gamma_3(\eta) \exp\left(-\gamma_4(\eta) n \lambda_{k_n}^4\right),$$

with $\gamma_3(\eta) > 0$, $\gamma_4(\eta) > 0$.

It remains to study $c_n(x)$. Using (8.72) and Lemma 8.4 we get

$$P(\sup_{\substack{x \in C_0 \\ \|x\|_{C_0} \leq 1}} \| c_n(x) \|_{C_0} \geq \eta) \leq \gamma_5(\eta) \exp\left(-\gamma_6(\eta) n \lambda_{k_n}^4\right),$$

with $\gamma_5(\eta)$ and $\gamma_6(\eta) > 0$, and the proof of Theorem 8.10 is now complete.

■

NOTES

Most of the results in this Chapter are new or are improvements of those in Bosq (1991-a) and Pumo (1999).

Theorem 8.9 is due to Mas (1999-a).

9

Implementation of Functional Autoregressive Predictors and Numerical Applications

In this final chapter we discuss implementation of predictors based on the ARB model and give examples of applications in various fields.

The first section deals with interpolation and smoothing of discrete functional data. In particular, spline smoothing is considered. For a study of functional data one may consult Ramsay and Silverman (1997).

Section 2 is devoted to identification and estimation of $ARH(p)$, while methods of prediction in continuous time are presented in Section 3.

Various numerical results appear in Section 4: simulations, prediction of electricity consumption, prediction of road traffic, climatic prediction (*el niño*), and prediction of electrocardiograms.

9.1 Functional data

In many different fields continuous-time processes provide convenient descriptions of mechanisms that generate observed systems.

If the system admits a natural seasonal component, the associated process may be interpreted as a sequence of random variables in a function space, as indicated in Section 2.1. The same interpretation takes place if the process is stationary, but in that case the function space is not canonical (see Example 6.2).

Now, if data are collected in **continuous time** within this framework, they constitute a finite sequence of curves indexed by **time-intervals of equal length**. These intervals may be **adjacent**, **disjoint**, or even **overlapping**.

For example, electricity consumption recorded in continuous time for one year generates 365 (or 366) adjacent time intervals. The same variable observed each Sunday generates 52 (or 53) disjoint intervals. Overlapping intervals can be used to counterbalance slight variations of "period." Physiological signals like electrocardiograms suffer this sort of variation.

Error in continuous time is often due to reaction delay of measuring instruments. If (y_t) is a recording of (x_t) (a realization of the underlying process (ξ_t)), this error can be characterized by writing

$$y_t = \int_{-\infty}^{+\infty} x_s \frac{1}{\delta} w\left(\frac{t-s}{\delta}\right) ds, \tag{9.1}$$

where the weight-function w is positive, bounded, with compact support, and such that $\int_{-\infty}^{+\infty} w(u)du = 1$; δ is positive.

Then, if x satisfies the Hölder condition

$$|x_t - x_s| \leq c|t-s|^\alpha, \tag{9.2}$$

where c is constant and $\alpha \in]0,1]$, we have

$$\left(\int_a^b |y_t - x_t|^2 dt\right)^{1/2} \leq (b-a) \sup_{t\in[a,b]} |y_t - x_t| \leq c\delta^\alpha \int_{-\infty}^{+\infty} |u|^\alpha w(u)du, \ a < b. \tag{9.3}$$

Thus L^2 and uniform errors are $O(\delta^\alpha)$.

In many situations functional data are observed in **discrete time** according to the scheme

$$y_j = x_{t_j} + \eta_j, \tag{9.4}$$

where y_j is a recording of x_{t_j} and η_j is the observational global error.

Another possible scheme is

$$y_{t_j} = x_{t_j} + \eta_{t_j}, \tag{9.5}$$

which means that η_{t_j} comes from some white-noise process.

Interpolation and smoothing

Interpolation and smoothing are techniques for converting discrete data into functional data.

Interpolation is used if, in (9.4), η_j may be considered as negligible. In Section 6.6 we saw that **linear interpolation** of (t_j, x_{t_j}) gives an $O(\delta^\alpha)$ error provided that (x_t) satisfies (9.2). **Spline interpolation** supplies smooth curves. A common choice is **cubic spline** : between two "knots" t_j

and t_{j+1}, a cubic spline is a polynomial of degree 3. At each knot two contiguous polynomials are required to have the same values of their derivatives of orders 1 and 2.

An analysis of **smoothing** arises in the book quoted above. Here we deal only with **spline smoothing** since it will be used in some numerical examples below.

Consider the model

$$y_j = x_{t_j} + \eta_j, \quad 1 \leq j \leq q, \quad 0 \leq t_1 < \ldots < t_q \leq 1, \tag{9.6}$$

where (x_{t_j}) is a discrete piece of $(x_t,\ 0 \leq t \leq 1)$.

Suppose that $t \mapsto x_t$ belongs to the Sobolev space

$$W^2[0,1] = \left\{ f : [0,1] \mapsto \mathbb{R}\ ;\ f,\ f' \text{ absolutely continuous}\ ;\ \int_0^1 [f''(t)]^2 dt < \infty \right\},$$

equipped with scalar product

$$< f, g > = \int_0^1 f(t)g(t)dt + \int_0^1 f''(t)g''(t)dt\ ;\ f, g \in W^2.$$

The spline approximation of x, say $\widehat{x}$, is the solution of the optimization problem

$$\min_{f \in W^2[0,1]} \left\{ \frac{1}{q} \sum_{j=1}^{q} (f(t_j) - y_j)^2 + \ell \int_0^1 [f''(t)]^2 dt \right\}, \tag{9.7}$$

where ℓ is a smoothing parameter.

If $\ell \to 0$, $\widehat{x}$ approaches the true function as much as possible when, if $\ell \to \infty$, it approximates linear regression. So the choice of ℓ is a compromise between regularity of $\widehat{x}$ and closeness to x.

Concerning accuracy, Ragozin (1983) has shown that, if $\max(t_j - t_{j-1}) = O(q^{-1})$, then

$$\int_0^1 (\widehat{x}(t) - x(t))^2 dt \leq c_0(q^{-4} + \ell) \int_0^1 [x''(t)]^2 dt, \tag{9.8}$$

where the constant c_0 does not depend on x.

Note that the above method can be employed even if x is not in $W^2[0,1]$, the aim being to obtain a tractable approximation of x.

Now in the context of functional autoregressive processes it is interesting to use a variant of spline smoothing. This method, introduced by Besse and Cardot (1997), is **spline smoothing with rank constraint** (SSRC).

Suppose that $X_1, \ldots, X_n$ are random functions defined on $[0,1]$ and observed at $t_1, \ldots, t_q$ $(0 \leq t_1 < \ldots < t_q \leq 1)$. We thus obtain nq observations

y_{ij}, $1 \le i \le n$, $1 \le j \le q$. The problem is to find approximations of $X_1, \ldots, X_n$ that lie in G_{k_n} a k_n-dimensional subspace of $W^2[0,1]$. This leads to the optimization problem

$$\min_{f_1,\ldots,f_n \in G_{k_n}} \left\{ \frac{1}{nq} \sum_{i=1}^{n} \left[\sum_{j=1}^{q} (y_{ij} - f_i(t_j))^2 + \ell \int_0^1 [f_i''(t)]^2 dt \right] \right\}. \tag{9.9}$$

This method is employed in example 9.5 (Section 4) in order to anticipate rank reduction of the autocorrelation operator estimate.

Finally by using spline or another method (Fourier series, Wavelets, etc.) one arrives at n smooth curves, which are realizations of functional random variables $\widehat{X}_1, \ldots, \widehat{X}_n$, which in turn are approximations of the "true" random variables $X_1, \ldots, X_n$.

9.2 Choosing and estimating a model

The infinite-dimensional linear process studied in this book covers a great variety of situations in continuous and discrete time ; see examples 3.3, 3.4, 6.2, 6.3, 6.4, 8.2, etc. However, note that in all its generality this model may be considered as **nonparametric** because parameter is infinite dimensional.

In practice, $ARH(p)$ and $ARB(p)$ processes will suffice to fit functional data adequately.

If sample paths are continuous, $ARC(p)$ and $ARL^2(p)$ are convenient : computations may be performed within the $ARL^2(p)$ model while the $ARC(p)$ structure ensures asymptotic properties with respect to uniform norm.

Before using the previous models it is sometimes necessary to eliminate the **trend**. This operation can be performed by discrete differentiation, that is, by substituting $y_j - y_{j-1}$ for y_j. Details and survey of other methods for eliminating trend appear in Brockwell and Davis (1994). Note that removing trend must be achieved prior to interpolation or smoothing.

Concerning **seasonality**, removing is in general injudicious, owing to the fact that seasonal component is genuinely incorporated into the model (see again example (6.4)).

Another precaution consists in testing independence of $\widehat{X}_1, \ldots, \widehat{X}_n$.

Testing independence

For convenience suppose that $\widehat{X}_1, \ldots, \widehat{X}_n$ are observed, with $n = 2p$, and consider the statistic

$$\widehat{\Delta}_n = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \widehat{X}_{2i-1} \otimes \widehat{X}_{2i}.$$

Take as null hypothesis

H_0 : (X_n) is a strong H-white noise and $E \parallel X_0 \parallel^4 < \infty$.

Then, under H_0, Theorem 2.7 applied to $(X_{2i-1} \otimes X_{2i},\ i \geq 1)$ yields

$$\Delta_n \xrightarrow{\mathcal{D}} N,$$

where $\Delta_n = \dfrac{1}{\sqrt{p}} \sum_{i=1}^{p} X_{2i-1} \otimes X_{2i}$ and $N \sim \mathcal{N}(0, \Gamma)$. Convergence takes place in the space $\mathcal{S}$ of Hilbert-Schmidt operators.

Γ has spectral decomposition

$$\Gamma = \sum_{\ell,k} \lambda_\ell \lambda_k v_\ell \otimes v_k,$$

where as usual (λ_j, v_j), $j \geq 1$ are the eigenelements of C (see Section 4.2).

Therefore

$$\parallel \Delta_n \parallel_{\mathcal{S}}^2 \xrightarrow{\mathcal{D}} \sum_{\ell,k} \lambda_\ell \lambda_k \xi_{\ell,k}^2,$$

where $(\xi_{\ell,k})$ is a family of independent $\mathcal{N}(0,1)$ random variables.

Now take as alternative

H_1 : (X_n) is a standard $ARH(p)$ associated with a strong H-white noise, $E(X_1 \otimes X_2) = D \neq 0$.

Then, using Theorem 5.11, it is easy to see that

$$\parallel \Delta_n \parallel_{\mathcal{S}}^2 \underset{p}{\longrightarrow} +\infty.$$

Hence the test of critical region:

$$\parallel \Delta_n \parallel_{\mathcal{S}}^2 > c.$$

In order to determine c one can use approximation, replacing $\sum_{\ell,k} \lambda_\ell \lambda_k \xi_{\ell k}^2$ with $\sum_{1 \leq \ell, k \leq m_n} \widehat{\lambda}_{\ell n} \widehat{\lambda}_{kn} \xi_{\ell k}^2$, where $(\widehat{\lambda}_{jn},\ j \geq 1)$ are eigenvalues of the empirical covariance operator associated with $\widehat{X}_1, \ldots, \widehat{X}_n$.

If independence is not satisfied, one may choose the $ARH(p)$ model with $p \geq 1$. This model is identified through **estimation of p.** A partial theoretical solution of this problem is supplied in Section 5.5.

A general heuristic method is selection by **predictive ability**. It is based on estimation of $\sigma^2 = \parallel C_\varepsilon \parallel_{\mathcal{N}}$ by $\widehat{\sigma}_n^2$, where $\widehat{\sigma}_n^2$ is defined by (8.27) but constructed from $\widehat{X}_1, \ldots, \widehat{X}_n$ instead of $X_1, \ldots, X_n$. In fact to each p one associates an estimator $\widehat{\sigma}_n^2(p)$ of the prediction error $\sigma^2(p)$ corresponding to the choice $ARH(p)$ (or equivalently $ARH^p(1)$). The resulting estimator $\widehat{p}_n$ of p is given by

$$\widehat{p}_n = \underset{1 \leq p \leq p_{\max}}{\arg\min} \ \widehat{\sigma}_n^2(p), \tag{9.10}$$

where $p_{\max}$ is selected by the practitioner.

We now turn to **parameter estimation**. Since $ARH(p)$ processes are equivalent to $ARH^p(1)$ processes we may deal with the basic $ARH(1)$ model. For ease we also suppose that $\widehat{X}_1, \ldots, \widehat{X}_n$ belong to $C[0,1]$.

The model is written under the appropriate form

$$X_n(t) - \mu(t) = [\rho(X_{n-1} - \mu)](t) + \varepsilon_n(t), \;\; 0 \le t \le 1, \; n \in \mathbb{Z}. \tag{9.11}$$

The natural estimator of μ is

$$\frac{\widehat{S}_n(t)}{n} = \frac{1}{n} \sum_{i=1}^{n} \widehat{X}_i(t) \;, \;\; 0 \le t \le 1, \tag{9.12}$$

and its L^2-error fulfills

$$\begin{aligned} E \int_0^1 \left(\frac{\widehat{S}_n(t)}{n} - \mu(t) \right)^2 dt \; & \le 2E \int_0^1 \left(\frac{\widehat{S}_n(t) - S_n(t)}{n} \right)^2 dt \\ & + 2E \int_0^1 \left(\frac{S_n(t)}{n} - \mu(t) \right)^2 dt. \end{aligned} \tag{9.13}$$

If the smoother is the spline described by (9.7), one may apply (3.30) and (9.8) to obtain the bound

$$E \int_0^1 \left(\frac{\widehat{S}_n(t)}{n} - \mu(t) \right)^2 dt = O(q^{-4} + \ell + n^{-1}), \tag{9.14}$$

and optimal rate is achieved if $q^{-1} = O(n^{-1/4})$ and $\ell = O(n^{-1})$.

