Probabilistic and Statistical Tools for Modeling Time Series

Publicações Matemáticas

Probabilistic and Statistical Tools for Modeling Time Series

Paul Doukhan
University Cergy-Pontoise &
French Universitary Institute

impa

$30^{\underline{o}}$ Colóquio Brasileiro de Matemática

Impresso no Brasil / Printed in Brazil
Capa: Noni Geiger / Sérgio R. Vaz

30º Colóquio Brasileiro de Matemática

- Aplicações Matemáticas em Engenharia de Produção - Leonardo J. Lustosa e Fernanda M. P. Raupp
- Boltzmann-type Equations and their Applications - Ricardo Alonso
- Dissipative Forces in Celestial Mechanics - Sylvio Ferraz-Mello, Clodoaldo Grotta-Ragazzo e Lucas Ruiz dos Santos
- Economic Models and Mean-Field Games Theory - Diogo A. Gomes, Levon Nurbekyan and Edgard A. Pimentel
- Generic Linear Recurrent Sequences and Related Topics - Letterio Gatto
- Geração de Malhas por Refinamento de Delaunay - Marcelo Siqueira, Afonso Paiva e Paulo Pagliosa
- Global and Local Aspects of Levi-flat Hypersurfaces - Arturo Fernández Pérez e Jiří Lebl
- Introdução às Curvas Elípticas e Aplicações - Parham Salehyan
- Métodos de Descida em Otimização Multiobjetivo - L. M. Graña Drummond e B. F. Svaiter
- Modern Theory of Nonlinear Elliptic PDE - Boyan Slavchev Sirakov
- Novel Regularization Methods for Ill-posed Problems in Hilbert and Banach Spaces - Ismael R. Bleyer e Antonio Leitão
- **Probabilistic and Statistical Tools for Modeling Time Series - Paul Doukhan**
- Tópicos da Teoria dos Jogos em Computação - R. C. S. Schouery, O. Lee, F. K. Miyazawa e E. C. Xavier
- Topics in Spectral Theory - Carlos Tomei

Distribuição: IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br
http://www.impa.br

ISBN: 978-85-244-0411-5

Cet ouvrage est dédié à Jean et Sarah, que j'ai participé à faire vivre et qui me le rendent au centuple...

Contents

Objectives

Time series appear naturally with data sampled in the time but many other physical situations let also appear such evolutions indexed by integers.
We aim at providing some tools for the study of such statistical models. The purpose of those lectures is introductory and definitely no systematic study will be proposed here.

Those notes are divided into 3 Parts including each 4 Chapters and an Appendix.

1. Independence and Stationarity.
   Even wether this part addresses mainly items of the independent world, the choice of subjects is biased and motivated by the fact that they easily extend to a dependent setting.

   (a) Independence.
   This is a main concept in those notes so we set some simple comments concerning independence as a separate chapter. For instance we mention all the elementary counter-examples invoking independence. Other examples relating orthogonality with independence may be found in the Chapter 8 and in Appendix, § A.2.1.

   (b) Gaussian convergence and moments.
   A special emphasis is set on Lindeberg method with easily extends to a dependent setting. Applications of the central limit theorems are proved in the independent setting. Moment and exponential inequalities related to Gaussian convergence are also derived.

   (c) Estimation concepts.
   Classical estimations techniques, as empirical ones, contrasts and non-parametric techniques are introduced. Kernel density estimates are described with some details as an application of previous results in view of their extension to time series in a further Chapter.

   (d) Stationarity.
   The notions of stationarity are essential for spectral analysis of time series. [Brockwell and Davis, 1991] use filtering techniques in order to return to such a simple stationary case. Indeed this assumption is not naturally observed. Weak and strong stationarity are considered together with examples.

Second order weak dependence or long range dependence are defined according the convergence of the series of covariances. Stationarity and an introduction to spectral techniques are provided after this. We precise the spectral representation for both a covariance and for the process itself. and we rapidly scan some applications of time series.

2. Models of time series.

   (a) Gaussian chaos.
   Due to the CLT the Gaussian case plays a central role in statistics. The first time series to be considered are Gaussian. We introduce the Gaussian chaos and Hermite polynomials as well as some of their properties. Gaussian processes and the methods of the Gaussian chaos are thus investigated. Namely Hermite representations and Mehler formula for functions of Gaussian processes are developed precisely while the diagram formula for higher order moments is simply considered. The fractional Brownian motion essential hereafter for the long range dependent setting is also introduced. The asymptotic theory for Gaussian functionals is also precisely stated. We also recall the 4th moment method based on Malliavin calculus.

   (b) Linear models.
   From Lindeberg's lemma, the linear case is the second case to consider after the Gaussian one. Eg. ARMA shortly dependent processes or long range dependent models such as FARIMA models are provided. See [Brockwell and Davis, 1991] for further information.

   (c) Nonlinear models.
   This central Chapter proposes a wide botanic for models of time series. Non linear models are naturally considered as extensions of the previous ones. After the elementary ideas of polynomials and chaoses we come to an algebraic approach of models explicit solutions of a recursion equation. Then more general and non explicit contractive iterative systems are introduced together with a variety of examples. Finally the abstract Bernoulli shifts yield a general and simple overview of those various examples; their correlation properties are explicitly provided. This class of general non linear functionals of independent sequences yields a large amount of examples.

   (d) Association.
   Associated processes are then rapidly investigated. This prop-

erty was introduced for reliability and for statistical physics. The association property admits a main common point with Gaussian case: independence and orthogonality coincide in both cases. This feature is exploited in the following chapter.

3. Dependences.

   (a) Ergodic theorem.
   As an extension of the strong law of large numbers, the ergodic theorem is the first result proposed in this chapter. In order to get confidence bounds for asymptotic distribution of the mean one first needs consistency of the empirical mean. Further needed asymptotic expansions are obtained from SRD/LRD properties.
   We then make a tour of the tools for the asymptotic theory under long range or short range dependence (resp. SRD and LRD).

   (b) Long Range Dependence.
   Under LRD the more elementary examples are are seen to get such asymptotic explicit expansion in distribution up to non-Gaussian limits. Gaussian and subordinated Gaussians are first considered as well as linear LRD models, anyway a rapid description of non linear LRD models is also included.

   (c) Short Range Dependence.
   In the SRD case we give a rapid idea of techniques. Namely the standard Bernstein blocks technique is proposed as a way to derive CLTs by using a recent dependent Lindeberg approach.

   (d) Moment methods.
   A last chapter is devoted to moment and cumulant inequalities developing the more standard spectral ideas of the second chapter.
   Such inequalities are needed in many occasions but first in order to derive CLTs, another application is for subsampling. This technique applies for the kernel density estimator.

Appendices.

   (a) Probability.
   A first Appendix recalls essential concepts of probability, including repartition functions and some Hoeffding inequalties.

   (b) Distributions.
   Useful examples of probability distributions are introduced in

relation with the dependence conditions. Standard Gaussians, Gaussian vectors and $\gamma-$type distributions are considered.

(c) Limit theory in probability.
Basic concepts of convergence in probability theory are recalled in this Appendix.

Applications of those techniques to spectral estimations are developed in an elegant way in monographs [Rosenblatt, 1985, Rosenblatt, 1991]. Relations with the asymptotic theory for kernel density estimation are also given.
[Azencott and Dacunha-Castelle, 1987] and [Rosenblatt, 1985] also lead to a large amount of additional developments. Functional estimation frames are synthetically described in [Rosenblatt, 1991].
[Doukhan et al., 2002b] provides a wide amount of directions for the study of LRD.
[Doukhan and Louhichi, 1999], [Dedecker and Doukhan, 2003] as well as [Dedecker et al., 2007] also consider the weakly dependent setting.

Paris, May 20, 2015

Paul Doukhan

Acknowledgements Preliminary versions of those notes were used in Universities of (Bogotá UNAL Colombia), Cergy-Pontoise & Paris 6 (France), Hong-Kong (HKU), Porto Alegre (Brazil), and Torún (Poland).
They Adam Jakubowski (Torun), Konstantinos Fokianos (Cyprus), Yvan Nourdin (Luxembourg), and Gabriel Lang (AgroParitech) provided me with many useful suggestions. Also I wish to thank Natalia Bahamonde (Valparaiso), Jean Marc Bardet (Paris), Emilie Chautru (Ecole des Mines), Alain Latour (Grenoble) and Xiaoyin Li (Cleveland) for their kind help to provide me with illustrative helpful and comprehensive figures. An anonymous referee also helped me for a more suitable organization of the included material.
My main thanks are for Silvia and Artur Lopes from Porto Alegre for their constant support and help in the final redaction of those notes. I finally thank IMPA (Rio de Janeiro) for the invitation to give an advanced course based on those notes.
This work has been developed within the MME-DII center of excellence (ANR-11-LABEX-0023-01).

Part I

Independence and stationarity

Chapter 1

Independence

Definition 1.1.1. *Events $A, B \in \mathcal{A}$ are called independent in case*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Definition 1.1.2. *The random variables $X_1, \ldots, X_n$ (with values for instance in the same space topological E) are said to be independent in case, for any $g_1, \ldots, g_n : E \to \mathbb{R}$ continuous and bounded:*

$$\mathbb{E}\Big(g_1(X_1) \times \cdots \times g_n(X_n)\Big) = \big(\mathbb{E}g_1(X_1)\big) \times \cdots \times \big(\mathbb{E}g_n(X_n)\big).$$

Definition 1.1.3. *Events $A_1, \ldots A_n$ are called independent if the random variables $X_1 = 1\!\!1_{A_1}, \ldots, X_n = 1\!\!1_{A_n}$ are independent.*
In other words for each $E \subset \{1, \ldots, n\}$

$$\mathbb{P}\Big(\bigcap_{i \in E} A_i\Big) = \prod_{i \in E} \mathbb{P}(A_i).$$

Definition 1.1.4. *The random variables $X_1, \ldots, X_n$ are called pairwise independent if each couple X_i, X_j is independent for $i \neq j$.*

In case the characteristic function is analytic around 0, and $E = \mathbb{R}$ previous remarks imply that the previous identity is enough to prove the the independence of $X_1, \ldots, X_n$ if

$$\phi_{X_1 + \cdots + X_n} = \phi_{X_1} \times \cdots \times \phi_{X_n}.$$

Assume now that X_j admits a density f_j with respect to some measure ν_j on E_j then the random vector $(X_1, \dots, X_d) \in E_1 \times \cdots \times E_d$ admits the density

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d), \quad (x_1, \dots, x_d) \in E_1 \times \cdots \times E_d$$

on the product space $E_1 \times \cdots \times E_d$ with respect to $\nu_1 \times \cdots \times \nu_d$.

If $A_1, \dots, A_d \in \mathcal{A}$ are events then simple random variables write $X_k = 1_{A_k} \in \{0, 1\}$ and the independence of couples (X_i, X_j) is easily proved to coincide with the independence of couples of events A_i, A_j. Anyway the independence of the family of events $A_1, \dots, A_d$ writes a bit differently, as:

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i), \qquad \forall I \subset \{1, \dots, d\}.$$

Example 1.1.1. *As a probability space consider a model* $(\Omega, \mathcal{A}, \mathbb{P})$ *for two (fair) independent dices*

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2, \qquad \mathcal{A} = \mathcal{P}(\Omega),$$

and $\mathbb{P}$ *is the uniform probability on this finite set with 36 elements. Let* A, B *be the events that the dices show an even number, then*

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$$

and those events are independent.
Now let C *be the event that the sum of the results in both dices is also even then* $A \cap B \subset C$ *and on the event* $A \cap C$ *the second dice is necessarily even too, so that* $A \cap C \subset B$.
Analogously $B \cap C \subset A$ *so that it is easy to check that* A, C *and* B, C *are independent pairs of events,*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C),\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C)$$

(those values all equal $\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$*).*
From the other hand $A \cap B \cap B = A \cap B$ *thus*

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{1}{8}$$

and the triplet of events (A, B, C) is not independent.
Then the events A, B, C are pairwise independent but not independent on this probability set with 36 elements equipped with the uniform law.