The empirical covariance function has the form

$$\widehat{c}_n(s,t) = \frac{1}{n} \sum_{i=1}^{n} (\widehat{X}_i(s) \; - \; \frac{\widehat{S}_n(s)}{n})(\widehat{X}_i(t) \; - \; \frac{\widehat{S}_n(t)}{n}), \;\; 0 \le s,t \le 1, \tag{9.15}$$

and **the empirical cross-covariance function** is

$$\widehat{d}_n(s,t) = \frac{1}{n-1} \sum_{i=1}^{n-1} (\widehat{X}_i(s) - \frac{\widehat{S}_n(s)}{n})(\widehat{X}_{i+1}(t) - \frac{\widehat{S}_n(t)}{n}), \;\; 0 \le s,t \le 1. \tag{9.16}$$

If the process (X_n) is bounded, we have

$$E \int_{[0,1]^2} [\widehat{c}_n(s,t) - c(s,t)]^2 \, dsdt = O\left(q^{-4} + \ell + \frac{1}{n} \right), \tag{9.17}$$

where $c(s,t) = E(X_0(s)X_0(t))$; a corresponding result takes place for $\widehat{d}_n$.

From $\widehat{c}_n$ and $\widehat{d}_n$ it is now possible to construct an approximation, say $\widehat{\rho}_n$, of ρ_n defined by (8.59). For a given k_n we have

$$\widehat{\rho}_n = \widehat{\pi}^{k_n} \widehat{D}_n \widehat{C}_n^{-1} \widehat{\pi}^{k_n}, \tag{9.18}$$

where $\widehat{\pi}^{k_n}$ is the orthogonal projector of $\mathrm{sp}\,(\widehat{v}_{1n}, \ldots, \widehat{v}_{k_n n})$.

$\widehat{v}_{1n}, \ldots, \widehat{v}_{k_n n}$ are eigenvectors of $\widehat{C}_n$ associated with the eigenvalues $\widehat{\lambda}_{1n}, \ldots, \widehat{\lambda}_{k_n n}$ $\left(\widehat{\lambda}_{1n} \geq \ldots \geq \widehat{\lambda}_{k_n n}\right)$. Determination of $\widehat{\rho}_n$ is classical since it is based on matrix calculus and computation of eigenelements. Note that if $\widehat{\lambda}_{k_n,n}$ is too small it is opportune to replace $\widehat{\rho}_n$ by an estimator $\widehat{\rho}_{n,a}$ similar to $\overline{\rho}_{n,a}$ (see (8.19) in Section 8.1).

A **predictor** of X_{n+1} follows by computing $\widehat{\rho}_n(\widehat{\boldsymbol{X}_n})$.

Clearly **determination of k_n** is crucial in this context. If k_n is too small, estimation of ρ is not accurate enough; if it is too large, prediction error should explode (see Example 9.6). Since this drawback is more serious than lack of accuracy, it seems preferable to choose a moderate-sized k_n.

Examples 8.5 and 8.6 give some ideas about a reasonable magnitude of k_n.

Now an automatic choice of k_n may be performed by using validation. Consider the **empirical prediction error** defined as

$$\widehat{\Delta}_n(k) = \frac{1}{n - n_0} \sum_{\nu=n_0}^{n-1} \| \widehat{\rho}_\nu^{(k)} \left(\widehat{X}_\nu\right) - \widehat{X}_{\nu+1} \|^2, \tag{9.19}$$

where $\widehat{\rho}_\nu^{(k)}$ is the estimator of ρ associated with $\widehat{X}_1, \ldots, \widehat{X}_\nu$ and the choice $k_\nu = k$. The variable n_0 is a given large enough integer (for example $n_0 = \left[\frac{n}{2}\right]$).

Then the choice of k_n is

$$\widehat{k}_n = \underset{1 \leq k \leq k_{\max}}{\arg\min} \ \widehat{\Delta}_n(k), \tag{9.20}$$

where $k_{\max}$ is at one's disposal. Note analogy with (9.10). Some partial theoretical results concerning $\widehat{k}_n$ appear in Pumo (1992).

Finally, the spline smoothing with rank constraint defined by (9.9) combines smoothing and validation as shown in Example 9.4.

9.3 Statistical methods of prediction

In this section we give an account of some well-known prediction methods. Empirical comparisons with functional autoregressive predictors are presented in the next section.

We first consider statistical predictors associated with a discrete-time real strictly stationary process $\xi = (\xi_t,\ t \in \mathbb{Z})$. The variables $\xi_1, \ldots, \xi_n$ are observed and a forecast of ξ_{n+1} is needed.

Naïve predictor

The naïve predictor of ξ_{n+1} is simply ξ_n (the last observation). It can be useful if ξ is associated with a phenomenon that varies only from time to time, like weather. From a theoretical point of view this predictor is optimal if ξ_n is a martingale, since in this case

$$E^{\sigma(\xi_n, \xi_{n-1}, \ldots)}(\xi_{n+1}) = \xi_n,$$

a very unusual property for a stationary process!

Sample mean

An unsophisticated predictor of ξ_{n+1} is the sample mean

$$\overline{\xi}_n = \frac{\xi_1 + \ldots + \xi_n}{n}. \tag{9.21}$$

It is optimal with respect to quadratic error if ξ is i.i.d. Gaussian but poor in a general situation because it does not take advantage of inner-correlation of ξ.

$AR(p)$ predictor

Let ξ be a real $AR(p)$ defined by

$$\xi_n = \rho_1 \xi_{n-1} + \ldots + \rho_p \xi_{n-p} + \varepsilon_n, \quad n \in \mathbb{Z},$$

where $\rho_1, \ldots, \rho_p$ are scalars and (ε_n) is a strong real white noise.

Then a natural statistical predictor is

$$\widehat{\xi}_{n+1}^{(p)} = \widehat{\rho}_1 \xi_n + \ldots + \widehat{\rho}_p \xi_{n-p+1}\ , \quad (n \geq p), \tag{9.22}$$

where the estimator $(\widehat{\rho}_1, \ldots, \widehat{\rho}_p)$ of $(\rho_1, \ldots, \rho_p)$ is computed by using (scalar) Yule-Walker equations (see Section 5.2). Quadratic prediction error satisfies

$$E\left(\widehat{E}_{n+1} - E(\xi_{n+1} \mid \xi_n, \ldots, \xi_1)\right)^2 = O(n^{-1}) \tag{9.23}$$

(cf. Yamamoto (1976)).

Concerning the more general SARIMA predictor the reader is referred to Box and Jenkins (1970) or Gourieroux and Monfort (1990).

Kernel predictor

Let p be a given positive integer ($p < n$). A p-dimensional **kernel** K is a bounded, symmetric, strictly positive density with respect to Lebesgue measure on $\mathbb{R}^d$ and such that

$$\| u \|^p K(u) \underset{\|u\|\to\infty}{\longrightarrow} 0 \text{ and } \int_{\mathbb{R}^p} \| u \|^2 K(u)du < \infty.$$

A typical example is $K(u) = (2\pi)^{-p/2} \exp\left(-\frac{\| u \|^2}{2}\right)$, where $\| \cdot \|$ is the Euclidean norm over $\mathbb{R}^p$.

The related **kernel predictor** is

$$\widehat{\xi}_{n+1}^{(K)} = \frac{\sum_{i=p+1}^{n} \xi_i K\left(\delta_n^{-1} Z_{i,n,p}\right)}{\sum_{i=p+1}^{n} K\left(\delta_n^{-1} Z_{i,n,p}\right)}, \tag{9.24}$$

where

$$Z_{i,n,p} = (\xi_n - \xi_{i-1}, \xi_{n-1} - \xi_{i-2}, \dots, \xi_{n-p+1} - \xi_{i-p}), \quad p+1 \leq i \leq n.$$

The **bandwidth** δ_n is a positive scalar at one's disposal.

This nonparametric predictor appears as a weighted mean with random coefficients

$$\widehat{\xi}_{n+1}^{(K)} = \sum_{i=p+1}^{n} p_{i,n}\xi_i, \tag{9.25}$$

where $p_{i,n} > 0$ and $\sum_{i=p+1}^{n} p_{i,n} = 1$.

An underlying model associated with this method should be a nonlinear $AR(p)$.

This predictor is accurate and more robust than $\widehat{\xi}_{n+1}^{(p)}$ because it is based on the "process history."

Asymptotic properties, references, and implementation appear in Bosq (1998). Also see Härdle (1990).

Functional kernel predictor

The above method can be extended to Hilbert spaces. If (ξ_n) is an H-valued process and K is a p-dimensional kernel, the predictor takes the form

$$\widehat{\xi}_{n+1}^{(K)} = \frac{\sum_{i=p+1}^{n} \xi_i K\left(\delta_n^{-1} Z'_{i,n,p}\right)}{\sum_{i=p+1}^{n} K\left(\delta_n^{-1} Z'_{i,n,p}\right)}, \tag{9.26}$$

with

$$Z'_{i,n,p} = (\| \xi_n - \xi_{i-1} \|, \dots, \| \xi_{n-p+1} - \xi_{i-p} \|), \quad p+1 \leq i \leq n.$$

Kernel predictor in continuous time

Now let $\xi = (\xi_t,\ t \in \mathbb{R})$ be a strictly stationary measurable real process. Data are $(\xi_t,\ 0 \leq t \leq T)$ and a forecast of ξ_{T+H} is needed, where $0 < H < T$.

For the sake of simplicity suppose that ξ is a measurable Markov process. The nonparametric predictor becomes

$$\widehat{\xi}_{T+H}^{(K)} = \frac{\displaystyle\int_H^T \xi_t K\left(\delta_T^{-1}(\xi_T - \xi_{t-H})\right) dt}{\displaystyle\int_H^T K\left(\delta_T^{-1}(\xi_T - \xi_{t-H})\right) dt}, \tag{9.27}$$

where K is a one-dimensional kernel and $\delta_T > 0$.

Under some rather stringent conditions the predictor has a surprising "parametric rate" viz

$$E\left(\widehat{\xi}_{T+H}^{(K)} - E(\xi_{T+H} \mid \xi_T)\right)^2 = O(T^{-1})$$

(cf. Bosq (1998)).

Local functional autoregressive predictor

Recall that the functional autoregressive predictor is $\rho_n(X_n)$, where ρ_n is defined by (8.53).

Its empirical form $\widehat{\rho}_n(\widehat{X}_n)$ is considered in the previous section. If $\widehat{X}_n$ is determined by spline smoothing it is called a **smooth FAP**.

A combination of FAP and kernel predictor has been proposed by Besse, Cardot, and Stephenson (1999).

They use **local estimators** of C and D defined as

$$\widehat{C}_n^{(K)} = \frac{\displaystyle\sum_{i=1}^{n} K\left(\delta_n^{-1} \| \widehat{X}_i - \widehat{X}_n \|\right) \widehat{X}_i \otimes \widehat{X}_i}{\displaystyle\sum_{i=1}^{n} K\left(\delta_n^{-1} \| \widehat{X}_i - \widehat{X}_n \|\right)} \tag{9.28}$$

and

$$\widehat{D}_n^{(K)} = \frac{\displaystyle\sum_{i=1}^{n-1} K\left(\delta_n^{-1} \| \widehat{X}_i - \widehat{X}_n \|\right) \widehat{X}_i \otimes \widehat{X}_{i+1}}{\displaystyle\sum_{i=1}^{n-1} K\left(\delta_n^{-1} \| \widehat{X}_i - \widehat{X}_n \|\right)}, \tag{9.29}$$

where K is a one-dimensional kernel, $\delta_n > 0$, and $\widehat{X}_1, \dots, \widehat{X}_n$ are determined via (9.9).

The associated estimator of ρ is given by

$$\widehat{\rho}_n^{(K)} = \widehat{\pi}_{(K)}^{k_n} \widehat{D}_n^{(K)} \left(\widehat{C}_n^{(K)}\right)^{-1} \widehat{\pi}_{(K_n)}^{k_n}, \tag{9.30}$$

with $\widehat{\pi}_{(K)}^{k_n}$ the orthogonal predictor of sp $\left(\widehat{v}_{1,n}^{(K)}, \dots, \widehat{v}_{k_n,n}^{(K)}\right)$, where $\left(\widehat{v}_{i,n}^{(K)}\right)$ are eigenvectors of $\widehat{C}_n^{(K)}$. The associated predictor is the local functional autoregressive predictor $\widehat{\rho}_n^{(K)}\left(\widehat{X}_n\right)$.

The values ℓ, k_n, and δ_n can be computed by validation.

This predictor is more robust than the simple FAP because of its resistance against slight variations of C and D.

9.4 Some numerical applications

We now present numerical applications performed in various fields by various authors.

Example 9.1: Simulation of $ARC(1)$
In a pioneering work, Pumo (1992) has studied empirical efficiency of the FAP by simulating an $ARC(1)$.
The white noise employed here is a special case of example 6.1, namely

$$\varepsilon_n(t) = W_{n+1} - W_n, \quad 0 \le t \le 1, \; n \in \mathbb{Z},$$

where W is a standard bilateral Wiener process (example 1.9).

Simulation of (ε_n) has been computed by using Karhunen-Loève expansion (see Theorem 1.5 and (1.28)).

$$W_t = \sum_{j=0}^{\infty} \eta_j \varphi_j(t), \quad 0 \le t \le 1,$$

where

$$\varphi_j(t) = \sqrt{2} \sin\left[(2j+1)\frac{\pi}{2}t\right], \quad 0 \le t \le 1, \; j \ge 0,$$

and (η_j) is a sequence of independent zero-mean Gaussian random variables with respective variances

$$\mu_j^2 = \frac{4}{(2j+1)^2\pi^2}, \quad j \ge 0.$$

Here the operator ρ is completely determined by

$$\rho(\varphi_j) = \beta_j \varphi_j, \quad j \ge 0,$$

where $0 < \sup_{j \geq 0} \beta_j < 1$, so that C_ε and ρ commute and

$$C(I - \rho^2) = C_\varepsilon.$$

This implies $v_j = \varphi_{j-1}$, $j \geq 1$, and

$$\lambda_j = \frac{\mu_{j-1}}{1 - \beta_{j-1}^2}, \quad j \geq 1.$$

Data are drawn on Figure 9.1. (71 time intervals, 64 points per interval).
Estimation of the first eigenvalues appear in table 9.1 below

j	β_j	λ_{j+1}	$\lambda_{j+1,n}$
0	0.45	0.5078	0.6106
1	0.90	0.2368	0.3440
2	0.34	0.0183	0.0208
3	0.45	0.0010	0.0076

Table 9.1
Estimated eigenvalues in Pumo example.

Comparison between X_{71} and its prediction from $(X_i,\ 1 \leq i \leq 70)$ is shown in Figures 9.1, 9.2, 9.3 and 9.4. The choice of k_n is clearly decisive.

Example 9.2: Simulation and smoothing of an $ARC_{q-1}(1)$
This simulation deals with Example 6.3, where ξ is defined by the stochastic differential equation

$$\sum_{\ell=0}^{q} a_\ell d\xi^{(\ell)}(t) = dW(t).$$

Recall that here $B = C_{q-1}([0, h])$.

Simulation has been performed by Besse and Cardot (1997).

The Wiener process is computed in a quicker way than in example 9.1 by using directly the fact that increments are stationary independent Gaussian random variables.

Simulation of (ξ_t) is obtained from the Milstein scheme (cf. Kloeden and Platen (1995)), which is easy to implement. An alternative method is the use of Bergström (1990) result which claims that a regular discretization of (ξ_t) gives an $ARMA(q, q-1)$ process. However, this method requires rather tedious computations.

Now smoothing is carried out with the SSRC described in Section 9.1. The obtained sample path appears in Figure 9.5.
An empirical study shows that the above procedure gives better forecasts than other smoothing methods.

Example 9.3: Prediction of harmonic levels in electrical networks
We now turn to forecasting electricity consumption in Bologna (Italy). The study is due to Cavallini, Montavari, Loggini, Lessi and Cacciani (1994).

Data are collected in the industrial area at Winter. The time-interval is one day, and from observation in continuous time during 29 days one construct, a forecast for the 30^{th} day. The value of k_n is determined automatically by (9.19), with $n_0 = 14$.

Figure 9.6 shows prediction of the fundamental current by sample mean (see Section 9.3) and by FAP. Due to importance of the deterministic component, both predictors have comparable performance.

Figure 9.7 deals with forecast of the 5^{th} harmonic current. Here the FAP appears to be more capable of taking into account stochastic modulation of recorded periods.

Example 9.4: Prediction of traffic
Besse and Cardot (1997) have examined prediction of traffic by FAP.

Data were collected hour by hour during one year at Saint-Arnoux tollbooth (France). They are represented in Figure 9.8. As in Example 9.2, smoothing is performed by SSRC.

In order to predict traffic on Wednesday, authors have used data corresponding to 30 Wednesdays and have compared prediction with data corresponding to the next 16 Wednesdays. The values ℓ and k_n are computed by validation.

Results and comparison with SARIMA prediction appear in Figure 9.9. FAP is globally better than the SARIMA method.

Example 9.5: Forecasting some climatic variations
In a recent work, Besse, Cardot, and Stephenson (1999) have used the local functional autoregressive prediction (cf. Section 9.3).

Data describe *el niño* : southern oscillation (ENSO). This is a natural phenomenon arising from interactions between atmosphere and ocean in the tropical Pacific Ocean. *el niño* (EN) is the ocean component of ENSO, involving major changes in tropical Pacific Ocean surface temperature, and the southern oscillation is its atmospheric counterpart. There is much interest in predicting ENSO because it may cause worldwide climate damage.

Data concern sea surface temperature index and sea level pressure. Forecasts of these two quantities appear in Figures 9.10 and 9.11, where various methods are compared.

Example 9.6: Prediction of electrocardiograms
This experiment has been performed by Bernard (1997). The question is adaptability of FAP to physiological data.

Bernard has considered an electrocardiogram consisting of 3000 complexes taken from a patient who suffers extrasystoles.

Figures below show that FAP can predict those extrasystoles. The best forecast is obtained if $k_n = 49$ (cf. Figures 9.12, 9.13, 9.14). The empirical quadratic error explodes if $k_n > 50$ (Figure 9.15).

This example confirms accuracy and flexibility of the functional autoregressive predictor.

NOTES

References are indicated within the text.

The independence test and the more general test "$\rho = \rho_0$" will appear in a work by Cardot, Mas and Sarda (2000).

Simulation and prediction of $ARH(1)$ processes

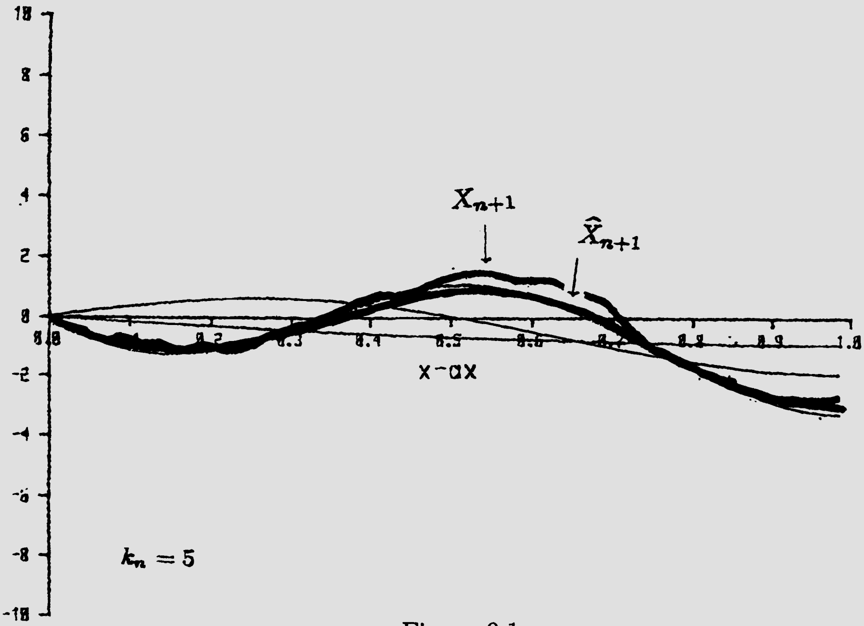


Figure 9.1

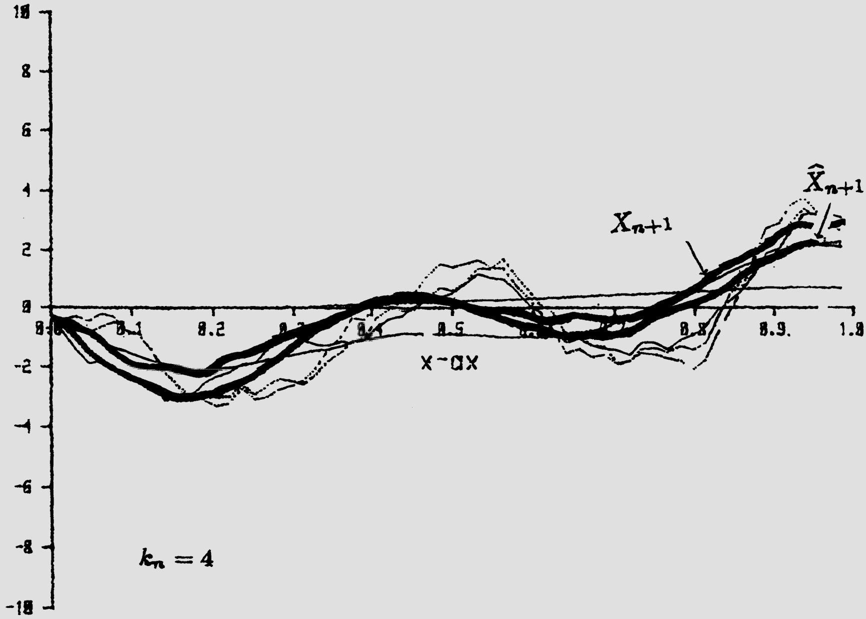


Figure 9.2

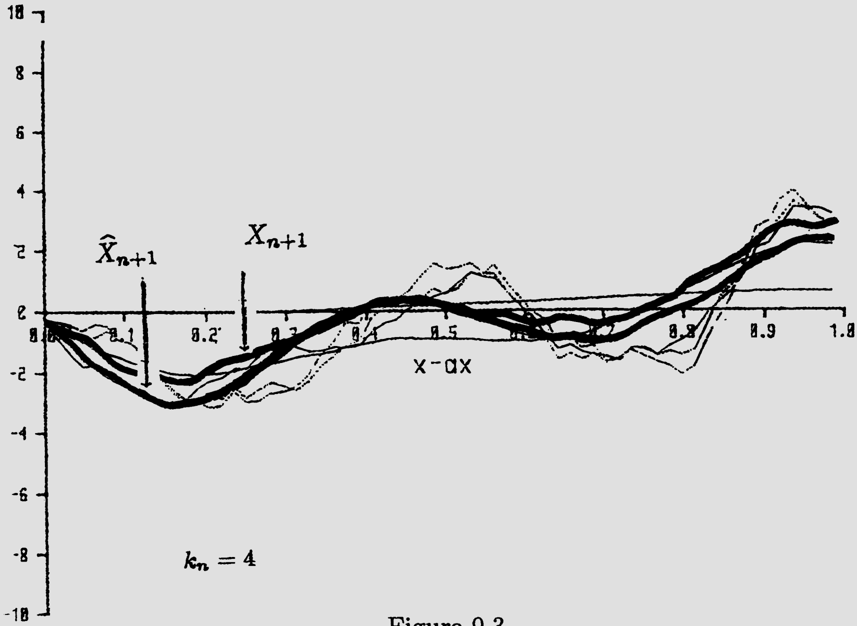


Figure 9.3

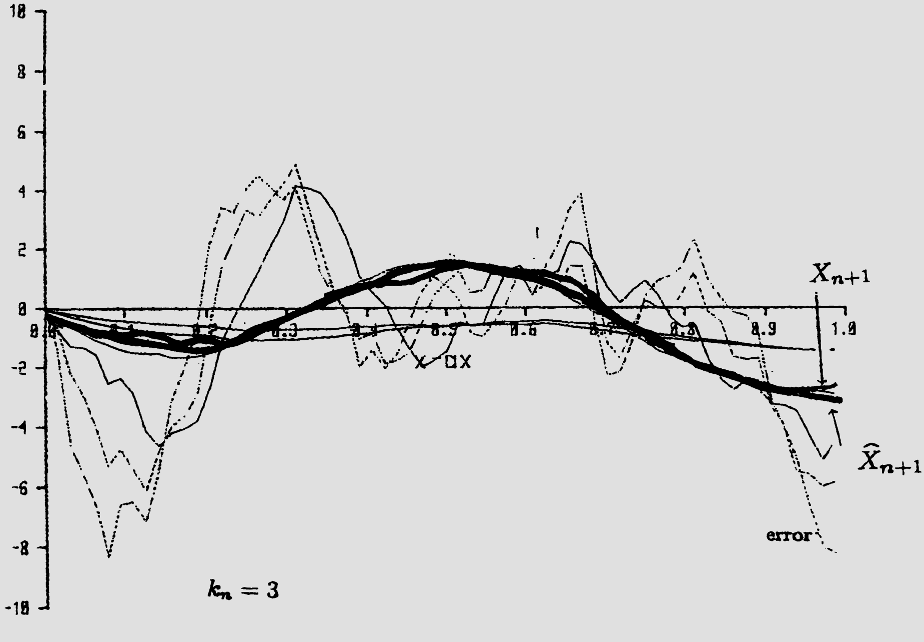
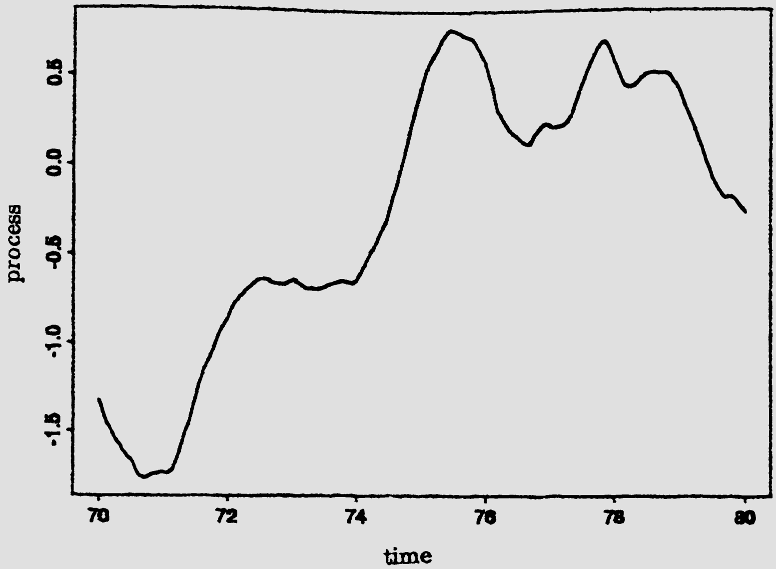
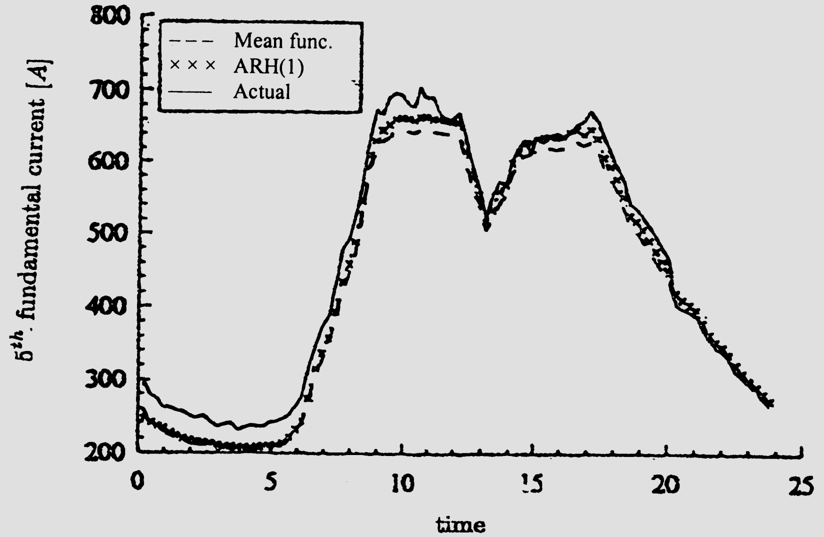