Remark 1.1.1.

- *As sketched in the previous Exercise, this is possible to find 3 pairwise independent random variables which are not independent ($X = \mathbb{1}_A$, $Y = \mathbb{1}_B$ and $Z = \mathbb{1}_C$). Pairwise independence should thus be carefully distinguished from independence.*
  Precisely for each p there exist a vector $X = (X_1, \ldots, X_p) \in \mathbb{R}^p$ which components are not independent but such that any vector with dimension strictly less than p and components among $X_1, \ldots, X_p$ is independent, [Derriennic and Klopotowski, 2000] and [Bradley and Pruss, 2009] for additionally a counter-example to the CLT.

- *Now quote that the Example A.1.3 provides us with a whole sequence of independent random variables with a given distribution on $\mathbb{R}$.*

- *Let $X_1, \ldots, X_n$ be independent $b(p)-$distributed random variables, then the calculation of generating functions implies that $X_1 + \cdots + X_n \sim B(n, p)$ admits a Binomial distribution.*

The following essential but very simple result is also stated as an Exercise:

Exercise 1. *Let $X, Y \in \mathbb{R}$ be real valued random variables with $\mathbb{E}X^2 + \mathbb{E}Y^2 < \infty$. If (X, Y) are independent then $Cov(X, Y) = 0$.*

Solution to Exercise 1. In case those variables are bounded, then independence allows indeed to assert that $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.
The general case is derived from a truncation by setting $X_M = X \vee (-M) \wedge M$ and the use of Lebesgue dominated convergence theorem with $M \uparrow \infty$. ∎

Exercise 2. *Let $X, R \in \mathbb{R}$ be independent random variables with X symmetric (i.e. $-X$ admits the same distribution as X), $\mathbb{E}X^2 < \infty$ and $\mathbb{P}(R = \pm 1) = \frac{1}{2}$, set $Y = RX$.*

- *$Cov(X,Y) = 0$,*
- *moreover if $|X|$ is not a.s. constant then X, Y are not independent.*

An important use of this Exercise in provided in Exercise 23.

Solution to Exercise 2. The first equality follows as well from independence in the case of bounded X and dominated convergence yields the general case as in Exercise 1.
The second result also follows since because $|X|$ is not a.s. constant there is an even function g such that $\mathrm{Var}\, g(X) \neq 0$, now we have: $\mathrm{Cov}\,(g(X), g(Y)) \neq 0$. ∎

Exercise 3. *If random variables $X, Y \in \{0,1\}$ satisfy $Cov(X,Y) = 0$, prove that X, Y are independent.*

Solution of Exercise 3. To prove the independence of those random variables one needs to prove the independence of events (A_a, B_b) for all $a, b \in \{0,1\}$, with $A_a = (X = a)$ and $B_b = (Y = b)$.

- Relation $\mathrm{Cov}\,(X,Y) = 0$ writes as the independence of the events A_1, B_1,
- Relation $\mathrm{Cov}\,(X, 1-Y) = 0$ writes as the independence of events A_1, B_0,
- Relation $\mathrm{Cov}\,(1-X, Y) = 0$ is independence of A_0, B_1,
- Relation $\mathrm{Cov}\,(1-X, 1-Y) = 0$ is independence of A_0, B_0. ∎

Quote that either Gaussian or associated vectors fit the same property see in Appendix A.2.1, and Chapter 8 respectively.
This Exercise 3 admits tight assumptions as suggests the following:

Exercise 4. *Exhibit random variables $X \in \{0, \pm 1\}, Y \in \{0,1\}$ not independent, but orthogonal anyway, i.e. $Cov(X,Y) = 0$.*

Solution for Exercise 4. Consider the uniform random variable X on the set $\{-1, 0, 1\}$ and $Y = \mathbb{1}_{\{X=0\}}$, then $\mathbb{E}X = 0$, $\mathrm{Cov}\,(X, Y) = \mathbb{E}XY = 0$ because $XY \equiv 0$ *(a.s.)* while those random variables are not independent.
Indeed with $f(x) = \mathbb{1}_{\{x=0\}}$ and $g(x) = x$ we derive

$$\mathbb{E}f(X)g(Y) = \mathbb{P}(X = 1) \neq \mathbb{E}f(X)\mathbb{E}g(Y) = \mathbb{P}^2(X = 1). \qquad \blacksquare$$

Chapter 2

Gaussian convergence and inequalities

The Chapter aims at processing the Gaussian limit theory, namely we precise some Central Limit Theorems together with applications and moment/exponential inequalities for partial sums behaving asymptotically as Gaussian random variables.

2.1 Gaussian Convergence

This is a well-know feature that accumulation of infinitesimal independent random effect are accurately approximated by the Gaussian distribution. The best illustration of this fact is explained by Lindeberg method.

Lemma 2.1.1 (Lindeberg)**.** *Assume that $U_1, \ldots, U_k$ are centered real valued random variables.*
Let $V_1, \ldots, V_k$ be independent random variables, independent of the random variables $U_1, \ldots, U_k$ and such that $U_j \sim \mathcal{N}(0, \mathbb{E}U_j^2)$ and $g \in C_b^3$.
Set $U = U_1 + \cdots + U_k$ and $V = V_1 + \cdots + V_k$ then we obtain the two

bounds:

$$\begin{aligned}
|\mathbb{E}(g(U)-g(V))| &\leq 3\sum_{i=1}^{k}\mathbb{E}\left(|U_i|^2\left(\|g''\|_\infty\wedge(\|g'''\|_\infty|U_i|)\right)\right). \\
&\leq 3\,\|g''\|_\infty^{1-\epsilon}\|g'''\|_\infty^{\epsilon}\sum_{i=1}^{k}\mathbb{E}|U_j|^{2+\epsilon}.
\end{aligned}$$

Proof of Lemma 2.1.1. Set $Z_j = U_1+\cdots+U_{j-1}+V_{j+1}+\cdots+V_k$ for $1\leq j\leq k$ then

$$\mathbb{E}(g(U)-g(V)) = \sum_{j=1}^{k}\mathbb{E}\left(g(Z_j+U_j)-g(Z_j+V_j)\right) = \sum_{j=1}^{k}\mathbb{E}\delta_j.$$

Set now $\delta = g(z+u)-ug'(z)-\frac{1}{2}u^2g''(z)$ then Taylor formula at order 2 entails $|\delta|\leq\frac{1}{2}u^2|g''(z)-g''(t)|$ for some $t\in]z,z+u[$.
This implies from either the mean value theorem or from a simple bound that

$$\begin{aligned}
|\delta| &\leq (u^2\|g''\|_\infty)\wedge(\frac{1}{2}|u|^3\|g'''\|_\infty) \\
&= (u^2\|g''\|_\infty)\left(1\wedge\left(\frac{1}{2}|u|\frac{\|g'''\|_\infty}{\|g''\|_\infty}\right)\right) \\
&\leq u^2\|g''\|_\infty\left(\frac{1}{2}|u|\frac{\|g'''\|_\infty}{\|g''\|_\infty}\right)^\epsilon \\
&= 2^{-\epsilon}|u|^{2+\epsilon}\|g''\|_\infty^{1-\epsilon}\|g'''\|_\infty^{\epsilon}
\end{aligned}$$

Apply the above inequality with $z = Z_j$ and $u = U_j$ or V_j. To conclude we also quote that $\mathbb{E}|V_j|^2 = \mathbb{E}|U_j|^2$ and thus $\mathbb{E}|V_j|^3 = \mathbb{E}|Z|^3\left(\mathbb{E}U_j^2\right)^{3/2}$ for a standard normal random variable $Z\sim\mathcal{N}(0,1)$.
Hölder inequality thus yields $\left(\mathbb{E}U_j^2\right)^{3/2}\leq\mathbb{E}|U_j|^3$.
An integration by parts implies $\mathbb{E}|Z|^3 = \frac{4}{\sqrt{2\pi}} < 2$. Hence from Jensen inequality (Proposition A.1.1) we derive $\mathbb{E}|V|^{2+\epsilon} < 2\mathbb{E}|V|^{2+\epsilon}$.
Now

$$\begin{aligned}
\mathbb{E}|\delta_j| &\leq 2^{-\epsilon}\|g''\|_\infty^{1-\epsilon}\|g'''\|_\infty^{\epsilon}\mathbb{E}\left(|U_j|^{2+\epsilon}+|V_j|^{2+\epsilon}\right) \\
&\leq 3\|g''\|_\infty^{1-\epsilon}\|g'''\|_\infty^{\epsilon}\mathbb{E}|U_j|^{2+\epsilon}
\end{aligned}$$

This yields the desired result. ∎

As a simple consequence of this result we derive:

Theorem 2.1.1 (Lindeberg)**.** *Let $(\zeta_{n,k})_{k\in\mathbb{Z}}$ be independent identically distributed sequences of centered random variable (for each n). Suppose*

$$\begin{aligned}\sum_k \mathbb{E}\zeta_{n,k}^2 &\to_{n\to\infty} \sigma^2 > 0,\\ \sum_k \mathbb{E}\zeta_{n,k}^2\, 1\!\!1_{\{|\zeta_{n,k}|>\epsilon\}} &\to_{n\to\infty} 0, \text{ for each } \epsilon > 0.\end{aligned}$$

Then:

$$\sum_k \zeta_{n,k} \xrightarrow{\mathcal{L}}_{n\to\infty} \mathcal{N}(0,\sigma^2).$$

Proof. In the first inequality from Lemma 2.1.1, set $U_k = \zeta_{n,k} 1\!\!1_{\{|\zeta_{n,k}|\le\epsilon\}}$ for a convenient $\epsilon > 0$ then the first assumption implies that

$$\sup_n \sum_k \mathbb{E}\zeta_{n,k}^2 \equiv C < \infty,$$

and thus setting $\zeta_n = \sum_k^n \zeta_{n,k}$, we derive

$$\sum_{k=1}^n \mathbb{E}|U_k|^3 \le C\cdot\epsilon.$$

Now from independence,

$$\mathbb{E}\left(\zeta_n - U\right)^2 \le \sum_k \mathbb{E}\zeta_{n,k}^2\, 1\!\!1_{\{|\zeta_{n,k}|>\epsilon\}} \equiv a_n(\epsilon).$$

Thus the triangular inequality implies $\sigma_n^2 = \mathbb{E}U^2 \to_{n\to\infty} \sigma^2$. Those bounds together imply for $Z \sim \mathcal{N}(0,1)$ a normal random variable:

$$\begin{aligned}|\mathbb{E}g(\zeta_n) - g(\sigma Z)| \le |\mathbb{E}g(\zeta_n) - g(U)|&\\ + |\mathbb{E}g(U) - g(\sigma_n Z)|&\\ + |\mathbb{E}g(\sigma_n Z) - g(\sigma Z)|&.\end{aligned}$$

To prove the result use $|\mathbb{E}g(\sigma_n Z) - g(\sigma Z)| \leq \|g'\|_\infty \mathbb{E}|Z||\sigma_n - \sigma|$ and select $\epsilon = \epsilon_n$ conveniently such that $\lim_n (a_n(\epsilon_n) + \epsilon_n) = 0$.
Then the result follows. ∎

In order to prove the power of this result the forthcoming subsections aim at deriving some other consequences of Lindeberg lemma, see the beautiful book [van der Vaart, 1998] for much more.
Namely the classical central limit Theorem 2.1.2 is a first consequence of this result.
Then the asymptotic behavior of empirical medians will derived in the Proposition 2.1.1 following the proof in [van der Vaart, 1998].
Finally the validity of the Gaussian approximation of binomial distributions is essential for example in order to assert the validity of χ^2-goodness-of-fit tests.

To conclude this section quote that we will present a simple dependent version of the Lindeberg lemma [Bardet et al., 2006] in Lemma 11.5.1 below.

2.1.1 Central Limit Theorem

Theorem 2.1.2. *The central limit theorem ensures the convergence*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) \to^{\mathcal{L}} \mathcal{N}(0, \mathbb{E}X_0^2),$$

for independent identically distributed sequences with finite variance.

Proof. This follows from Theorem 2.1.1. Set $\zeta_{n,k} = X_k/\sqrt{n}$ the only point to check is now $\lim_{n\to\infty} \mathbb{E}X_1^2 \, \mathbb{1}_{|X_1| \geq \epsilon\sqrt{n}} = 0$, which follows from $\mathbb{E}X_1^2 < \infty$ ([1]). ∎

---

[1]An alternative proof may be given by using Lemma 2.1.1 with $k = n$ and $U_j = X_j/\sqrt{n}$.
To prove it simply note that such a random variable X_0 satisfies the tightness condition

$$\mathbb{E}|X_0|^2 \wedge \left(\frac{|X_0|^3}{\sqrt{n}}\right) \to_{n\to\infty} 0.$$

Indeed we let it as an exercise that if $\mathbb{E}X_0^2 < \infty$ then there exists a function $H : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{x\to\infty} H(x)/x^2 = \infty$, $\mathbb{E}H(|X_0|) < \infty$ (symmetric

2.1.2 Empirical median

Suppose that the number of observations $n = 2N+1$ is even; we consider here an independent identically distributed $n-$sample $Y_1, \dots, Y_n$ with median M

$$\mathbb{P}(Y_1 < M) \le \frac{1}{2} \le \mathbb{P}(Y_1 > M).$$

To simplify notations assume this law is continuous.
The empirical median of the sample is the value M_n of the order statistic with rank $N+1$.

Proposition 2.1.1. *Assume that (X_k) is an atomless identically distributed and independent sequence. If the cumulative repartition function F of Y_1 admits a derivative γ at point M then*

$$\sqrt{n}(M_n - M) \xrightarrow{\mathcal{L}}_{n\to\infty} \mathcal{N}\left(0, \frac{1}{4\gamma^2}\right).$$

Proof. Notice that $\mathbb{P}(\sqrt{n}(M_n - M) \le x) = \mathbb{P}(M_n \le M + x/\sqrt{n})$ is the probability that $N+1$ observations Y_i (among the $n = 2N+1$ considered) satisfy $Y_i \le M + x/\sqrt{n}$:

$$\mathbb{P}(\sqrt{n}(M_n - M) \le x) = \mathbb{P}\Big(\sum_{i=1}^n 1\!\!1_{\{Y_i \le M + x/\sqrt{n}\}} \ge N+1\Big).$$

Setting $p_n = \mathbb{P}(Y_1 \le M + x/\sqrt{n})$ and

$$X_{i,n} = \frac{1\!\!1_{\{Y_i \le M + x/\sqrt{n}\}} - p_n}{\sqrt{np_n(1-p_n)}},$$

yields

$$\mathbb{P}(\sqrt{n}(M_n - M) \le x) = \mathbb{P}\Big(s_n \le \sum_{i=1}^n X_{i,n}\Big), \qquad s_n = \frac{N+1-np_n}{\sqrt{np_n(1-p_n)}}.$$

---

and non decreasing on $\mathbb{R}^+$). For each $k > 0$ there exists $M_k > 0$ that we may choose non decreasing and such that $\mathbb{E}|X_0| 1\!\!1_{\{|X_0| \ge M_k\} \le \frac{1}{k^2}}$. Set $H(x) = kx^2$ for $M_k \le |x| < M_{k+1}$.

The continuity of Y_1 distribution at point M implies $p_n \to \frac{1}{2}$ and its derivability yields $s_n \to -2x\gamma$.
Lindeberg theorem thus yields $\sum_{i=1}^n X_{i,n} \to \mathcal{N}(0,1)$, which allows to conclude. ∎

Remark 2.1.1. *If instead of the continuity of X_0's cdf (the atomless assumption) we deal with more general properties then only the regularity around the median is really required.*

2.1.3 Gaussian approximation for binomials

Theorem 2.1.3. *Let $S_n \sim B(n,p)$ and fix some $\epsilon \in (0,1]$, then*

$$\sup_{np(1-p)\epsilon>1} \sup_{u\in\mathbb{R}} \Delta_{n,p}(u) = \mathcal{O}\Big((np(1-p))^{-\frac{1}{8}}\Big),$$

with

$$\Delta_{n,p}(u) = \Big|\mathbb{P}_p\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \le u\Big) - \Phi(u)\Big|.$$

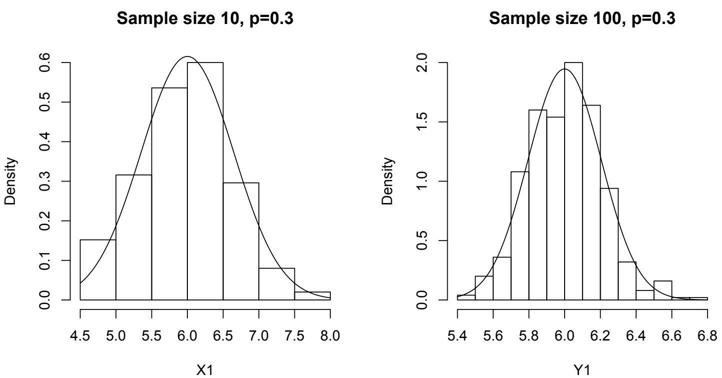


Figure 2.1: Accuracy of Gaussian approximation from binomials.

Remark 2.1.2. *This result is not optimal and exponent* $\frac{1}{8}$ *may be replaced by* $\frac{1}{2}$, *see [Petrov, 1975].*
Anyway it allows to validate the Gaussian approximation if the product $np(1-p)$ *is large as a classical heuristics tells us: in statistics* $np \geq 5$ *is the condition used to use a Gaussian approximation of binomials. Figure 2.1 demonstrates the evolution of this Gaussian approximation.*

Proof. Use Lemma 2.1.1. Rewrite $S_n = b_1 + \cdots + b_n$ with iid $b_1, b_2, \ldots \sim b(p)$. Set

$$X_i = \frac{b_i - p}{\sqrt{np(1-p)}}, \qquad 1 \leq i \leq n.$$

Then $X_1, \ldots, X_n$ are centered independent identically distributed and

$$\mathbb{E}_p b_i^3 = \mathbb{E}_p (b_i - p)^2 = p(1-p).$$

Let $0 < p \leq 1 - \epsilon$ then for $f \in \mathcal{C}^3$ we get from the Lemma 2.1.1, with some $Z \sim \mathcal{N}(0,1)$:

$$\begin{aligned}
\Delta_n(f) &= \left|\mathbb{E}_p f\left(\frac{S_n - np}{\sqrt{n\theta(1-p)}}\right) - f(Z)\right| \\
&\leq \frac{\|f'''\|_\infty}{2} \sum_{i=1}^n \mathbb{E}|X_i|^3 \\
&\leq \frac{4\|f'''\|_\infty}{\epsilon} \frac{1}{\sqrt{np(1-p)}}
\end{aligned}$$

In order to conclude one needs to prove *Exercise* 5 below.
Using $\mathbb{P}(Z \in [u, u+\eta]) \leq \eta/\sqrt{2\pi}$ we then derive

$$\Delta_n(f_{u-\eta,\eta}) + \mathbb{P}(Z \in [u, u-\eta]) \leq \Delta_{n,p}(u) \leq \Delta_n(f_{u,\eta}) + \mathbb{P}(Z \in [u, u+\eta])$$

Thus

$$\Delta_{n,p}(u) \leq C\left(\frac{1}{\eta^3\sqrt{np(1-p)}} + \eta\right),$$

for some constant non depending on n, η, ϵ and p.
The choice $\eta = (np(1-p))^{-1/8}$ allows to conclude. ∎

Exercise 5. *For each $\eta > 0$, $u \in \mathbb{R}$ there exists function $f_{u,\eta} \in \mathcal{C}_b^3$ with*

$$1\!\!1_{[u+\eta,\infty[} \leq 1\!\!1_{[u,\infty[} \leq f_{u,\eta} \quad \text{and} \quad \|f'''_{u,\eta}\|_\infty = \mathcal{O}\left(\eta^{-3}\right).$$

1. *Set first $u = 0, \eta = 1$. Then we set $g(x) = 0$ if $x \notin]0,1[$ and:*

   (a)
   $$g(x) = x^4(1-x)^4, \qquad x \in]0,1[.$$
   Then $g \in \mathcal{C}_b^3$.

   (b)
   $$g(x) = \exp\left(-\frac{1}{x(1-x)}\right), \qquad x \in]0,1[.$$
   Then $g \in \mathcal{C}_b^\infty$.
   Indeed each of g's derivative writes as $g^{(k)}(x) = F(x)g(x)$ for some rational function F with no pole excepted for $0, 1$. In this case the function is $\mathcal{C}_b^\infty$.

   A convenient function is defined as $f(x) = G(x)/G(0)$ where we set
   $$G(x) = \int_x^1 g(s)ds, \quad \text{for } 0 \leq x < 1,$$
   and $f(x) = 0$ for $x \geq 1$ with g as above.

2. *General case. With f as before set $f_{u,\eta}(x) = f(u + x/\eta)$:*
   $$f_{u,\eta}^{(k)}(x) = \frac{1}{\eta^k}\left(u + \frac{x}{\eta}\right) \leq \frac{\|f^{(k)}\|_\infty}{\eta^k}, \quad \text{for } k = 0, 1, 2 \text{ or } 3.$$
   For the second function k may be chosen arbitrarily large. This allows to conclude. ∎

2.2 Quantitative results

2.2.1 Moment inequalities

We now derive two important moment inequalities respectively called Marcinkiewicz-Zygmund and Rosenthal moment inequalities.

Later in those notes alternative proofs of those results will be obtained.