Figure 9.4



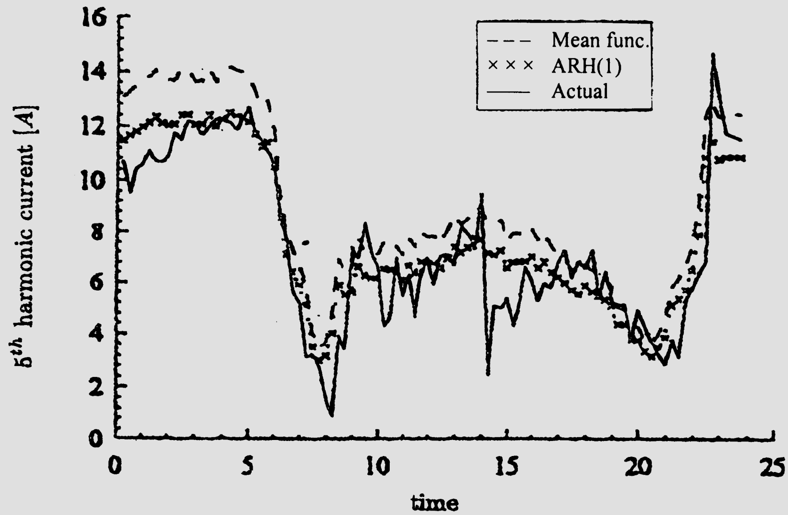
SSRC smoothing of a sample path of $ARB(1)$ associated with a linear stochastic differential equation of order q

Figure 9.5



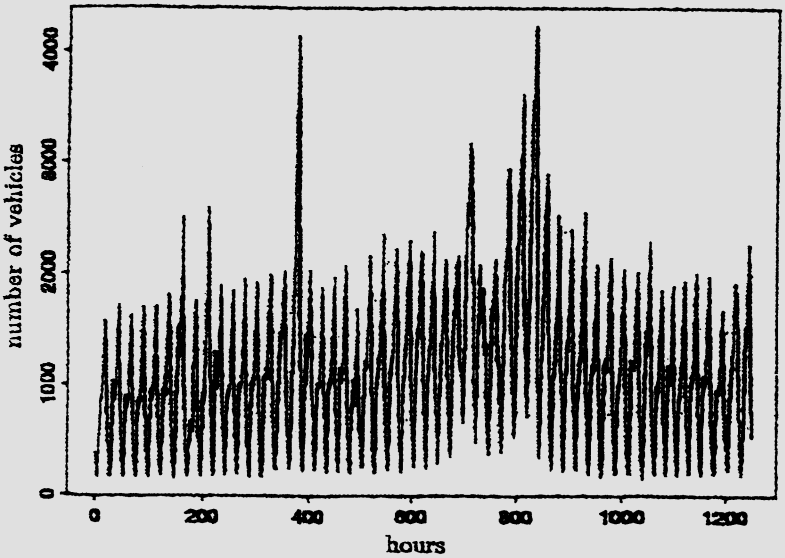
Actual and forecasting (obtained by either $ARH(1)$ or mean estimate) amplitude of 28^{th} period of the fundamental harmonic current measured at ILBUS

Figure 9.6



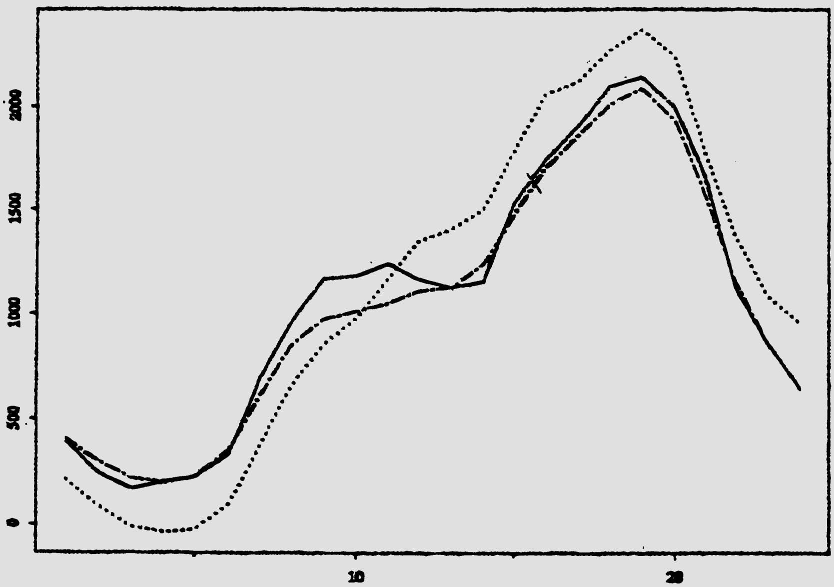
Actual and forecasted (obtained by either $ARH(1)$ or mean estimate) amplitude of 28^{th} period of the 5^{th} harmonic current measured at ILBUS

Figure 9.7



Traffic at Saint Arnoux tollbooth on Wednesday during 1990

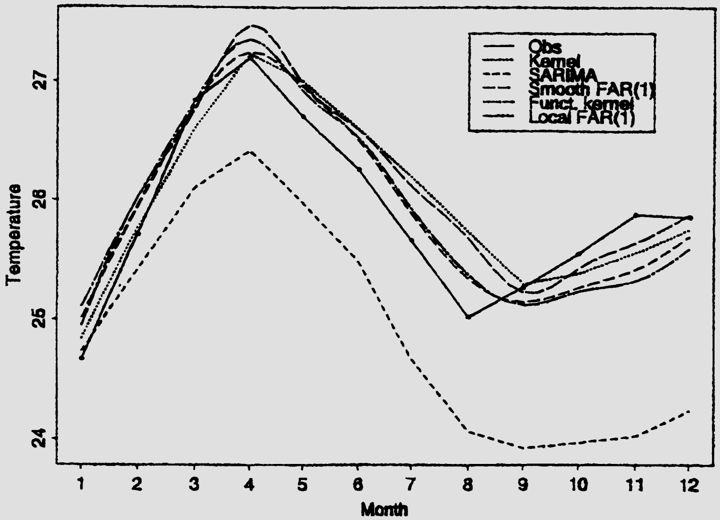
Figure 9.8



Prediction of traffic by functional autoregressive and methods
Solid: data. Dotted: SARIMA predictor
Dashed: functional predictor.

Figure 9.9

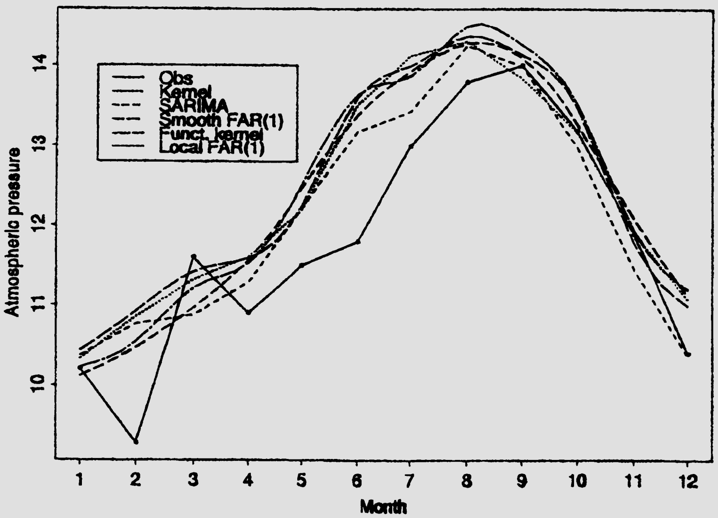
El niño



The raw ENSO time series of Pacific surface temperatures during 1986 and its forecasts

Figure 9.10

Tahiti 1986



Southern oscillation pressure in Tahiti during 1986 and its different forecasts.

Figure 9.11

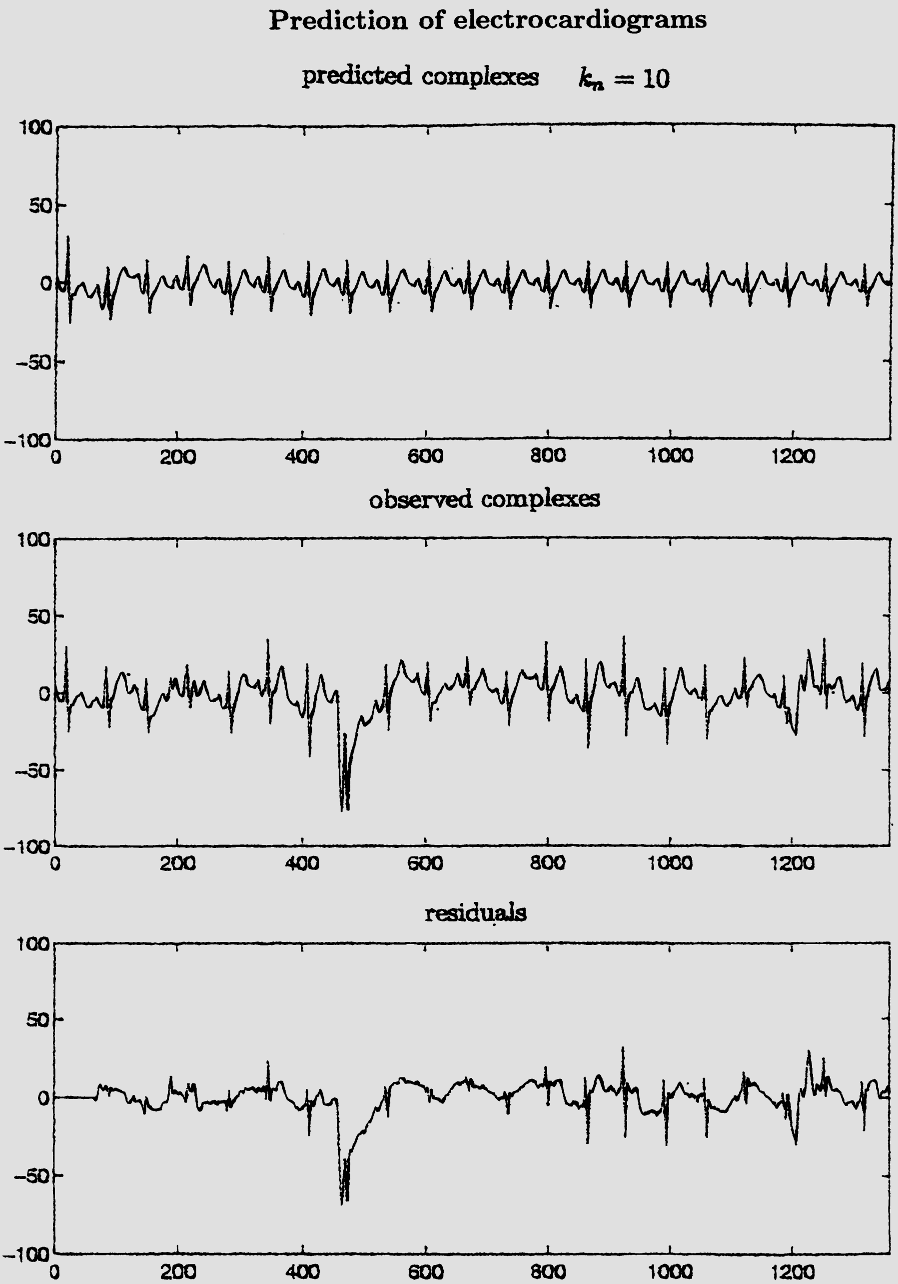

Prediction of electrocardiograms
predicted complexes $k_n = 10$
observed complexes
residuals
100
50
0
-50
-100
0
200
400
600
800
1000
1200