Lemma 2.2.1. *Let X_n be a sequence of independent centered random variables with finite moment of order $2p$ for some $p \in \mathbb{N}^*$, then there exists a constant $C > 0$ which only depends on p such that*

- *Marcinkiewicz-Zygmund inequality holds:*

$$\mathbb{E}(X_1 + \cdots + X_n)^{2p} \leq Cn^p EX^{2p}.$$

- *Rosenthal inequality for $p = 2$:*

$$\mathbb{E}(X_1 + \cdots + X_n)^4 \leq C((n\mathbb{E}X^2)^2 + n\mathbb{E}X^4).$$

Remark 2.2.1. *The second inequality also extends to all $p \geq 2$. There exists a constant C only depending on p such that*

$$\mathbb{E}|X_1 + \cdots + X_n|^p \leq C((n\mathbb{E}X^2)^{\frac{p}{2}} + n\mathbb{E}|X|^p).$$

Proof. Simple combinatoric arguments yield:

$$\begin{aligned}
\mathbb{E}(X_1 + \cdots + X_n)^{2p} &= \sum_{i_1,..,i_{2p}=1}^{n} \mathbb{E}X_{i_1} \cdots X_{i_{2p}} \\
&= \sum_{i_1,..,i_{2p}=1}^{n} T(i_1, \ldots, i_{2p}) \\
&\leq \sum_{i_1,..,i_{2p}=1}^{n} |T(i_1, \ldots, i_{2p})| \\
&\leq (2p)! \sum_{1 \leq i_1 \leq \cdots \leq i_{2p} \leq n} |T(i_1, \ldots, i_{2p})|.
\end{aligned}$$

Now from centering conditions we see that terms T vanish except for cases when $i_1 = i_2, \ldots, i_{2p-1} = i_{2p}$, since else an index i would be isolated and the corresponding term vanishes by using independence. Among $A = \{i_2, i_4, \ldots, i_{2p}\}$ which take precisely n^p values one needs to make summations according to $\mathrm{Card}(A)$.
If all those indices are equal $T = \mathbb{E}X_0^{2p}$ and there are n such terms,

and if they are all different, it is $(\mathbb{E}X_0^2)^p$.
For $p = 2$ we thus get the second point in this Lemma.
For any $p \geq 1$, just use Hölder inequality to derive the first result. ∎

Exercise 6 (Weierstrass theorem). *Weierstraß theorem states that a continuous function over the interval is the uniform limit of some sequence of polynomials.*
Let $g : [0, 1] \to \mathbb{R}$ be a continuous function we recall that

$$w(t) = \sup_{|x-y|<t} |g(x) - g(y)|$$

satisfies $\lim_{t\downarrow 0} w(t) = 0$ since Heine theorem (recalled below) entails that the function g is uniformly continuous.
Let $X_{1,x}, X_{2,x}, \ldots$ be iid $b(x)-$random variables (Bernoulli distributed with the parameter x), we denote

$$S_{n,x} = \frac{1}{n}(X_{1,x} + \cdots + X_{n,x}).$$

Set $g_n(x) = \mathbb{E}g(S_{n,x})$:

1. *Prove that g_n is a polynomial with degree n with respect to the variable p.*

2. *Prove the bound of $Var\, g(S_{n,x}) = \frac{1}{n} Var\, X_{1,x} \leq \frac{1}{4n}$.*

3. *Apply Markov inequality to derive to prove that*

$$\lim_{n\to\infty} \sup_{0\leq x\leq 1} |g_n(x) - g(x)| = 0.$$

4. *If g is a Hölder function, hence if there exist constants $c, \gamma > 0$ with $|g(x) - g(y)| \leq c|x - y|^\gamma$ for each $x, y \in [0, 1]$, precise convergence rates in the Weierstraß approximation theorem.*

5. *Now use Lemma 2.2.1 for moment inequalities with even order $2m$, then $\mathbb{E}(S_{n,x} - g_n(x))^{2m} \leq cn^{-m}$ for a constant which does not depend on $x \in [0, 1]$.*

6. *Use the previous hight order moment inequality to derive alternative convergence rates in Weierstraß theorem.*

Hints.

1.
$$g_n(x) = \sum_{k=0}^{n} \binom{k}{n} x^k (1-x)^{n-k} g\left(\frac{k}{n}\right).$$

2. Prove that $x(1-x) \leq \frac{1}{4}$ if $0 \leq x \leq 1$.

3. Set $t > 0$ arbitrary, an $A_{n,p} = (|S_{n,p} - p| > t)$ then:
$$\begin{aligned} g_n(x) - g(x) &= \mathbb{E}(g(S_{n,x}) - g(x)) \\ &= \mathbb{E}(g(S_{n,x}) - g(x))1_{A_{n,x}} \\ &\quad + \mathbb{E}(g(S_{n,x}) - g(p))1_{A^c_{n,x}}. \end{aligned}$$
From Markov inequality and the second point
$$\mathbb{P}(A_{n,x}) \leq \frac{1}{4nt^2},$$
thus a bound of the first term in the previous inequality is
$$\frac{\|g\|_\infty}{2nt^2},$$
and from definitions the second term is bounded above by $w(t)$. Let n tend to infinity first to conclude.

4. Here $w(t) \leq ct^\gamma$ and the previous inequality writes
$$\|g_n - g\|_\infty \leq \frac{\|g\|_\infty}{2nt^2} + ct^\gamma.$$
Setting $t^{2+\gamma} = \frac{\|g\|_\infty}{2cn}$ provides a rate $n^{-\frac{\gamma}{2+\gamma}}$.

5. From Lemma 2.2.1 $\mathbb{E}(S_{n,x} - g_n(x))^{2m} \leq c\mathbb{E}X_{1,x}^{2m}\, n^{-m}$.

6. Now
$$\|g_n - g\|_\infty \leq \frac{c\|g\|_\infty}{2nt^{2m}} + ct^\gamma$$
set $t^{2m+\gamma} = \frac{C}{n^m}$ then a rate is $n^{-\frac{m\gamma}{2m+\gamma}}$ is now provided. ∎

Recall that continuity at point $x_0 \in [0,1]$ and uniform continuity of $g : [0,1] \to \mathbb{R}$ write respectively

$$\begin{array}{rcl} \forall\epsilon > 0, \exists\alpha > 0,\ \forall x \in [0,1] :\ |x - x_0| < \eta & \Rightarrow & |g(x) - g(x_0)| < \epsilon \\ \forall\epsilon > 0, \exists\alpha > 0, \forall x, y \in [0,1] :\ |x - y| < \eta & \Rightarrow & |g(x) - g(y)| < \epsilon. \end{array}$$

In the latter case η does thus not depend on x_0.

Exercise 7. *The function $x \mapsto g(x) = x^2$ is not uniformly continuous over $\mathbb{R}$.*

Hint. Reasoning by absurd. Set $x = n$ and $y = n + \frac{1}{2n}$, then

$$g(y) - g(x) = 1 + \frac{1}{4n^2} \quad \text{does not tend to zero as} \quad n \uparrow \infty. \blacksquare$$

This fundamental result (see eg. [Doukhan and Sifre, 2001]) writes:

Theorem 2.2.1 (Heine). *Let $g : K \to \mathbb{R}$ be a continuous function defined on a compact metric space (K, d) then g is uniformly continuous.*

2.2.2 Exponential inequalities

Below we develop two exponential inequalities which yelds reasonable bounds for the tail of partial sums of independent identically distributed random variables. From the Central Limit Theorem we first check the Gaussian case.

Exercise 8. *Let $N \sim \mathcal{N}(0,1)$ be a standard normal random variable, then use integrations by part and Markov inequality to derive:*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \le \overline{\Phi}(t) = \mathbb{P}(\mathcal{N}(0,1) > t) \le \frac{1}{t}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}.$$

Analogously we obtain:

Lemma 2.2.2 (Hoeffding). *Let $R_1, \ldots, R_n$ be independent Rademacher random variables (i.e. $\mathbb{P}(R_i = \pm 1) = \frac{1}{2}$).*
For real numbers $a_1, \ldots, a_n$ set

$$\xi = \sum_{i=1}^n a_i R_i, \qquad \textit{and we assume that} \qquad \sum_{i=1}^n a_i^2 \le c.$$

Then:

1. $\mathbb{P}(\xi \geq x) \leq e^{-\frac{x^2}{2c}}$, *for all* $x \geq 0$,

2. $\mathbb{P}(|\xi| \geq x) \leq 2e^{-\frac{x^2}{2c}}$, *for all* $x \geq 0$, *and*

3. $\mathbb{E}e^{\frac{\xi^2}{4c}} \leq 2$.

Proof. If $s \in \mathbb{R}$ first prove that

$$\mathbb{E}e^{sR_1} \leq e^{s^2/2}. \tag{2.1}$$

The inequality (2.1) is rewritten

$$\mathrm{ch}\ s \equiv \frac{1}{2}\left(e^s + e^{-s}\right) \leq e^{s^2/2}.$$

Indeed the two previous functions may be expanded as analytic functions on the real line, $\mathbb{R}$, and:

$$\mathrm{ch}\ s = \sum_{k=0}^{\infty} \frac{s^{2k}}{(2k)!}, \qquad e^{s^2/2} = \sum_{k=0}^{\infty} \frac{s^{2k}}{2^k \cdot k!}.$$

Inequality (2.1) thus follows from the relation $(2k)! \geq 2^k \cdot k!$ simply restated as

$$(k+1)(k+2)\cdots(k+k) \geq (2\cdot 1)(2\cdot 1)\cdots(2\cdot 1) = 2^k.$$

Markov inequality now implies

$$\mathbb{P}(\xi \geq x) \leq e^{-tx}\mathbb{E}e^{t\xi}, \qquad \forall t \geq 0,$$

because eqn. (2.1) entails

$$\mathbb{E}e^{t\xi} = \prod_{i=1}^{n} \mathbb{E}e^{ta_i R_i} \leq e^{t^2 c/2}.$$

For $t = x/c$ we derive the point *1)*.
Point *2)* come from the observation that ξ is a symmetric random

variable and thus $\mathbb{P}(|\xi| \geq x) \leq 2\mathbb{P}(\xi \geq x)$.
Point *3)* is derived after the forthcoming calculations:

$$
\begin{aligned}
\mathbb{E}e^{\frac{\xi^2}{4c}} - 1 &= 4c\mathbb{E}\int_0^{\xi^2} \exp\left(\frac{t}{4c}\right) dt \\
&= 4c\mathbb{E}\int_0^{\infty} \mathbb{1}_{\{t \leq \xi^2\}} \exp\left(\frac{t}{4c}\right) dt \\
&= 4c\int_0^{\infty} \mathbb{E}\,\mathbb{1}_{\{t \leq \xi^2\}} e^{\frac{t}{4c}}\, dt \\
&= 4c\int_0^{\infty} \mathbb{P}(\xi^2 \geq t) e^{\frac{t}{4c}}\, dt \\
&\leq 4c\int_0^{\infty} e^{-\frac{t}{4c}}\, dt = 1
\end{aligned}
$$

Here Fubini-Tonnelli justifies the first inequalities while the last inequality is consequence of relation *2)*. ∎

Remark 2.2.2. *Let $R \in [-1, 1]$ be a centered random variable then*

$$
\mathbb{E}e^{tR} \leq \frac{1}{2}\left(e^t + e^{-t}\right),
$$

thus Hoeffding instantaneously extends to sums $\sum_i a_i R_i$ for R_i with values in $[-1, 1]$, centered independent random variables.