Figure 9.12

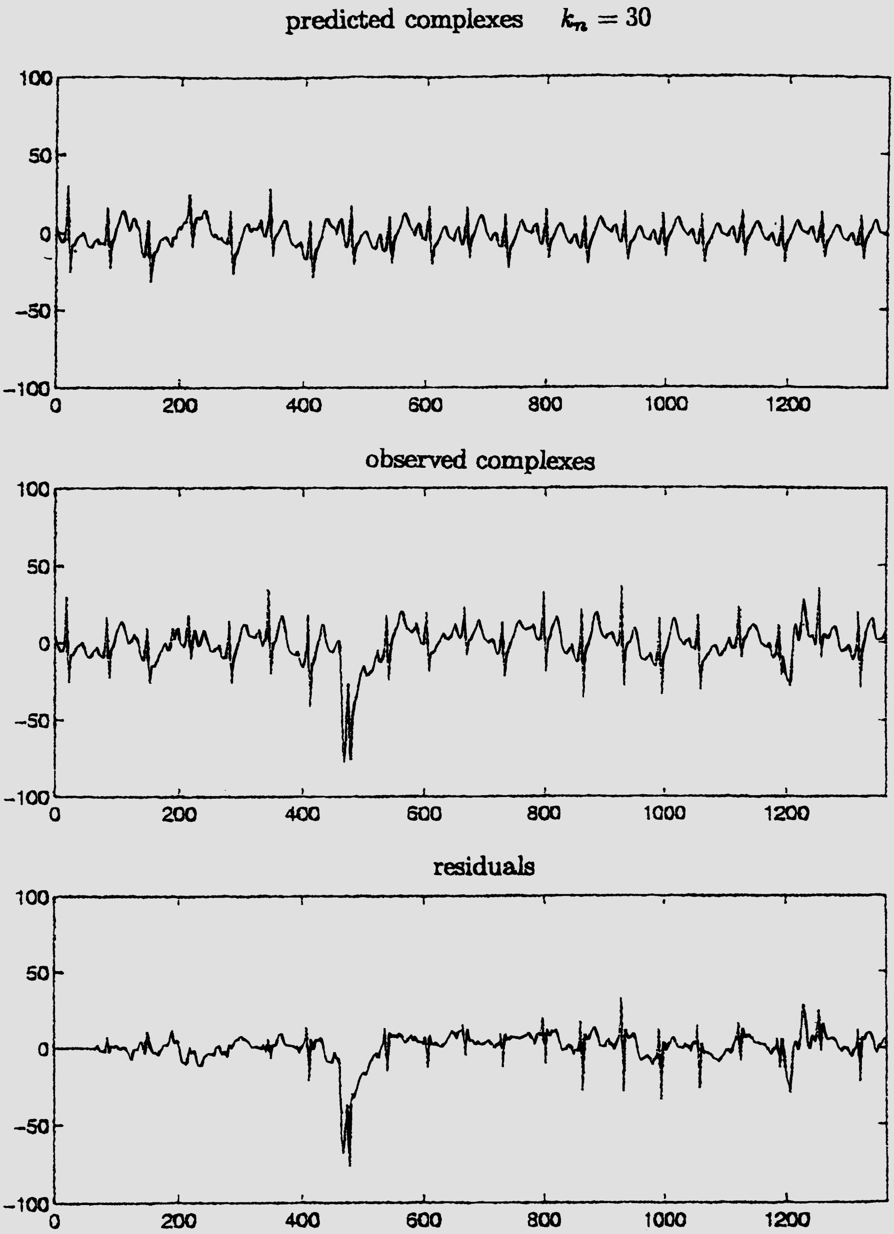

predicted complexes $k_n = 30$
100
50
0
-50
-100
0
200
400
600
800
1000
1200
observed complexes
100
50
0
-50
-100
0
200
400
600
800
1000
1200
residuals
100
50
0
-50
-100
0
200
400
600
800
1000
1200

Figure 9.13

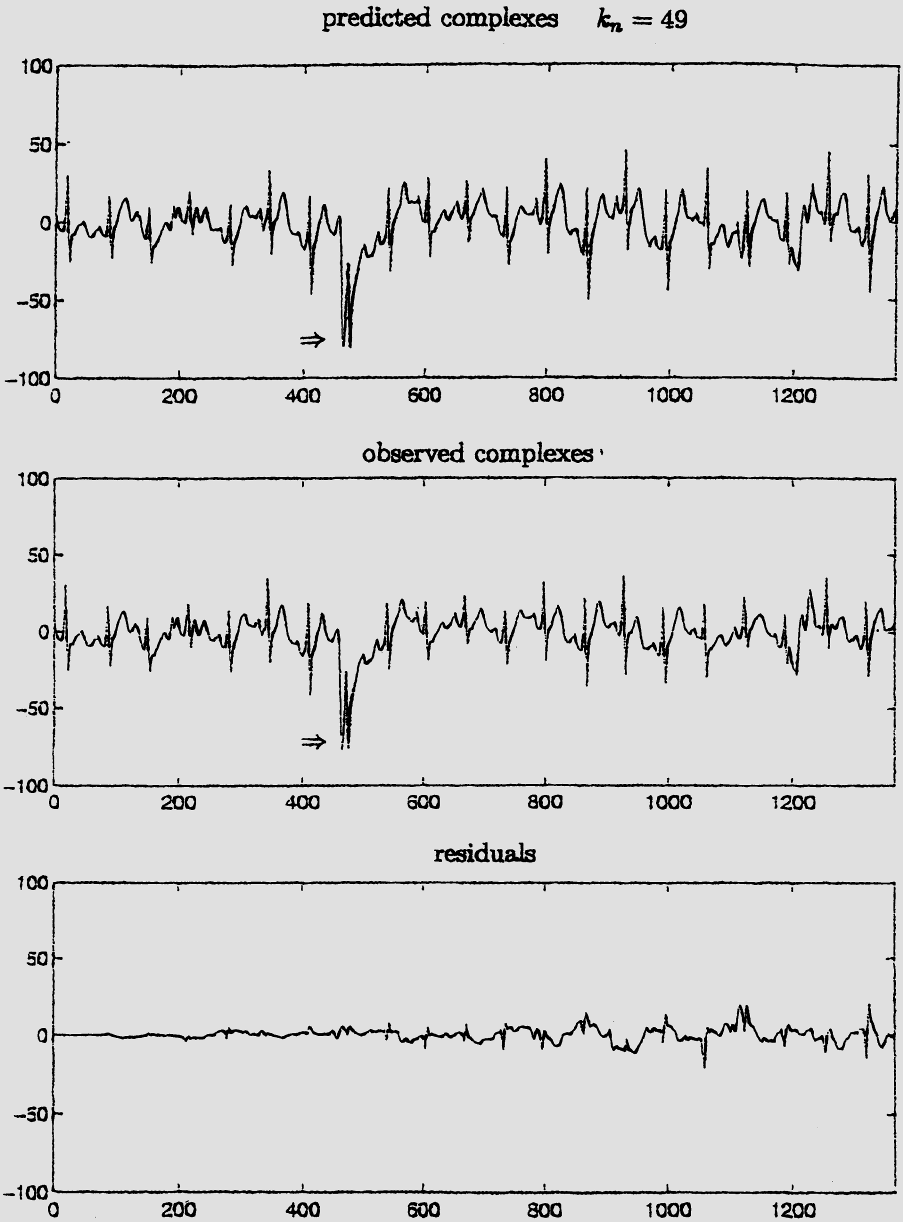


Figure 9.14

Quadratic empirical error in function of k_n

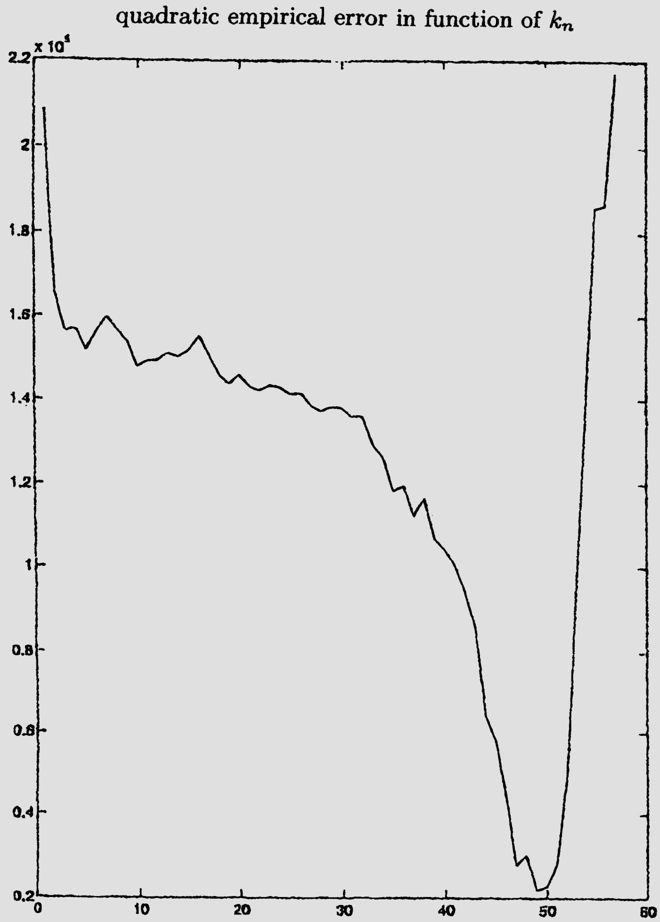


Figure 9.15

Appendix

1 Measure and probability

A **σ-algebra** (or σ-field) over a non-empty set Ω is a family $\mathcal{A}$ of subsets of Ω such that $\Omega \in \mathcal{A}$, $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$ and if $(A_n,\ n \geq 1)$ is a countable family of elements in $\mathcal{A}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. A **measure** over $(\Omega, \mathcal{A})$ is a mapping $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}_+ := [0, +\infty]$ such that if $A_n,\ n \geq 1$ is a disjoint countable family of elements of $\mathcal{A}$ then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A **Probability** P is a measure such that $P(\Omega) = 1$. If P is a Probability over $(\Omega, \mathcal{A})$, $(\Omega, \mathcal{A}, P)$ is called a **Probability space**. A Probability space is **complete (with respect to P)** if for all $A \in \mathcal{A}$ with $P(A) = 0$, every subset of A is also in $\mathcal{A}$ i.e. each **P-negligible** set is in $\mathcal{A}$.

Let Ω be a metric space, the **Borel σ-algebra** $\mathcal{B}_\Omega$ of Ω is the smallest σ-algebra which contains all open sets. $\mathcal{B}_\mathbb{R}$ is also the smallest σ-algebra which contains all open intervals.

The **Lebesgue measure** λ over $(\mathbb{R},\ \mathcal{B}_\mathbb{R})$ is characterized by $\lambda(]a, b]) = b - a$; $a, b \in \mathbb{R}$, $a < b$. It is invariant by translation.

A measure μ is **σ-finite** if there exists $(A_n) \subset \mathcal{A}$ such that $A_n \subset A_{n+1}$, $n \geq$, $\bigcup_{n=1}^{\infty} A_n = \Omega$ and $\mu(A_n) < \infty$, $n \geq 1$.

Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ $i = 1, 2$ two measure spaces with μ_1 and μ_2 σ-finite. Their **product** is defined as $(\Omega_1 \times \Omega_2,\ \mathcal{A}_1 \otimes \mathcal{A}_2,\ \mu_1 \otimes \mu_2)$ where $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the smallest σ-algebra which contains all $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$ and $\mu_1 \otimes \mu_2$ is the unique measure over $(\Omega_1 \times \Omega_2,\ \mathcal{A}_1 \otimes \mathcal{A}_2)$ such that

$$(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)\ ;\ A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2.$$

2 Random variables

A real **random variable** defined over the Probability space $(\Omega, \mathcal{A}, P)$ is a mapping $X : \Omega \to \mathbb{R}$ such that $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}_{\mathbb{R}}$ i.e. a $\mathcal{A} - \mathcal{B}_{\mathbb{R}}$ **measurable** mapping. The **distribution** of X is a Probability on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ defined by

$$P_X(B) = P(X^{-1}(B)),\ \ B \in \mathcal{B}_{\mathbb{R}}.$$

A simple real random variable has the form

$$X = \sum_{j=1}^{k} x_j 1_{A_j}$$

where $x_1, \ldots, x_k \in \mathbb{R}$ and $A_1, \ldots, A_k \in \mathcal{A}$ and are disjoint. Let X be a simple real random variable. Its integral or expectation is defined as

$$EX = \sum_{j=1}^{k} x_j P(A_j).$$

Now if X is any positive random variable one sets

$$EX = \sup\{EY,\ \ 0 \leq Y \leq X,\ Y \in \Sigma\}$$

where Σ denote the family of simple random variables defined over $(\Omega, \mathcal{A}, P)$.

Let X be a real random variable, if $E|X| < \infty$ then X is said to be **integrable** and

$$EX = EX^+ - EX^-$$

where $X^+ = \sup(0, X)$ and $X^- = \sup(0, -X)$.

If $EX^2 < \infty$ one defines the **variance** of X as $VX = E(X - EX)^2$.

Similarly one defines $\mathbb{R}^d$-valued random variables : $X = (X_1, \ldots, X_d)'$.

Their expectation is defined coordinatewise. Variance is replaced by the **covariance matrix** : if $EX_i^2 < \infty$; $i = 1, \ldots, d$ the covariance matrix of X is defined as the $d \times d$-matrix

$$C_X = [E((X_i - EX_i)(X_j - EX_j))] := [\mathrm{Cov}(X_i, X_j)].$$

The **correlation matrix** of X is $\rho_X = (\rho_{ij})_{1\le i,j\le d}$ where ρ_{ij} is the **correlation coefficient** of X_i and X_j defined as

$$\rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{VX_i}\sqrt{VX_j}} := \mathrm{Corr}(X_i, X_j)$$

provided X_i and X_j are not degenerated.

Finally if X and Y are two $\mathbb{R}^d$-valued random variables such that $EX_i^2 < \infty$ and $EY_i^2 < \infty$; $i = 1, \dots, d$ their **cross-covariance and cross-correlation matrices** are the $d \times d$-matrices

$$C_{X,Y} = [\mathrm{Cov}(X_i, Y_j)] \quad , \quad C_{Y,X} = [\mathrm{Cov}(Y_i, X_j)],$$

$$\rho_{X,Y} = [\mathrm{Corr}(X_i, Y_j)] \quad , \quad \rho_{Y,X} = [\mathrm{Corr}(Y_i, X_j)].$$

3 Function spaces

A (real) **Banach space** B is a linear space over the field of real numbers equipped with a norm $\| \cdot \|$ for which it is complete. The (topological) **dual space** of all continuous linear functional is denoted by B^*. The norm on B^*, also denoted by $\| \cdot \|$ is defined by

$$\| x^* \| = \sup_{\substack{x \in B \\ \|x\| \le 1}} |x^*(x)| \ , \quad x^* \in B^*.$$

A Banach space is **separable** if it has a countable dense subset.

A **Hilbert space** H is a Banach space whose norm is associated with an inner product (or scalar product) $< \cdot, \cdot >$, thus $\| x \| = (< x, x >)^{1/2}$, $x \in H$.

The following **Cauchy-Schwarz inequality** is then valid :

$$| < x, y > | \le \| x \| \, \| y \| \ ; \quad x, y \in H$$

H^* may be identified to H in the sense that to each x^* in H^* one may associate a unique y in H such that

$$x^*(x) = < y, x > \ , \quad x \in H.$$

Let H' be a closed subspace of H, its **orthogonal projector** is the mapping $\pi : H \mapsto H'$ characterized by

$$< x - \pi(x), \ z > = 0 \ , \quad z \in H',$$

or by

$$\| x - \pi(x) \| = \inf\{ \, \| x - z \|, \ z \in H' \}.$$

A set $(e_i,\ i \in I)$ of elements in H is **orthonormal** if

$$< e_i, e_j >= 1 \ \text{ if } i = j, \ = 0 \text{ if } i \neq j.$$

Let H be a separable Hilbert space then there exist countable orthonormal sets such that

$$\sum_{i=1}^{\infty} < x, e_i >^2 = \| x \|^2 \ , \quad x \in H.$$

Such a set is an **orthonormal basis** of H or a **complete orthonormal set**.

4 Basic function spaces

For $1 \leq p \leq \infty$ let ℓ^p be the space of all real **sequences** $x = (x_i)$ for which $\| x \|_p = \left(\sum_i |x_i|^p \right)^{1/p} < \infty$ ($\sup_i |x_i| < \infty$ if $p = \infty$) then ℓ^p is a separable Banach space. ℓ^2 is a Hilbert space with scalar product

$$< x, y >= \sum_i x_i y_i \ ; \ (x_i),\ (y_i) \in \ell^2.$$

Let μ be a measure over $(\Omega, \mathcal{A})$. Identifying f and g if $\mu(f \neq g) = 0$ one may define the spaces $\boldsymbol{L^p(\Omega, \mathcal{A}, \mu)}$ where $1 \leq p \leq \infty$ of functions $f : \Omega \mapsto \mathbb{R}$ such that $\| f \|_p = \left(\int |f|^p d\mu \right)^{1/p} < \infty$

$$(\| f \|_\infty = \inf\{c : \mu\{|f| > c\} \text{ if } p = \infty).$$

For each p, $L^p(\Omega, \mathcal{A}, \mu)$ is a Banach space. If $1 < p < +\infty$ then $(L^p(\Omega, \mathcal{A}, \mu))^*$ may be identified to $L^q(\Omega, \mathcal{A}, \mu)$ where $\frac{1}{p} + \frac{1}{q} = 1$. $L^2(\Omega, \mathcal{A}, \mu)$ is a Hilbert space with scalar product

$$< f, g >= \int f g d\mu \ ; \ f, g \in L^2(\Omega, \mathcal{A}, \mu).$$

If μ is Lebesgue measure over a Borel set of $\mathbb{R}$ these spaces are separable.

Let $k \geq 0$ be an integer and $[a, b]$ a compact interval in $\mathbb{R}$. The family $\boldsymbol{C_k[a, b]}$ of functions $x : [a, b] \mapsto \mathbb{R}$ with k continuous derivatives is a separable Banach space under the norm

$$\| x \| = \sum_{\ell=0}^{k} \sup_{0 \leq t \leq 1} |x^{(\ell)}(t)|.$$

5 Conditional expectation

Let $(\Omega, \mathcal{A}, P)$ be a Probability space and let $\mathcal{A}_0$ be a sub σ-algebra of $\mathcal{A}$. **Conditional expectation** with respect to $\mathcal{A}_0$ is the mapping $E^{\mathcal{A}_0}$: $L^1(\Omega, \mathcal{A}, P) \mapsto L^1(\Omega, \mathcal{A}_0, P)$ characterized by

$$\int_B E^{\mathcal{A}_0} X dP = \int_B X dP \ , \quad B \in \mathcal{A}_0,$$

$X \in L^1(\Omega, \mathcal{A}, P)$.

Considered as a mapping from $L^2(\Omega, \mathcal{A}, P)$ to $L^2(\Omega, \mathcal{A}_0, P)$ it becomes the **orthogonal projector of $\boldsymbol{L^2}$ $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}_o}, \boldsymbol{P})$** i.e. for each X in $L^2(\Omega, \mathcal{A}, P)$ $E^{\mathcal{A}_0} X$ is the best approximation of X by a square-integrable $\mathcal{A}_0$-measurable random variable. In particular if $\mathcal{A}_0 = \sigma(X_i, \ i \in I)$ i.e. $\mathcal{A}_0$ is the smallest σ-algebra such that each X_i is $\mathcal{A}_0 - \mathcal{B}_{\mathbb{R}}$ measurable then $E^{\mathcal{A}_0}(X)$ is the **conditional expectation of $\boldsymbol{X}$ given $(\boldsymbol{X_i}, \ \boldsymbol{i \in I})$**.

If $X = \mathbf{I}_A$ where $A \in \mathcal{A}$ then one sets

$$P^{\mathcal{A}_0}(A) = E^{\mathcal{A}_0}(1_A).$$

$P^{\mathcal{A}_0}$ is called conditional Probability with respect to $\mathcal{A}_0$ (or relative to $\mathcal{A}_0$).

6 Stochastic integral

Let $W = (W_t, \ t \geq 0)$ be a standard Wiener process defined on $(\Omega, \mathcal{A}, P)$ (see example 1.6). Set

$$\mathcal{F}_t = \sigma(W_s, \ 0 \leq s \leq t) \quad , \quad t \geq 0$$

and define the class $\mathcal{I}$ of stochastic processes (Section 1.1) $\xi = (\xi_t, 0 \leq t \leq a)$ such that

(1) $(t, \omega) \mapsto \xi_t(\omega)$ is in $L^2([0, a] \times \Omega, \ \mathcal{B}_{[0,a]} \otimes \mathcal{A}, \ \lambda \otimes P)$,

(2) ξ is **adapted** to $(\mathcal{F}_t)$ (or **nonanticipative** with respect to $(\mathcal{F}_t)$) i.e. $\omega \mapsto \xi_t(\omega)$ is $\mathcal{F}_t - \mathcal{B}_{\mathbb{R}}$ measurable for all t in $[0, a]$

Now let $\mathcal{E}$ be the subclass of $\mathcal{I}$ constituted by the **step processes** : $\xi \in \mathcal{E}$ if $\xi \in \mathcal{I}$ and

$$\xi_t(\omega) = \sum_{i=0}^{k-1} f_i(\omega) \mathbf{I}_{[t_i, t_{i+1})}(t), \ \omega \in \Omega, \ 0 \leq t \leq a,$$

where $0 = t_0 < t_1 < \ldots < t_k = a$, $[t_i, t_{i+1})$ is open on the right except $[t_{k-1}, t_k)$ which is closed, f_i is a member of $L^2 \left(\Omega, \mathcal{F}_{t_i}, P\right)$, $0 \leq i \leq k-1$.

For such a process one defines the **stochastic integral (in Ito sense)** as

$$I(\xi) := \int_0^a \xi_t dW_t := \sum_{i=0}^{k-1} f_i \left(W_{t_{i+1}} - W_{t_i} \right).$$

$I(\xi)$ is a square-integrable random variable. Basic properties of the Ito integral are linearity and

$$E\, I(\xi) = 0, \quad E[I(\xi)\, I(\eta)] = \int_0^a E(\xi_t \eta_t) dt\ ;\ \xi,\ \eta \in \mathcal{E}.$$

This shows that $\xi \mapsto I(\xi)$ defines a linear **isomorphism** from $\mathcal{E}$ to $L^2(\Omega, \mathcal{A}, P)$.

Now $\mathcal{E}$ is dense in $\mathcal{I}$ thus, by continuity, I has a unique linear extension to $\mathcal{I}$ which is called **stochastic integral** on $[0, a]$ with respect to W and denoted by

$$I(\xi) = \int_0^a \xi_t dW_t \ , \quad \xi \in \mathcal{I}.$$

The basic properties remain valid. In particular $I(\xi)$ is a zero-mean square integrable random variable and

$$E[I(\xi)^2] = \int_0^a \left(E\xi_t^2 \right) dt.$$

The above definition extends to intervals of infinite length.

Now it is possible to give a specific sense to the **stochastic differential equations** (1.7), (1.14) and (6.10). For example (1.7) means

$$\xi_t - \xi_0 = \int_0^t m(\xi_s, s) ds + \int_0^t \sigma(\xi_s, s) dW_s, \quad t \geq 0.$$

References

Akhiezer, N.I., and I.M. Glazman (1961). *Theory of linear operators in Hilbert spaces.* Ungar, New York

Allahverdiev, D.E. (1957). On the rate of approximation of completely continuous operators by finite dimensional operators.*Azerbaidjan Gos. Univ. Učen. Zap. Ser. Fiz. Mat : Him. Nauk* 2, 27-35 (in Russian).

Antoniadis, A., and J.H. Beder (1989). Joint estimation of the mean and the covariance of a Banach valued Gaussian vector. *Statistics* 20, 1, 77-93.

Ash, R., and M. Gardner (1975). *Topics in Stochastic processes.* Academic Press, New York

Beck, A., and Warren P. (1974). Strong laws of large numbers for weakly orthogonal sequences of Banach space-valued random variables. *Annals of Probab.*, 2, 5, 918-925.

Bensaïd, N. (1992). Méthode Hilbertienne pour la prévision non paramétrique d'un processus ponctuel.*C.R. Acad. Sci. Paris* I, 314, 11, 865-870.

Bergström, A.R. (1990). *Continuous time econometric modelling. Oxford.*

Bernard, P. (1997). *Analyse de signaux physiologiques.* Mémoire Univ. Cathol. Angers.

Besse, P., and H. Cardot (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. of Statist.* 24, 467-487.

Besse, P., H. Cardot, and D. Stephenson (1999). Autoregressive forecasting of some climatic variations. To appear in *Scandin. J. of Statistics.*

Billingsley, P. (1968). *Convergence of probability measures.* Wiley. New York

Bosq, D. (1983). Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires. *Z. Wahrscheinlichkeit. und Verw. Geb.* 64, 541-553.

Bosq, D. (1989). Propriétés des opérateurs de covariance empiriques d'un processus stationnaire Hilbertien. *C.R. Acad. Sci. Paris*, 309, I, 873-875.

Bosq, D. (1990). Sur les processus autorégressifs dans les espaces de Hilbert. *Publ. Instit. Stat. Paris*, XXXV, 2, p. 3-17.

Bosq, D. (1991-a). Modelization, nonparametric estimation and prediction for continuous time processes. *Nato Asi Series C*, 335, 509-529, ed. Roussas.

Bosq, D. (1991-b). Prévision d'une fonction aléatoire. *Statistica LI*, 2, 153-164.

Bosq, D. (1996). Limit theorems for Banach-valued autoregressive processes. Application to real continuous time processes. *Bull. Belg. Math. Soc.* 9, 537-555.

Bosq, D. (1998). Nonparametric statistics for stochastic processes. Estimation and prediction (2^{nd} edition) *Lecture Notes in Statistics*, 110. Springer-Verlag, New York

Bosq, D. (1999-a). Représentation autorégressive de l'opérateur de covariance empirique d'un $ARH(1)$. Applications. *C.R. Acad. Sci. Paris*, t. 329, Ser. I, 531-534.

Bosq, D. (1999-b). Autoregressive Hilbertian processes. *Technical report LSTA*. Univ. Paris 6. 1, 31 pages.

Bosq, D., and M. Delecroix (1985). Nonparametric prediction of a Hilbert space valued random variable. *Stoch. Proc. and Appl.* 19, 271-280.

Bosq, D., and T. Mourid (1999). On the equivalence of the measures induced by Banach valued Gaussian autoregressive processes. *Stoch. Anal. and Appl.* 17, 2, 137-144.

Box, G.E., and G.M. Jenkins (1970). *Time Series Analysis : Forecasting and Control.* Holden-day, San Francisco.

Bradley, R.C. (1997). On quantiles and the central limit theorem question for strongly mixing sequences. *J. Theor. Probab.*, 10, 507-555.

Brockwell, P., and R. Davis (1991). *Time series : Theory and methods* (2^{nd} ed). Springer-Verlag, New York

Cardot, H. (1997). Contribution à l'estimation et à la prévision statistique de données fonctionnelles. *Ph. D. thesis University of Toulouse 3.*

Cardot, H., F. Ferraty, and P. Sarda (1999). Functional linear model. *Probability Letters*, 45, 11-22.

Cardot, H., A. Mas, and P. Sarda (2000). Testing the autocorrelation operator of an $ARH(1)$. *To appear.*

Cavallini, A., G.C. Montanari, M. Loggini, 0. Lessi, and M. Cacciari (1994). Nonparametric prediction of harmonic levels in electrical networks. *Proceed. IEEE ICHPS VI*, Bologna, 165-171.

Cramer, H., and M.R. Leadbetter (1967). *Stationary and related stochastic processes.* Wiley, New York.

Dauxois, J., A. Pousse, and Y. Romain (1982). Asymptotic theory for the PCA of a vector random function. Some applications to statistical inference. *J. of Mult. Anal.* 12, 136-154.

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Washrscheinlichkeistheor und Verw. Geb.* 63, 393-432.

Denisevskii, N.A. (1986). The strong law of large numbers for a linear process in Banach space (in Russian) *Dokl. Akad. Nauls. Ukrain. SSR*, A, 12, 57-61.

Denisevskii, N.A. (1990). The central limit theorem for a linear process in Banach space. *Theor. Probab. and Math. Statist.* 40, 1990.

Denisevskii, N.A., and Y.A. Dorogovtsev (1988). On the law of large numbers for a linear process in Banach space. *Soviet Math. Dokl.* 36, 1, 47-50.

Dippon, J. (1993). Asymptotic confidence spheres in Banach spaces via covariance operators. *J. of Multiv. Anal.* 47, 48-58.

Doob, J. (1953). *Stochastic processes.* Wiley, New York.