Lemma 2.2.3 (Bennett)**.** *Let $Y_1, \ldots, Y_n$ be independent centered random variables with $|Y_i| \leq M$ for $1 \leq i \leq n$ denote*

$$
V = \sum_{i=1}^{n} \mathbb{E}Y_i^2.
$$

If $\xi = \sum_{i=1}^{n} Y_i$ then for each $x \geq 0$ the Bennett inequality holds:

$$
\mathbb{P}(|\xi| \geq x) \leq 2e^{-\frac{x^2}{2V}B\left(\frac{Mx}{V}\right)}, \text{ with } B(t) = \frac{2}{t^2}\left((1+t)\log(1+t) - t\right).
$$

Bernstein inequality also holds:

$$\mathbb{P}(|\xi| \geq x) \leq 2\exp\left(-\frac{x^2}{2\left(V + \frac{1}{3}Mx\right)}\right).$$

Proof. The proof is again based upon Markov inequality.
We shall make use of the independence of $Y_1, \ldots, Y_n$.
We first need to bound above Laplace transform of Y_i:

$$\begin{aligned}
\mathbb{E}e^{tY_i} &= \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}Y_i^k \\
&\leq 1 + \mathbb{E}Y_i^2 \sum_{k=2}^{\infty} \frac{t^k}{k!}\mathbb{E}Y_i^k \\
&= 1 + \mathbb{E}Y_i^2 g(t), \\
&\quad \text{with} \;\; g(t) = \frac{e^{tM} - 1 - tM}{M^2} \\
&\leq e^{\mathbb{E}Y_i^2 g(t)}
\end{aligned}$$

The first inequality follows from $\mathbb{E}Y_i = 0$ and $|\mathbb{E}Y_i^k| \leq M^{k-2}\mathbb{E}Y_i^2$ for each $k > 1$.
Both from independence and from Markov inequality we then obtain:

$$\mathbb{P}(\xi \geq x) \leq e^{Vg(t)-xt}.$$

Optimize this bound with respect to V yields $Vg'(t) = x$ hence

$$t = \frac{1}{M}\log\left(1 + \frac{xM}{V}\right) > 0,$$

and $Vg(t) - xt = x/M - t(V/M + x)$ yields Bennett inequality.
Bernstein inequality follows from the relation

$$(1+t)\log(1+t) - t \geq \frac{t^2}{2(1+t/3)}.$$

rewritten $(1 + t/3)B(t) \geq 1$. To prove it quote that the function

$$t \mapsto f(t) = t^2((1 + t/3)B(t) - 1)$$

satisfies moreover

$$f'(0) = 0, \text{ and } f''(t) = \frac{1}{3}((1+t)\log(1+t) - t) \geq 0,$$

entails $f(0) = 0$ and $f'(t) \geq 0$, ∎

Remark 2.2.3. *Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be an a.s. derivable non-decreasing function, then Fubini-Tonnelli entails*

$$\mathbb{E}g(|\xi|) = \int g(z)\mathbb{P}_{|\xi|}(dz) = \int g'(z)\mathbb{P}(|\xi| > z)\, dz.$$

Set $A = 3V/M$ then from Bernstein inequality in Lemma 2.2.3 we get

$$\mathbb{E}g(|\xi|) \leq g(3V/M) + 2\int_{3V/M}^{\infty} g'(z)e^{-3z/4M}\, dz,$$

and

$$\mathbb{E}g(|\xi|) \leq g(3V/M) + 8M/3\int_{9V/4M^2}^{\infty} g'(4Mx/3)e^{-x}\, dx,$$

with $x = 3z/4M$.
Hence if $g(x) = |x|^p$ for some $p > 0$,

$$\mathbb{E}g(|\xi|) \leq (3V/M)^p + 2p(4M/3)^p\int_{9V/4M^2}^{\infty} x^{p-1}e^{-x}\, dx.$$

This is a more general form of the Rosenthal inequality in Lemma 2.2.1.

Chapter 3

Estimation concepts

Many statistical procedures are restatements of probabilistic inequalities and results but in several occurrences such procedures need much more as this will be initiated in this section for the independent case. We begin the section with applications of the previous moment inequalities in Lemma 2.2.1 useful for empirical procedures and then describe some empirical estimates, contrast estimates and non parametric estimates.
The developments are not given with mention to their specific interest but rather with respect to the dependent development provided in those notes under dependence.

3.1 Empirical estimates

The behavior of empirical means are deduced from the behavior of partial sums, and below we shall restate such results in a statistical setting.

Corollary 3.1.1. *Let $(X_n)_{n\geq 0}$ be and independent identically distributed sequence. If $\mathbb{E}X_0^4 < \infty$ then,*

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \to_{n\to\infty} \mathbb{E}X_0, \qquad a.s.$$

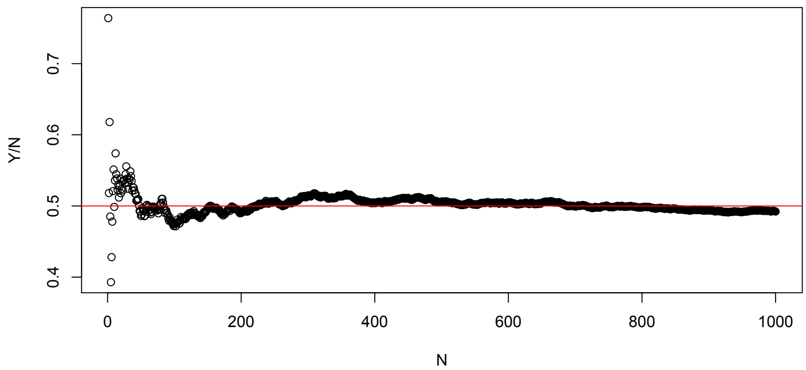


Figure 3.1: Convergence in a law of large numbers.

Remark 3.1.1. *Quote the the ergodic Theorem 9.1.1 proves that the simple assumption $\mathbb{E}|X_0| < \infty$ ensures indeed this SLLN. We set this result as a simple consequence of the previous Marcinkiewicz-Zygmund inequality in Lemma 2.2.1 for clarity of the exposition. Convergence in this LLN is simulated in the Figure 3.1*

Proof. Let $\epsilon > 0$ be arbitrary then Markov inequality entails

$$\mathbb{P}(|\overline{X}| \geq \epsilon) \leq C\frac{\mathbb{E}X_0^4}{\epsilon^4 n^2}.$$

Thus

$$\sum_n \mathbb{P}(|\overline{X}_n| \geq \epsilon),$$

is a convergent series. Hence the *a.s.* convergence is a consequence of Borel-Cantelli Lemma. ∎

Now in case $\mathbb{E}X_0^2 < \infty$ Markov inequality yields $\mathbb{L}^2-$convergence of $\overline{X}$; indeed $\mathrm{Var}\,(\overline{X}) = \mathrm{Var}\,(X_0)/n$, thus convergence in probability also holds.

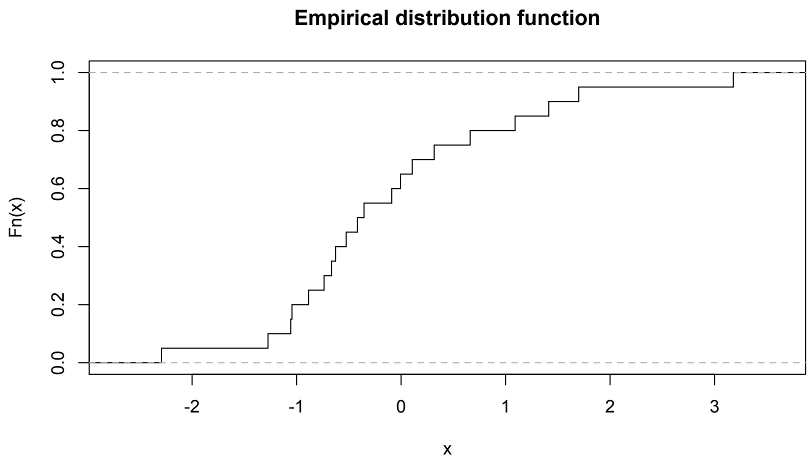


Figure 3.2: A cumulative distribution function.

This also allows to prove a first statistical fundamental result:

Theorem 3.1.1. *Let (Y_n) be a real valued and independent identically distributed sequence such that Y_0 admits cumulative distribution function $F(y) = \mathbb{P}(Y_0 \le y)$ on $\mathbb{R}$.*
Define the empirical cumulative distribution

$$F_n(y) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{Y_j \le y\}}.$$

Then $\mathbb{E}F_n = F$ (the estimator is said to be unbiased) and

$$\sup_{y \in R} |F_n(y) - F(y)| \to_{n\to\infty} 0, \quad a.s.$$

Remark 3.1.2. *This is interesting to check graphically this convergence as reported by Figure 3.2.*

Proof. The previous SLLN (in Corollary 3.1.1) implies the convergence.

Uniform convergence is implied by the variant of Dini theorem in Exercise 9. ∎

Exercise 9 (variant of Dini theorem). *Assume that a sequence of cdf satisfies* $\lim_{n\to\infty} F_n(x) = F(x)$ *for each* $x \in \mathbb{R}$*. If* F *is a continuous cdf then the convergence is uniform.*

Proof. Let $\epsilon > 0$ be arbitrary from the properties of a cdf there exists $A > 0$ such that if $x > A$ then $1 - F(x) < \epsilon/2$ and $x < -A$ implies $F(x) < \epsilon/3$.
From Heine Theorem 2.2.1, there exist $x_1 = -A < x_2 < \cdots < x_p = A$ such that if $x \in [x_i, x_{i+1}]$ then $F(x_{i+1}) - F(x) < \epsilon/3$ and $F(x) - F(x_i) < \epsilon/3$, if $i = 1, \ldots, p-1$. Set $x_0 = -\infty$ and $x_{p+1} = \infty$, thus the oscillation of F is less that $\epsilon/3$ over each interval $J_i = ([x_i, x_{i+1})$ for each $i = 0, \ldots, p$ (limits are included for each finite extremity).
From the relation $\lim_{n\to\infty} F_n(x_i) = F(x_i)$ for $i = 1, \ldots, p$ this is possible to exhibit N such that if $n > N$ then $|F_n(x_i) - F(x_i)| < \epsilon/3$.
Each $x \in \mathbb{R}$ belongs to some interval J_i so that in case $i \neq 0$:

$$|F_n(x) - F(x)| \le |F_n(x) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_i) - F(x)| < \epsilon$$

For $i = 0$ one should replace $x_0 = -\infty$ by $x_1 = -A$ in the above inequality to conclude. ∎

3.2 Contrasts

Assume that an independent identically distributed sample with values in a Banach space E and admits a marginal distribution in a class $(P_\theta)_{\theta\in\Theta}$.

Definition 3.2.1. *A function* $\rho : E \times \Theta \to \mathbb{R}$ *is a contrast if the expression* $\theta \mapsto D(\theta_0, \theta) = \mathbb{E}_{\theta_0}\rho(X, \theta)$ *is well defined and admits a unique minimum* θ_0*.*

If $X \sim P_{\theta_0}$ then $\rho(X, \theta)$ is an unbiased for the function $g(\theta_0) = D(\theta_0, \theta)$ (for each $\theta \in \Theta$). In case we have only one realization X of this experiment the true parameter θ_0 is estimated by a minimizer $\widehat{\theta}(X)$ of the contrast $\theta \mapsto \rho(X, \theta)$.

$$\widehat{\theta}(X) = \mathrm{Argmin}_{\theta \in \Theta}\ \rho(X, \theta) \tag{3.1}$$

If $\Theta \subset \mathbb{R}^d$ is open and such that the function $\theta \mapsto \rho(X, \theta)$ be differentiable the estimate $\widehat{\theta}(X)$ of the parameter θ_0 satisfies

$$\nabla \rho(X, \widehat{\theta}(X)) = 0 \tag{3.2}$$

(usually this is easier to check than (3.1)).

Example 3.2.1. *This situation occurs eg. if:*

- *Maximum Likelihood Estimator (MLE)* $\rho(x, \theta) = -\log f_\theta(X)$ *with* f_θ *the density of* P_θ. *If* $X = (X_1, \ldots, X_n)$ *for an independent identically distributed sample with marginal densities* $p_\theta(x)$ *then*
$$\rho(x, \theta) = -\sum_{k=1}^n \log f_\theta(X_k).$$
The contrast assumption relies on identifiability:
$$f_{\theta_1} = f_{\theta_2}\ (a.s.)\ \Rightarrow \theta_1 = \theta_2.$$
- *Least Squares (LSE)* $X = G(\theta) + \sigma(\theta)\xi$ *and* $\rho(x, \theta) = \|X - G(\theta)\|^2/\sigma^2(\theta)$. *If* $\xi = (\xi_1, \ldots, \xi_n)$ *are independent identically distributed random variables and* $G(\theta) = (g(\theta, z_1), \ldots, g(\theta, z_n))$ *then this is a regression model with fixed design.*

Remark 3.2.1 (Model selection)**.** *A huge part of nowadays statistics extends on such contrast techniques in case the statistical model itself is unknown but in a class of models* $\mathcal{M}$, *precisely each of those models* $M \in \mathcal{M}$ *is indexed by a parameter set* Θ_M *and one knows a contrast* $(\rho_M(X, \theta))_{\theta \in \Theta_M}$ *(this is model selection).*
However the price to pay for using the model M *is a penalization* $p(M)$ *which increases with the complexity of the model, then one may estimate the model* M *and the parameter* $\theta \in \Theta_M$ *as:*

$$\mathit{Argmin}\left\{ p(M) + \inf_{\theta \in \Theta_M} \rho_M(X, \theta), \quad M \in \mathcal{M} \right\}.$$

We choose in this presentation to avoid a precise presentation of those techniques essentially introduced by Pascal Massart, see eg. in [Massart, 2007].
Indeed very tough concentration inequalities are needed in this fascinating setting. Also the dependent case is always in our mind: no completely satisfactory extension exists yet.

3.3 Functional estimation

We now introduce another standard tool of statistics related to function estimation; a great presentation is that in the lecture notes [Rosenblatt, 1991].
Let (Y_j) be an independent identically distributed sequence with a common marginal density f.
In order to fit $f = F'$ the simple plug-in technique consists to derive a an estimate of the cumulative repartition. This does not work since derivation is not a continuous function in the space $D[0,1]$, moreover F_n's derivative is 0 (a.s.).
A reasonable estimate is the histogram; divide the space of values into pieces with a small probability then we may count the proportion of occurrences of Y_j's in an interval to fit f by a step function.
Formally this means that

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{j,m}(x)e_{j,m}(X_i), \quad e_{j,m}(x) = \frac{\mathbb{1}_{\{x\in I_{j,m}\}}}{\sqrt{\mathbb{P}(X_0 \in I_{j,m})}}$$

for a partition $I_{1,m}\bigcup\cdots\bigcup I_{m,m} = \mathbb{R}$.

Remark 3.3.1. *A problem is that histograms are not smooth even if they are aimed at estimating possibly smooth densities.*
Thus, more generally $(e_{j,m})_{1\le j\le m}$ may be chosen as an orthonormal system of $\mathbb{L}^2(\mathbb{R})$, such as a wavelet basis; in [Doukhan, 1988] we initially introduced a simple wavelet type estimates.
Any orthonormal system $e_{j,m} = e_j$ for $1 \le j \le m$ may also be considered. Note also that

$$\widehat{f}(x) = \sum_{j=1}^{m} \widehat{c}_j e_{j,m}(x)$$

where

$$\widehat{c}_j = \frac{1}{n}\sum_{i=1}^{n} e_{j,m}(X_i)$$

empirical unbiased estimate of $\mathbb{E}e_{j,m}(X_0)$ *(which means that the mean of those estimates is the coefficient to be fitted).*
Such estimators are empirical estimators of the orthogonal projection f_m *of* f *of the vector space spanned by* $(e_{j,m})_{1\le j\le m}$*; they are know as projection density estimations of* f*. In order to make them consistent one needs to choose a sequence of parameter* $m \equiv m_n \uparrow \infty$*. Such general classes of estimates are thus reasonable and may be proved to work.*

Anyway we propose to develop an alternative smoothing technique since in this case an asymptotic expansion of the bias may be exhibited (quote that the wavelet type estimator corrects this real problem of projection estimators) .
Let g_h be an approximation of the Dirac measure as $h \downarrow 0$, we derive $F_n \star g_h(x)$ to get the forthcoming kernel type estimates of the density:

Definition 3.3.1. *Let* (Y_n) *be a real valued and independent identically distributed sequence such that* Y_0 *admits a density* f *on* $\mathbb{R}$*. If* $K : \mathbb{R} \to \mathbb{R}$ *denotes a function such that:*

$$\int_{\mathbb{R}} (1+|K(y)|)|K(y)|dy < \infty, \qquad \int_{\mathbb{R}} K(y)dy = 1,$$

a kernel estimator of f *is defined through a sequence* $h = h_n \to_{n\to\infty} 0$ *by:*

$$\widehat{f}(y) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{Y_j - y}{h}\right).$$

Figure 3.3 reports competitive behaviors for an histogram and for a kernel density estimate ($n = 32, h = 2.477$). A first result allows to bound the bias of such estimators:

Lemma 3.3.1. *Let* g *denote a bounded density for some probability distribution with moments up to order* $p \in \mathbb{N}^*$*, then there exists a polynomial* P *with degree* $\le p$ *such that* $K = Pg$ *is a kernel satisfying*

$$\int_{\mathbb{R}} y^s K(y)dy = \begin{cases} 1, & \text{if } \ j = 0,\ p, \\ 0, & \text{if } \ 1 \le j < p. \end{cases}$$

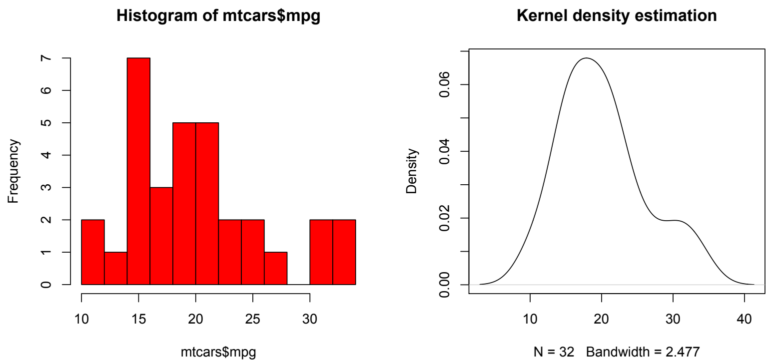


Figure 3.3: Histogram versus a kernel density estimate.

Remark 3.3.2. *Such functions are called $p-$th order kernels.*
If $p > 2$ such kernel are not non-negative.
For $p = 1$ and g symmetric ($g(-y) = g(y)$) this is simple to see that $P = 1$ satisfies the previous relations but maybe not $\int y^2 g(y)dy = 1$, anyway this expression is positive.

Proof. It is simple to use the fact that the quadratic form associated to the square matrix $A = (a_{i+j})_{0 \le i,j \le n}$ with $a_k = \int_{\mathbb{R}} y^k g(y)dy$ with order $(p + 1)$ is symmetric positive definite. Indeed if $x = (x_0, \ldots, x_p)' \in \mathbb{R}^{p+1}$

$$x'Ax = \int_{\mathbb{R}} \Big(\sum_{j=0}^{p} x_j y^j \Big)^2 g(y)dy \ge 0.$$

If the previous expression vanishes the fact that $g \ne 0$ on a set with positive measure implies that this set infinite and thus that the polynomial

$$y \mapsto \sum_{j=0}^{p} x_j y^j,$$

vanishes on an infinite set.
Thus it must have null coefficients. ∎

The change of variable $u = (v - y)/h$ together with a simple application of Taylor formula, left as an exercise, proves Proposition 3.3.1:

$$\mathbb{E}\widehat{f}(y) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{v-y}{h}\right) f(v)dv = \int_{\mathbb{R}} K(u)f(y-hu)\,du.$$

Proposition 3.3.1. *Assume that $h \to_{n\to\infty} 0$.*
Assume that the function f admits p continuous and bounded derivatives, then if K is a $p-$th order kernel:

$$\sup_y \left|\mathbb{E}\widehat{f}(y) - f(y)\right| \le \frac{h^p}{p!}\|f^{(p)}\|_\infty \int |u|^p|K(u)|du.$$

Remark 3.3.3. *Some precisions and improvements are needed here.*

- *Independence of (Y_k) is not necessary here but only the fact that Y_k's are identical for $1 \le k \le n$.*
- *The uniformity over $\mathbb{R}$ may be omitted if K admits a compact support, then*

$$\left|\mathbb{E}\widehat{f}(y) - f(y)\right| \le \frac{h^p}{p!}\sup_{u\in y+V}|f^{(p)}(u)|\int |u|^p|K(u)|du,$$

  for a neighborhood V of 0 such that $hV \subset$ Support(K).
- *In fact using analytic functions ideas this is even possible to describe infinite-order kernels as in [Politis, 2003].*

Use the previous results together with Lindeberg theorem with $k = n$ and

$$U_j = \frac{1}{\sqrt{nh}}\left(K\left(\frac{Y_j - y}{h}\right) - \mathbb{E}K\left(\frac{Y_j - y}{h}\right)\right).$$

Theorem 3.3.1. *Let now $h = h_n \downarrow 0$ as $n \to \infty$. Assume that $nh_n \to_{n\to\infty} \infty$:*

$$\begin{aligned} nh\,Var\,\widehat{f}(y) \quad &\to_{n\to\infty} \quad f(y)\int_{\mathbb{R}} K^2(u)du, \\ \sqrt{nh}(\widehat{f}(y) - \mathbb{E}\widehat{f}(y)) \quad &\to^{\mathcal{L}}_{n\to\infty} \quad \mathcal{N}\left(0, f(y)\int_{\mathbb{R}} K^2(u)du\right). \end{aligned}$$

Hence if conditions of the previous Proposition hold and if we assume also that $h \to_{n\to\infty} 0$ then:

$$\mathbb{E}(\widehat{f}(y) - f(y))^2 \quad \sim_{n\to\infty} \quad \frac{1}{nh} f(y) \int_{\mathbb{R}} K^2(u)du + \left(\frac{h^p}{p!} f^{(p)}(y) \int |u|^p K(u)du \right)^2.$$

Thus convergence in probability holds for such estimators and a CLT is also available.
The usual minimax rates of such estimates is $O(n^{\frac{p}{2p+1}})$ is obtained by minimizing this expression wrt $h = h_n$ or by equating the squared bias and variance of the estimator.
Now if $nh_n \to \infty$ using Rosenthal moment inequalities from Remark 2.2.1 implies

$$\mathbb{E}|\widehat{f}(y) - \mathbb{E}\widehat{f}(y)|^p \leq \frac{C}{(nh)^{\frac{p}{2}}}.$$