Doukhan, P. (1994). Mixing. Properties and examples. *Lecture Notes in Statistics*, 85, Springer-Verlag, New York.

Doukhan, P., J. Léon, and F. Portal (1985). Calcul de la vitesse de convergence dans le théorème central limite vis à vis des distances de Prokorov, Dudley et Lévy dans le cas de variables aléatoires dépendantes. *Probab. Math. Statist.* 6, 1, 19-27.

Dunford, N., and J.T. Schwartz (1958). *Linear operators I*, Wiley, New York.

Falb, P.L. (1967). Infinite-dimensional filtering : The Kalman-Bucy filter in Hilbert-space. *Inform. Control*, 11, 102-137.

Fortet, R. (1995). *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert.* Hermes. Paris.

Geffroy, J. (1959). Quelques extensions du théorème de P. Lévy sur la convergence p.s. des séries aléatoires à termes indépendants. *C.R. Acad. Sci. Paris* 24, 1180-1182.

Gohberg, I.C., and M.G. Krein (1971). *Opérateurs linéaires non autoadjoints dans un espace Hilbertien.* Dunod, Paris

Gourieroux, C., and A. Monfort (1990). *Séries temporelles et modèles dynamiques.* Economica. Paris.

Grenander, U. (1981). *Abstract inference.* Wiley, New York.

Guelfand, I.M., and N.Y. Vilenkin (1967). *Les distributions* t. 4. Dunod, Paris.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press.

Hannan, E.J. (1970). *Multiple time series.* Wiley, New York.

Heinkel, B. (1979). Relation entre théorème central limite et loi du logarithme itéré dans les espaces de Banach. *Z. Wahrscheinlichkeistheor. und Verw. Geb.* 41, 211-220.

Hoffmann-Jorgensen, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* 52, 159-186.

Ito, K., and M. Nisio (1968). On the convergence of sums of independent Banach space valued random variables. *Osaka J. Math.* 5, 35-48.

Jain, N.C., and M.B. Marcus (1975). Central limit theorem for $C(S)$-valued random variables. *J. Funct. Anal.* 19, 216-231.

Jakubowski, A. (1988). Tightness criteria for random measures with application to the principle of conditioning in Hilbert spaces. *Probab. Math. Statist.*, 9, 1, 95-114.

Karatzas, I., and S. Shreve (1991). *Brownian motion and stochastic calculus* (2^{nd} ed). Springer-Verlag, New York.

Karhunen, K. (1946). Zur spektraltheorie stochasticher Prozesse. *Ann. Ac. Sci. Fennicae*, AI, 34.

Kloeden, P., and E. Platen (1995). *Numerical solution of stochastic differential equations* (2^{nd} ed). Springer-Verlag, Berlin.

Koval, V. (1993). A functional limit theorem for solutions of stochastic difference equation in a Banach space. *Proceed. 6^{th} int. Vilnius conf. on Probab. and Stat.*, 203-204.

Kubrusly, C.S. (1985). Mean square stability for discrete bounded linear systems in Hilbert space. *Siam J. Control and Optimization.* 23, 1, 19-29.

Kuelbs, J. (1975). The law of the iterated logarithm and related strong convergence theorems for Banach space valued random variables. Ecole d'été de Probabilités de St-Flour. *Lecture Notes in Mathematics*, 539, 225-314. Springer-Verlag, Berlin.

Kukush, A.G. (1990). On the probability of large deviations of an estimator of a nonlinear regression parameter in Hilbert space. *Th. Prob. Math. Stat.*

40, 51-58.

Kulikova, T. (1996). Estimation of the correlation function of a stationary random process by linear splines. *Russian Phys. J.* 39, 4, 308-314.

Lecoutre, J.P. (1990). Uniform consistency of a class of regression function estimators for Banach space valued random variables. *Stat. Prob. Letters* 10, 145-149.

Ledoux, M., and M. Talagrand (1991). *Probability in Banach spaces.* Springer-Verlag, Berlin.

Lodges, W. (1986). Estimation of parameters for Hilbert space valued partially observable stochastic processes. *J. of Mult. Anal.* 20, 1, 161-174.

Loève, M. (1948). Fonctions aléatoires du second ordre. Suppl. to P. Lévy. *Processus stochastiques et mouvement Brownien.* p. 462-470. Gauthier-Villars. Paris.

Mas, A. (1999-a) Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus $ARH(1)$. *C. R. Acad. Sci. Paris*, t. 329, Ser. I, 899-902.

Mas, A. (1999-b) Personal communication.

Maurey, B. (1973). Espaces de cotype p, $0 < p \leq 2$. *Séminaire Maurey-Schwartz 1972-73*, Ecole Polytechnique, Paris.

Mercer, T. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Transac. London Phil. Soc. A*, 209, 415-446.

Merlevède, F. (1995). Sur l'inversibilité des processus linéaires à valeurs dans un espace de Hilbert. *C.R. Acad. Sci. Paris*, 321, I, 477-480.

Merlevède, F. (1996, a). Processus linéaires Hilbertiens : inversibilité, théorèmes limites. Estimation et prévision. *Ph.d. thesis. University of Paris 6.*

Merlevède, F. (1996, b). Central limit theorem for linear process with values in a Hilbert space. *Stoch. Proc. and Appl.* 65, 103-114.

Merlevède, F. (1997). Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires Hilbertiens. *C.R. Acad. Sci. Paris*, I., 324, 573-576.

Merlevède, F., M. Péligrad, and S. Utev (1997). Sharp conditions for the CLT of linear processes in a Hilbert space. *Journal of Theor. Probab.*, 10, 3, 681-693.

Merlevède, F., and M. Péligrad (1999). Personal communication.

Mourid, T. (1993). Processus autorégressifs d'ordre supérieur. *C.R. Acad. Sci. Paris*, t. 317, I, 1167-1172.

Mourid, T. (1995). Contribution à la statistique des processus autorégressifs à temps continu. *Thesis. University of Paris 6.*

Mourier, E. (1953). Elements aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré*, 13, 161-244.

Nsiri, S., and R. Roy (1993). On the invertibility of multivariate linear processes. *J. of Time Series Anal.* 14, 3, 305-316.

Parthasarathy, K.R. (1967). *Probability measures on metric spaces.* Academic Press, New York.

Paulauskas, V., and A. Rackauskas (1989). *Approximation in the central limit theorem. Exact results in Banach space.* Reidel, Dordrecht.

Pinelis, I.F., and I. Sakhanenko (1985). Remarks on inequalities for large deviation probabilities. *Theory of Probab. Appl.* 30, 143-148.

Prakasa Rao, B.L.S. (1988). Statistical inference from sampled data for stochastic processes. *Contemporary Math.*, 80, 249-284.

Prokhorov, Y.V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probab. Appl.* 1, 157-214.

Pugachev, V.S.(1959). A method of finding eigenvalues or eigenfunctions for a class of linear integral equations. *Prikladmet i. meth. XXIII*, 3, 527-533.

Pumo, B. (1992). Estimation et prévision de processus autorégressifs fonctionnels. Application aux processus à temps continu. *Ph. D. Thesis University of Paris 6.*

Pumo, B. (1999). Prediction of continuous time processes by $C[0,1]$-valued autoregresive process. *Statist. Infer. for Stoch. Processes* 3, 1-13.

Ragozin, D. (1983). Error bounds for derivatives estimates based on spline smoothing of exact or noisy data. *J. of Approx. Theory*, 37, 335-355.

Ramsay, J.O., and B.W. Silverman (1997). *Functional data analysis.* Springer-Verlag, New York.

Revuz, D., and M. Yor (1991). *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin.

Rice, J., and B. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. of Royal. Statist. Soc. B*, 63, 233-243.

Riesz, F., and B. Nagy (1955). *Leçons d'analyse fonctionnelle.* Gauthier-Villars, Paris.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Ac. Sci. USA* 42, 43-47.

Skorokhod, A.V. (1956). Limit Theorems for stochastics processes. *Theor. Probab. Appl.* 1, 261-290.

Sobczyk, K. (1991). *Stochastic differential equations.* Kluwer Acad. Publ., Dordrecht.

Stoica, G. (1992). Functional LIL's for the Ornstein-Uhlenbeck process. *C.R. Acad. Sci. Paris* 315, I, 1089-1093.

Tikhomirov, A.N. (1991). On the accuracy of the normal approximation of the probability of sums of weakly dependent Hilbert-space-valued random variables hitting a ball. *Theory of Probab. and Appl.* 36, 4, 738-751.

Utev, S. (1995). Personal communication.

Vakhania, N.N., V.I. Tarieladze, and S.A. Chobanyan (1987). *Probability distributions on Banach spaces.* Reidel, Dordrecht.

Varadhan, S.R.S. (1962). Limit theorems for sums of independent random variables with values in a Hilbert space. *Sankhya*, A.24, 213-238.

Weron, A. (1975). Prediction theory in Banach spaces. *Winter school of Probab. Karpacz* 207-226. Springer's LNM.

Yamamoto, T. (1976). Asumptotic mean square prediction error for an autoregressive model with estimated coefficients. *Appl. Statist.* 25, 2, 123-127.

Yamamoto, T. (1981). Predictions of multivariate autoregressive-moving average models. *Biometrika*, 68, 2, 485-492.

Yurinskii, V.V. (1976). Exponential inequalities for sums of random vectors. *J. of Mult. Analysis* 6, 473-499.

Yurinskii, V.V. (1982). On the accuracy of normal approximation of the probability of hitting a ball. *Theor. of Probab. and Appl.* 27, 2, 280-290.

Zhurbenko, I.G., and T.M. Zuparov (1986). The central limit theorem for stationary processes with mixing in a Hilbert space. *Soviet. Math. Dokl.* 33, I, 66-69.

Index

Adapted 63, 267
Addition of variances 7
Adjoint 34
$ARB(1)$, $ARH(1)$, $ARH(p)$ see autoregressive process
$ARMA(p, q)$ 4
$ARMAH(p, q)$ 187
Autocorrelation 20
    operator 80, 149,
Autocovariance 20, 133
    partial 134
Autoregression order 143
Autoregressive process (AR)
    in Banach spaces (ARB) 10, 148, 161, 168
    in $C[0, 1]$ 168
    in Hilbert spaces (ARH) 7, 73, 127
    of order p 127
    real 4, 17
Autoregressive representation 9, 20, 96, 166

Banach space 2, 265
Bandwidth 245
Bergström 152
Berry-Esseen type bound 54, 91
Borel σ-algebra 262
Bounded linear operator 5
B-random variable 21
B-white noise 148, 161

Cauchy-Schwarz inequality 264
Central limit theorem (CLT) 51, 64, 90, 159, 172, 193, 195
Characteristic functional 32
Classification of linear operators 5

Compact operator 34,
Compensated Poisson process 19
Complete
 probability space 262
 space 264
Conditional expectation 5, 29, 33, 266
Consistent predictor 227
Contractive property 29
Convergence
 almost sure 44
 in distribution 45
 in probability 44
 weak 45
Correlation matrix 264
Cotype 50
Covariance
 function 18, 30
 matrix 263
 operator 6, 30, 36, 95
 empirical operator 8, 96, 164
Cross correlation matrix 264
Cross covariance
 function 32
 operator 6, 31, 37, 112
 empirical operator 9, 112
 matrix 264

Data analysis 102
Dependent 54
Diffusion process 18
Dirac measure 76
Distribution 263
 finite-dimensional 16
Dominated convergence 29
Dual 22, 264
Dudley-Skorokhod Theorem 46

Eigenelements 102, 141
Electricity consumption 249
Electrocardiograms 249
El niño southern oscillation (ENSO) 249
Empirical
 covariance function 242
 covariance operator 96, 164
 cross-covariance function 242

cross-covariance operators 9, 112
eigenelements 9, 102, 104
prediction error 243
Equivalent processes 16
Estimation of
autocorrelation operator 12, 203
autocovariance 95, 140, 164
autoregression order 8, 143
covariance operator 8, 13, 95, 140, 194
cross-covariance operators 9, 112
dimensions of eigensubspaces 108
eigenelements 9, 102, 111, 117
mean 13
Expectation 5, 28
conditional 5, 29, 267
in Banach space 27
in $\mathbb{R}$ 263
weak 27

FAP 227, 246
Filtration 1, 63
Finite dimensional 12, 204
Forecasting 2
Fourier series 240
Fourier transform 32
Functional
autoregressive predictor 227, 246
autoregressive process 2
data 237
kernel predictor 245
Function space 1, 21, 265

Gaussian 18; 32, 39, 136, 193, 195
Geometrically strongly mixing (GSM) 58
GIN Theorem 45
$\mathcal{G}$-prediction 40

Harmonic levels 249
Hermetically closed 38
Hilbert-Schmidt operator 5, 34
Hilbert space 2, 264
Hölder condition 11

Identification of p 13
Implementation 237

Independence test 13, 240
Independent stationary increments 18
Indistinguishable 16
Innovation 3, 72, 149, 162
  weak 162
Integrable 27, 263
Interpolation 13, 238
Invertibility 11, 181, 184
Invertible 4
Ito stochastic integral 267

Karnhunen-Loève expansion 25
Kernel
  operator 35
  predictor 245

Langevin 20
Large deviation inequalities 48, 89, 99, 137, 170, 192, 205, 216, 225, 227
Law of large number 55, 86, 87, 136, 194
  strong 47, 90, 194
Law of the iterated logarithm 54, 93, 160, 177
$\mathcal{L}$-closed subspace 6, 38
Lebesgue measure 16, 262
LIL 54, 93, 160, 177
Limit in distribution 118, 221
Linear
  approximation 6
  interpolation 174
  operator 5
  prediction 40
  process 3, 4, 5, 11, 17, 181, 183
Local functional autoregressive predictor 246
LPB 183
LPH 184

Markovian representation 12,128, 188
Markov process
  strict sense 73
  wide sense 73
Markov stationary Gaussian process 20
Martingale difference 63, 97, 162, 166
Mean function 17
Measurability in Banach space 21
Measurable 21, 263
Measure 262