This with Markov inequality and Borel-Cantelli implies the result

Proposition 3.3.2. *For f continuous around the point y, $\widehat{f}(y) \to f(y)$ a.s. if*

$$\sum_n \frac{1}{(nh_n)^{\frac{p}{2}}} < \infty.$$

Remark 3.3.4. *In fact better results can be proved by using the Bernstein exponential inequality but the present section was only introductory in order to provide some statistical applications to be developed later under dependence.*

Exercise 10. *Prove the bound:*

$$\mathbb{P}(\sqrt{nh}|f_{n,h}(x) - \mathbb{E}f_{n,h}(x)| \geq t) \leq 2e^{-t^2/(cf(x))}, \qquad (3.3)$$

for all $c \leq 2\int_{\mathbb{R}} K^2(u)du$ for n large enough.
Then from integration derive that for each $p \geq 1$:

$$\|\widehat{f}(x) - \mathbb{E}\widehat{f}(x)\|_p = \mathcal{O}\left(\frac{1}{\sqrt{nh}}\right), \qquad \text{if } nh \geq 1.$$

Proofs. The results rely on simple integration tricks. We will use Bernstein inequality in Lemma 2.2.3.
Write
$$\sqrt{nh}(f_{n,h}(x) - \mathbb{E}f_{n,h}(x)) = Z_1 + \cdots + Z_n$$
with
$$Z_j = U_j - \mathbb{E}U_j, \qquad U_j = K\left((Y_j - y)/h\right)/\sqrt{nh}.$$
Then the relations
$$\|Z_j\|_\infty \le 2\|K\|_\infty/\sqrt{nh},$$
and
$$\mathbb{E}Z_j^2 \sim \sqrt{\frac{h}{n}}f(y)\int K^2(s)\,ds,$$
completing the proof of the first inequality.
The moment inequality relies on the fact that setting $u = t/\sqrt{f(x)}$ yields:
$$2\int t^p e^{-t^2/(cf(x))}dt = 2f(x)^{\frac{p+1}{2}}\int u^p e^{-u^2/c}du < \infty.$$
This allows to conclude. ∎

Remark 3.3.5 (Uniform convergence)**.** *A.s. uniform convergence over a compact interval I may also be derived under uniform continuity.*
Typically:
Divide $I = [0,1]$ into m i, thenntervals $I_1, \ldots, I_m$ with measure $1/m$.
Then if the chosen kernel is Lispchitz the oscillation of $\widehat{f}$ over each such interval is less that C/mh^2 for some suitable constant.
Thus if $mh^2 > C'$ is large enough for a constant $C' \equiv C'(\epsilon, C, f)$, the oscillation of the function $\widehat{f} - f$ over each interval I_k will be less that some fixed valued ϵ.
Choose now $x_k \in I_k$ for each $1 \le k \le m$.
Then for each $\epsilon > 0$:
$$\mathbb{P}\left(\sup_{x\in[0,1]}|\widehat{f}(x) - f(x)| > 2\epsilon\right) \le m\max_{1\le k\le m}\mathbb{P}(|\widehat{f}(x_k) - f(x_k)| > \epsilon).$$

Calibrating more precisely h and m yield a.s. uniform results for the convergence of $\widehat{f}$ extending Proposition 3.3.2. Quote that this reinforcement of a.s. convergence is not related to independence, see Exercise 28.

Example 3.3.1. *Other functions of interest may be fitted through kernel estimates.*

- *A very useful example is that of non-parametric regression. The natural estimate of a mean is the empirical mean, but think now of an independent sequence*

$$Y_k = r\left(\frac{k}{n}\right) + \xi_k, \qquad k = 1, \dots, n \tag{3.4}$$

  for some independent identically distributed sequence (ξ_k) and a smooth regression function r.
  A natural estimate would be a local mean

$$\widehat{r}(x) = \frac{\sum_{k=1}^{n} 1\!\!1_{\{h|x-k/n|<1\}} Y_k}{\sum_{k=1}^{n} 1\!\!1_{\{h|x-k/n|<1\}}}.$$

  This is easily generalized as a kernel regression estimate in the previous fixed regression design:

$$\widehat{r}(x) = \frac{1}{nh} \sum_{k=1}^{n} Y_k K\left(\frac{x - k/n}{h}\right).$$

- *Random regression designs write $Y_k = r(X_k) + \xi_k$, for $k = 1, \dots, n$, where (X_k) is an independent identically distributed sequence.*
  Here the Nadaraya-Watson estimate writes:

$$\widehat{r}(x) = \begin{cases} \frac{\widehat{g}(x)}{\widehat{f}(x)}, & \text{if } \widehat{f}(x) \neq 0 \\ 0, & \text{if } \widehat{f}(x) = 0 \end{cases} \tag{3.5}$$

  with

$$\widehat{g}(y) = \frac{1}{nh} \sum_{j=1}^{n} X_j K\left(\frac{Y_j - y}{h}\right).$$

  The functions f and g estimated are respectively the derivative and $g = rf$.

- *Derivative also can also be estimated. For example to estimate f' one just may derivate f's estimate if K is a smooth function. This makes a change in rates since eg.*

$$\widehat{f'}(y) = \frac{-1}{nh^2} \sum_{j=1}^{n} K'\left(\frac{Y_j - y}{h}\right).$$

One may indeed check that each term in the sum above admits a variance equivalent to its second moment and the usual change in variable $u = y + th$ yields:

$$\begin{aligned} \mathbb{E}\left(K'\left(\frac{Y_j - y}{h}\right)\right)^2 &= h\int K'^2(t)f(y+th)\,dt \\ &\sim_{h\to 0} hf(y)\int K'^2(t)\,dt. \end{aligned}$$

We derive analogously

$$\mathbb{E}K'\left(\frac{Y_j - y}{h}\right) = \mathcal{O}(h)$$

thus from independence we obtain

$$Var\,\widehat{f'}(y) \sim_{h\to 0} \frac{1}{nh^3} f(y) \int K'^2(t)\,dt.$$

The variance of this estimator, $\mathcal{O}(1/(nh^3))$, admits a different rate that for $\widehat{f}(y)$, which makes convergence rates pretty distinct.

- *The same phenomenon occurs for*

$$r' = \frac{g'f - gf'}{f^2}$$

or for higher order derivatives of f or of r.

Such estimates may be analogously controlled. Anyway we shall not develop more their theory in those notes.

Exercise 11. *Let $[a,b]$ be a compact interval. Consider the previous regression setting. Provide bounds for*

$$\sup_{x\in[a,b]} \mathbb{E}|\widehat{g}(x)-g(x)|^p.$$

Deduce convergence results for the Nadaraya-Watson estimation of a regression function.

Hint. Proceed as in Remark 3.3.5. ∎

3.4 Division trick

This section is a new visit to a result of interest for statistics, as it will be stressed below. This *ratio-trick* was initiated in [Colomb, 1977] and reformulated in [Doukhan and Lang, 2009], which we try to follow and to simplify. This result does not assert any independence which give it a place in this Chapter.
Setting $D=\mathbb{E}D_n$, and $N=\mathbb{E}N_n$ where N_n, D_n are random quantities, this is an interesting question to get evaluations for centered moments of ratios in some special cases, when this may be expected that

$$\mathbb{E}\left|\frac{N_n}{D_n}-\frac{N}{D}\right|^p \le \mathcal{O}\Big(\mathbb{E}|N_n-N|^p+\mathbb{E}|D_n-D|^p\Big) \tag{3.6}$$

Assume that this ratio appears as a weighted sum where $V_i \ge 0$

$$D_n = a_n\sum_{i=1}^n V_i, \qquad N_n = a_n\sum_{i=1}^n U_iV_i.$$

Maybe more simply, set

$$w_i = \frac{V_i}{\sum_{j=1}^n V_j},$$

then one may rewrite

$$R_n = \frac{N_n}{D_n} = \sum_{i=1}^n w_iU_i, \quad \text{with} \quad \sum_{i=1}^n w_i = 1,\ w_i \ge 0. \tag{3.7}$$

In the the general case for the previous relation (3.6) to hold we prove:

Theorem 3.4.1. *Assume that for some numbers* $1 \leq 2m \geq p \leq q \leq t$ *and that the sequence*(v_n) *is such that* $v_n \downarrow 0$ *(as* $n \uparrow \infty$*) there exists an absolute constant* $M > 0$ *such that:*

$$\begin{aligned} \max_{1\leq i\leq n} \|V_i\|_t + v_n^a n^{\frac{1}{t}} &\leq M && (3.8)\\ \|D_n - D\|_p + \|N_n - N\|_q &\leq M v_n && (3.9)\\ \textit{with: } a = \frac{(m-p)s+pm}{p(s-m)} & \end{aligned}$$

Then the relation (3.6) holds.

A useful main Lemma follows:

Lemma 3.4.1. *For each* $z \in \mathbb{R}$, *and* $0 \leq a \leq 1$ *the following inequality holds:*

$$\left|\frac{1}{1-z} - 1\right| \leq |z| + \frac{|z|^{1+a}}{|1-z|}$$

Proof. From relations

$$\frac{1}{1-z} = 1 + \frac{z}{1-z} = 1 + z + \frac{z^2}{1-z}$$

thus since $0 \leq a \leq 1$:

$$\begin{aligned} \left|\frac{1}{1-z} - 1\right| &\leq \max(|z| + \frac{|z|^2}{|1-z|}, |\frac{|z|}{|1-z|})\\ &\leq |z| + \frac{|z|\max(1,|z|)}{|1-z|}\\ &\leq |z| + |\frac{|z|^{1+a}}{|1-z|} \end{aligned}$$

The last inequality follows from the elementary eqn. (12.16). ∎

In the previous Lemma set $z = (D - D_n)/D$ then quote with $R =$

N/D that

$$\begin{aligned} |R_n - R| &\leq \left|R_n - \frac{N_n}{D}\right| + \frac{1}{D}|N_n - N| \\ &= |N_n|\left|\frac{1}{D_n} - \frac{1}{D}\right| + \frac{1}{D}|N_n - N| \\ &= \frac{|N_n|}{D}\left|\frac{D}{D_n} - 1\right| + \frac{1}{D}|N_n - N| \\ &= \frac{|N_n|}{D}\left|\frac{1}{1-z} - 1\right| + \frac{1}{D}|N_n - N| \\ &= \frac{|zN_n|}{D} + |R_n||z|^{1+a} + \frac{1}{D}|N_n - N| \end{aligned}$$

Thus

$$\begin{aligned} \|R_n - R\|_m &\leq A + B + C \\ A &= \frac{1}{D^2}\|(D_n - D)N_n\|_m \\ B &= \frac{1}{D^{2+a}}\|R_n|D_n - D|^{1+a}\|_m \\ C &= \frac{1}{D}\|N_n - N\|_m \end{aligned}$$

and for constants denoted $c, c', c'' \ldots > 0$:

$$\begin{aligned} A &\leq \frac{N}{D^2}\|D_n - D\|_m + \frac{1}{D^2}\|(D_n - D)(N_n - N)\|_m \\ &\leq c(v_n + \|(D_n - D)(N_n - N)\|_m) \\ &\leq c(v_n + \|N_n - N\|_p\|D_n - D\|_q) \\ &\leq c'(v_n + v_n^2) \qquad (\text{since} \ \ p, q \geq 2m) \\ &\leq c'''v_n \end{aligned}$$