Mercer lemma 24
Modification 16, 23
Moving average 4, 79
Multidimensional case 4

Negligible 262
Nilpotent 79
Nonanticipative 266
Nonparametric 11
Norm
  Hilbert-Schmidt 34
  nuclear 35
  uniform 34
Nuclear operator 6, 35
  in Banach space 156
  in Hilbert space 6

Operator
  bounded 34
  compact 34
  Hilbert-Schmidt 34
  linear 34
  nuclear 35, 96, 112, 156
Optimal rates 8
Ornstein-Uhlenbeck process 10, 20, 27, 76, 150, 178
Orthogonal complement 40
Orthogonality 39
Orthonormal basis 5, 62, 265

Partial autocovariance 134
Poisson process 19
Portmanteau Theorem 45
Positive 18, 32, 35
Prediction
  electricity consumption 13, 249
  electrocardiogram 13, 249
  el niño 13, 249
  road traffic 13, 249
Predictive ability 241
Predictor 243
  functional autoregressive 246
  functional kernel 245
  kernel 245
  local functional autoregressive 246
  naïve 244

Probability space 3, 262
Process
  continuous-time 16
  discrete-time 16
Prokhorov's
  metric 46, 122
  Theorem 46

Random
  function 21
  variable 5, 263
Regular 3
Robustness 13

Sampled data 174
Sample mean 244
Sample path 16
Sample space 15
$SBWN$ 148
Seasonal component 152, 240
Second order process 17
Separable 21, 264
σ-algebra 262
σ-field 262
σ-finite 262
Simulations 13, 247
Smooth FAP 246
Smoothing 239
Spectral decomposition 34, 98
Spline
  cubic 238
  smoothing 174, 239
  with rank constraint (SSRC) 239
Standard
  $ARB(1)$ process 153
  $ARH(1)$ process 79
  $ARH(p)$ process 133
  linear process 183
  Wiener process 18
State space 15
Stationary 19, 65
Statistical methods of prediction 243
Statistical prediction error 227
Stochastic differential equation 10, 18, 151, 267
Stochastic integral 268

Stochastic process 1, 5, 15
Strong law of large numbers (SLLN) 47, 86, 136, 191, 194
Strongly mixing 58, 101,
Support 77
Symmetric operator 35, 82

Test 13, 240
Tight 46
Time-interval(s)
  adjacent 237
  disjoint 237
  overlapping 237
Traffic 249
Trend 13, 240
Type 50

Unbiased estimator 96, 111, 164
Utev's Lemma 153

Validation 243

Wavelets 240
Weakly orthogonal 39
White noise
  in Banach space 148
  in Hilbert space 72
  in $\mathbb{R}$ 3, 17
  strong 72
  weak 161
Wiener process 18, 26, 72
  bilateral 20
  standard 18
Wold decomposition 3, 181

Yule-Walker equations 133

Lecture Notes in Statistics

For information about Volumes 1 to 75,
please contact Springer-Verlag

Vol. 76: L. Bondesson, Generalized Gamma Convolutions and Related Classes of Distributions and Densities. viii, 173 pages, 1992.

Vol. 77: E. Mammen, When Does Bootstrap Work? Asymptotic Results and Simulations. vi, 196 pages, 1992.

Vol. 78: L. Fahrmeir, B. Francis, R. Gilchrist, G. Tutz (Eds.), Advances in GLIM and Statistical Modelling: Proceedings of the GLIM92 Conference and the 7th International Workshop on Statistical Modelling, Munich, 13-17 July 1992. ix, 225 pages, 1992.

Vol. 79: N. Schmitz, Optimal Sequentially Planned Decision Procedures. xii, 209 pages, 1992.

Vol. 80: M. Fligner, J. Verducci (Eds.), Probability Models and Statistical Analyses for Ranking Data. xxii, 306 pages, 1992.

Vol. 81: P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search. xxiii, 526 pages, 1993.

Vol. 82: A. Korostelev and A. Tsybakov, Minimax Theory of Image Reconstruction. xii, 268 pages, 1993.

Vol. 83: C. Gatsonis, J. Hodges, R. Kass, N. Singpurwalla (Editors), Case Studies in Bayesian Statistics. xii, 437 pages, 1993.

Vol. 84: S. Yamada, Pivotal Measures in Statistical Experiments and Sufficiency. vii, 129 pages, 1994.

Vol. 85: P. Doukhan, Mixing: Properties and Examples. xi, 142 pages, 1994.

Vol. 86: W. Vach, Logistic Regression with Missing Values in the Covariates. xi, 139 pages, 1994.

Vol. 87: J. Müller, Lectures on Random Voronoi Tessellations.vii, 134 pages, 1994.

Vol. 88: J. E. Kolassa, Series Approximation Methods in Statistics. Second Edition, ix, 183 pages, 1997.

Vol. 89: P. Cheeseman, R.W. Oldford (Editors), Selecting Models From Data: AI and Statistics IV. xii, 487 pages, 1994.

Vol. 90: A. Csenki, Dependability for Systems with a Partitioned State Space: Markov and Semi-Markov Theory and Computational Implementation. x, 241 pages, 1994.

Vol. 91: J.D. Malley, Statistical Applications of Jordan Algebras. viii, 101 pages, 1994.

Vol. 92: M. Eerola, Probabilistic Causality in Longitudinal Studies. vii, 133 pages, 1994.

Vol. 93: Bernard Van Cutsem (Editor), Classification and Dissimilarity Analysis. xiv, 238 pages, 1994.

Vol. 94: Jane F. Gentleman and G.A. Whitmore (Editors), Case Studies in Data Analysis. viii, 262 pages, 1994.

Vol. 95: Shelemyahu Zacks, Stochastic Visibility in Random Fields. x, 175 pages, 1994.

Vol. 96: Ibrahim Rahimov, Random Sums and Branching Stochastic Processes. viii, 195 pages, 1995.

Vol. 97: R. Szekli, Stochastic Ordering and Dependence in Applied Probability. viii, 194 pages, 1995.

Vol. 98: Philippe Barbe and Patrice Bertail, The Weighted Bootstrap. viii, 230 pages, 1995.

Vol. 99: C.C. Heyde (Editor), Branching Processes: Proceedings of the First World Congress. viii, 185 pages, 1995.

Vol. 100: Wlodzimierz Bryc, The Normal Distribution: Characterizations with Applications. viii, 139 pages, 1995.

Vol. 101: H.H. Andersen, M.Højbjerre, D. Sørensen, P.S.Eriksen, Linear and Graphical Models: for the Multivariate Complex Normal Distribution. x, 184 pages, 1995.

Vol. 102: A.M. Mathai, Serge B. Provost, Takesi Hayakawa, Bilinear Forms and Zonal Polynomials. x, 378 pages, 1995.

Vol. 103: Anestis Antoniadis and Georges Oppenheim (Editors), Wavelets and Statistics. vi, 411 pages, 1995.

Vol. 104: Gilg U.H. Seeber, Brian J. Francis, Reinhold Hatzinger, Gabriele Steckel-Berger (Editors), Statistical Modelling: 10th International Workshop, Innsbruck, July 10-14th 1995. x, 327 pages, 1995.

Vol. 105: Constantine Gatsonis, James S. Hodges, Robert E. Kass, Nozer D. Singpurwalla(Editors), Case Studies in Bayesian Statistics, Volume II. x, 354 pages, 1995.

Vol. 106: Harald Niederreiter, Peter Jau-Shyong Shiue (Editors), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. xiv, 372 pages, 1995.

Vol. 107: Masafumi Akahira, Kei Takeuchi, Non-Regular Statistical Estimation. vii, 183 pages, 1995.

Vol. 108: Wesley L. Schaible (Editor), Indirect Estimators in U.S. Federal Programs. viii, 195 pages, 1995.

Vol. 109: Helmut Rieder (Editor), Robust Statistics, Data Analysis, and Computer Intensive Methods. xiv, 427 pages, 1996.

Vol. 110: D. Bosq, Nonparametric Statistics for Stochastic Processes. xii, 169 pages, 1996.

Vol. 111: Leon Willenborg, Ton de Waal, Statistical Disclosure Control in Practice. xiv, 152 pages, 1996.

Vol. 112: Doug Fischer, Hans-J. Lenz (Editors), Learning from Data. xii, 450 pages, 1996.

Vol. 113: Rainer Schwabe, Optimum Designs for Multi-Factor Models. viii, 124 pages, 1996.

Vol. 114: C.C. Heyde, Yu. V. Prohorov, R. Pyke, and S. T. Rachev (Editors), Athens Conference on Applied Probability and Time Series Analysis Volume I: Applied Probability In Honor of J.M. Gani. viii, 424 pages, 1996.

Vol. 115: P.M. Robinson, M. Rosenblatt (Editors), Athens Conference on Applied Probability and Time Series Analysis Volume II: Time Series Analysis In Memory of E.J. Hannan. viii, 448 pages, 1996.

Vol. 116: Genshiro Kitagawa and Will Gersch, Smoothness Priors Analysis of Time Series. x, 261 pages, 1996.

Vol. 117: Paul Glasserman, Karl Sigman, David D. Yao (Editors), Stochastic Networks. xii, 298, 1996.

Vol. 118: Radford M. Neal, Bayesian Learning for Neural Networks. xv, 183, 1996.

Vol. 119: Masanao Aoki, Arthur M. Havenner, Applications of Computer Aided Time Series Modeling. ix, 329 pages, 1997.

Vol. 120: Maia Berkane, Latent Variable Modeling and Applications to Causality. vi, 288 pages, 1997.

Vol. 121: Constantine Gatsonis, James S. Hodges, Robert E. Kass, Robert McCulloch, Peter Rossi, Nozer D. Singpurwalla (Editors), Case Studies in Bayesian Statistics, Volume III. xvi, 487 pages, 1997.

Vol. 122: Timothy G. Gregoire, David R. Brillinger, Peter J. Diggle, Estelle Russek-Cohen, William G. Warren, Russell D. Wolfinger (Editors), Modeling Longitudinal and Spatially Correlated Data. x, 402 pages, 1997.

Vol. 123: D. Y. Lin and T. R. Fleming (Editors), Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. xiii, 308 pages, 1997.

Vol. 124: Christine H. Müller, Robust Planning and Analysis of Experiments. x, 234 pages, 1997.

Vol. 125: Valerii V. Fedorov and Peter Hackl, Model-oriented Design of Experiments. viii, 117 pages, 1997.

Vol. 126: Geert Verbeke and Geert Molenberghs, Linear Mixed Models in Practice: A SAS-Oriented Approach. xiii, 306 pages, 1997.

Vol. 127: Harald Niederreiter, Peter Hellekalek, Gerhard Larcher, and Peter Zinterhof (Editors), Monte Carlo and Quasi-Monte Carlo Methods 1996, xii, 448 pages, 1997.

Vol. 128: L. Accardi and C.C. Heyde (Editors), Probability Towards 2000, x, 356 pages, 1998.

Vol. 129: Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov, Wavelets, Approximation, and Statistical Applications, xvi, 265 pages, 1998.

Vol. 130: Bo-Cheng Wei, Exponential Family Nonlinear Models, ix, 240 pages, 1998.

Vol. 131: Joel L. Horowitz, Semiparametric Methods in Econometrics, ix, 204 pages, 1998.

Vol. 132: Douglas Nychka, Walter W. Piegorsch, and Lawrence H. Cox (Editors), Case Studies in Environmental Statistics, viii, 200 pages, 1998.

Vol. 133: Dipak Dey, Peter Müller, and Debajyoti Sinha (Editors), Practical Nonparametric and Semiparametric Bayesian Statistics, xv, 408 pages, 1998.

Vol. 134: Yu. A. Kutoyants, Statistical Inference For Spatial Poisson Processes, vii, 284 pages, 1998.

Vol. 135: Christian P. Robert, Discretization and MCMC Convergence Assessment, x, 192 pages, 1998.

Vol. 136: Gregory C. Reinsel, Raja P. Velu, Multivariate Reduced-Rank Regression, xiii, 272 pages, 1998.

Vol. 137: V. Seshadri, The Inverse Gaussian Distribution: Statistical Theory and Applications, xi, 360 pages, 1998.

Vol. 138: Peter Hellekalek, Gerhard Larcher (Editors), Random and Quasi-Random Point Sets, xi, 352 pages, 1998.

Vol. 139: Roger B. Nelsen, An Introduction to Copulas, xi, 232 pages, 1999.

Vol. 140: Constantine Gatsonis, Robert E. Kass, Bradley Carlin, Alicia Carriquiry, Andrew Gelman, Isabella Verdinelli, Mike West (Editors), Case Studies in Bayesian Statistics, Volume IV, xvi, 456 pages, 1999.

Vol. 141: Peter Müller, Brani Vidakovic (Editors), Bayesian Inference in Wavelet Based Models, xi, 394 pages, 1999.

Vol. 142: György Terdik, Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach, xi, 258 pages, 1999.

Vol. 143: Russell Barton, Graphical Methods for the Design of Experiments, x, 208 pages, 1999.

Vol. 144: L. Mark Berliner, Douglas Nychka, and Timothy Hoar (Editors), Case Studies in Statistics and the Atmospheric Sciences, x, 208 pages, 1999.

Vol. 145: James H. Matis and Thomas R. Kiffe, Stochastic Population Models, viii, 220 pages, 2000.

Vol. 146: Wim Schoutens, Stochastic Processes and Orthogonal Polynomials, xiv, 163 pages, 2000.

Vol. 147: Jürgen Franke, Wolfgang Härdle, and Gerhard Stahl, Measuring Risk in Complex Stochastic Systems, xvi, 272 pages, 2000.

Vol. 148: S.E. Ahmed and Nancy Reid, Empirical Bayes and Likelihood Inference, x, 200 pages, 2000.

Vol. 149: D. Bosq, Linear Processes in Function Spaces: Theory and Applications, xiv, 296 pages, 2000.