Now quote that since (3.7) writes R_n as a convex combination, we obtain $|R_n| \leq \max_{1\leq i\leq n}|U_i|$; thus an idea of Gilles Pisier entails:

$$\mathbb{E}|R_n|^s \leq (\mathbb{E}|R_n|^t)^{\frac{s}{t}} \leq (nM^t)^{\frac{s}{t}}, \qquad \text{for } 1 \leq s \leq t.$$

Now use Hölder inequality writes

$$\|YZ\|_m \leq \|Y\|_{um}\|Z\|_{vm}, \qquad \text{if} \quad \frac{1}{u} + \frac{1}{v} = 1, \tag{3.10}$$

thus with $Y = R_n$, $Z = |D_n - D|^{1+a}$, $um = s$ $vm = (1+a)p$:

$$B \le \frac{1}{D^{2+a}} \|R_n \cdot |D_n - D|^{1+a}\|_m \le \|R_n\|_s \|D_n - D|\|_q^{1+a} \le c''' n^{\frac{1}{t}} v_n^{1+a}.$$

Quote that

$$\frac{m}{s} + \frac{m}{(1+a)p} = 1 \Longrightarrow \frac{1}{(1+a)p} = \frac{1}{m} - \frac{1}{s} = \frac{s-m}{ms},$$

thus

$$1 + a = \frac{ms}{p(s-m)} \Longrightarrow a = \frac{(m-p)s + pm}{p(s-m)},$$

we need $v_n^a n^{\frac{1}{t}} = \mathcal{O}(1)$, the result follows since $m \le q$, $C \le v_n/D$. ∎

Example 3.4.1. *Relations (3.6) are needed in many cases, examples are provided below:*

1. **Empirical estimate for censored data.** *Here one intends to fit the mean of the incompletely observed iid sequence $(U_t)_{t\ge 0}$. Namely we suppose that this is according to the fact that an independent sequence Bernoulli distributed sequence $V_t \sim b(p)$ take the value 1. The observed variables are thus $X_t = U_t V_t$ and there number is $D_n = \sum_{i=1}^n V_i$. Now with $a_n = n, v_n = 1/\sqrt{n}$ we calculate $D = p$ and $N = p\mathbb{E}V_0$ so that $R = \mathbb{E}V_0$.*

2. **Regression with random design.** *For the previous Nadaraya-Watson estimate (3.5) we may complete exercise 11. This is indeed important to bound centered moments $\|\widehat{r}(x) - r(x)\|_m$ as well as uniform moments $\|\sup_{x\in[a,b]} |\widehat{r}(x) - r(x)|\|_m$. For clarity we will only address the first question but the other one is handled analogously as in the exercise 11.*
   Assume the functions to have regularities $r = 2$ to allows the use of non-negative kernels. Then from the section above, the biases of $\widehat{f}(x)$ and $\widehat{g}(x)$ admit order h^2.
   The previous relation make the hard part of the job since with $N_n = \widehat{g}(x)$ and $D_n = \widehat{f}(x)$ and here $a_n = 1$ ([1]) and $v_n = \frac{1}{\sqrt{nh}}$,

[1] An alternative choice is $a_n = 1/nh$ and $N_n = nh\widehat{g}(x)$ and $D_n = nh\widehat{f}(x)$.

it should imply:

$$\begin{aligned}
\|\widehat{r}_n(x) - r(x)\|_m &\le \left\|\widehat{r}_n(x) - \frac{\widehat{g}(x)}{\widehat{f}(x)}\right\|_m + \left\|\frac{\widehat{g}(x)}{\widehat{f}(x)} - \frac{g(x)}{f(x)}\right\|_m \\
&\le C\left(\frac{1}{\sqrt{nh}} + h^2\right) \\
&\le 2Cn^{-\frac{2}{5}}, \qquad \textit{with a choice} \quad h = n^{-\frac{1}{5}}
\end{aligned}$$

Exercise 12 (3.4.1-1, continued)**.** *Make precise the assumptions in in Example 3.4.1 item 1.*

Hint. First conditions (3.8) follow from independence, and $v_n^a n^{\frac{1}{t}} = n^{\frac{1}{t}-\frac{a}{2}}$ if bounded in case

$$t \ge \frac{2p(s-m)}{(m-p)s+pm}. \qquad \blacksquare$$

Exercise 13 (3.4.1-2, continued)**.** *Make precise the assumptions in Example 3.4.1 item 2.*

3.5 A semi-parametric test

In case the model is indexed by a class of functions but the only parameter of interest is a constant in $\mathbb{R}^d$, the frame is semi-parametric. An example of such semi-parametric estimation is provided here.
Let $w : \mathbb{R} \to \mathbb{R}$ be a weight function such that the following integral converges. We estimate the energy $\theta = \int f^2(x)w(x)\,dx$ from a plug-in estimator of the density f.

$$\widehat{\theta}_n = \int f_{n,h}^2(x)w(x)dx \tag{3.11}$$

here $h = h_n \downarrow 0$ will also satisfy additional conditions described later. Set $\overline{\theta}_n = \int (\mathbb{E} f_{n,h})^2 (x) w(x) dx$, then

$$\begin{aligned}
\widehat{\theta}_n - \theta &= \widehat{\theta}_n - \overline{\theta}_n + \overline{\theta}_n - \theta \\
&= \int (f_{n,h}(x) - \mathbb{E} f_{n,h}(x))^2 \, w(x) \, dx \\
&+ \int (f_{n,h}(x) - \mathbb{E} f_{n,h}(x)) \, (2\mathbb{E} f_{n,h}(x)) w(x)) \, dx \\
&+ \int \left(\mathbb{E} f_{n,h}^2(x) - f^2(x)\right) w(x) dx \\
&= \int (f_{n,h}(x) - \mathbb{E} f_{n,h}(x))^2 \, w(x) \, dx + \mathcal{O}\left(\frac{1}{nh} + h^2\right)
\end{aligned}$$

The expressions $\mathcal{O}$ obtained are in $\mathbb{L}^1$ and thus in probability too. We use the previous bounds in section 3.3.

Theorem 3.5.1. *Under the previous assumptions, if* $nh_n^2 \to 0$, $nh_n^4 \to \infty$*:*

$$\sqrt{n} \left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}}_{n\to\infty} \mathcal{N}(0, V), \quad V = 4\, Var\, (f(X_1) w(X_1)).$$

Proof. Set $v(x) = 2f(x)w(x)$, the remarks above yield to the study of

$$\int (f_{n,h}(x) - \mathbb{E} f_{n,h}(x)) \, v(x) \, dx = \frac{1}{n} \sum_{i=1}^{n} (v(X_i) - \mathbb{E} v(X_i) + \Delta_i - \mathbb{E}\Delta_i)$$

the above sums is decomposed as sums of independent random variables with $\Delta_i = \int K(s)(v(X_i + sh) - v(X_i))ds$, since conditions over $h = h_n$ entail that remainder terms may by neglected. To conclude use the Central Limit Theorem for the iid random variables $v(X_i)$. The dominated convergence theorem implies $\mathbb{E}\Delta_i^2 \to 0$.
Thus:

$$\frac{1}{n} \sum_i \mathbb{E}(\Delta_i - \mathbb{E}\Delta_i)^2 \to 0. \qquad \blacksquare$$

Example 3.5.1 (some other parameters of interest)**.**

- *Fisher information* $I(f) = \int f'^2/f$ *may also be estimated under comparable conditions.*
  For this we leave as an exercise that $f'_{n,h}$ *is also a convergent estimate of* f' *and is asymptotically Gaussian (with normalization* $\sqrt{nh^3}$*).*

  Differentiability of the map $(u,v) \mapsto u^2/v$ *yields an affine approximation of this non linear functional of the couple* $\big(f_{n,h}, f'_{n,h}\big)$.

- *Using bivariate iid samples* (X_n, Y_n) *yields estimation of the regression function:*

  $$r(x) = \mathbb{E}(Y_0|X_0 = x).$$

  We already mentioned that $\widehat{r} = \widehat{g}/\widehat{f}$ *with*

  $$\widehat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right),$$

  accurately estimates r*; this is Nadaraya-Watson estimate.*

- *If one is involved to test the linearity of* r*,*

  $$r'' = \frac{D(f,g,f',g',f'',g'')}{f^3} = 0,$$

  or analogously $D(f,g,f',g',f'',g'') = 0$ *where this expression is a polynomial wrt the derivatives of* f *and* g*.*
  Since the function D *is a polynomial, a Taylor expansion is easy to derive.*
  Tests of linearity for r *by considering a CLT for the conveniently renormalized expressions:*

  $$\int D^2(f,g,f',g',f'',g'')w.$$

Exercise 14. *Extend ideas in last item of Example 3.5.1 to propose a linearity test for a regression function.*

Chapter 4

Stationarity

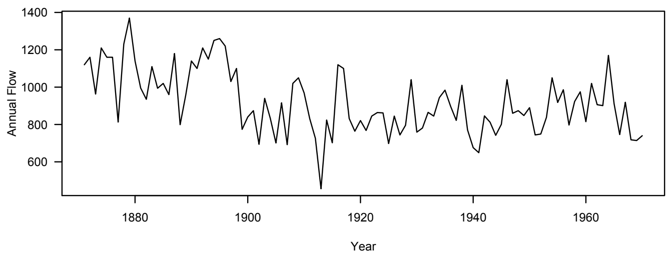


Figure 4.1: Annual flow of the river Nile at Ashwan 1871-1970.

Some bases for the theory of time series are given below. Time series are sequences $(X_n)_{n\in\mathbb{Z}}$ of random variables defined on a probability space (always denoted by $(\Omega, \mathcal{A}, \mathbb{P})$) and with values in a measured space $(E, \mathcal{E})$.
Another extension mainly avoided in these notes is that of random fields $(X_n)_{n\in\mathbb{Z}^d}$.
This means that we will never hesitate to assume that sequences of independent random variables can be defined on the same probability

space.
A classical time series is used as a classically non-linear one, this is Nile flooding data, see Figure 4.1 and some more financial data are designed in Figures 4.2 (with daily and longer duration data) for which stationarity seems more problematic.

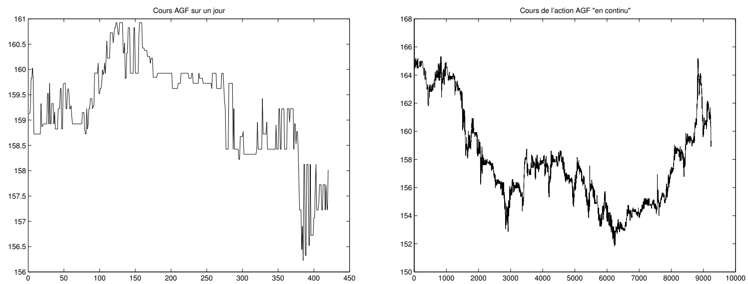


Figure 4.2: AGF stock values.

4.1 Stationarity

Definition 4.1.1 (strict stationarity)**.** *A random sequence* $(X_n)_{n\in\mathbb{Z}}$ *is strictly stationary if, for each* $k \geq 0$*, the distribution of the vector* $(X_l, \ldots, X_{l+k})$ *does not depend on* $l \in \mathbb{Z}$*.*

Definition 4.1.2 (weak stationarity)**.** *A random sequence* $(X_n)_{n\in\mathbb{Z}}$ *is second order stationary if* $\mathbb{E}X_l^2 < \infty$ *and if only:*
$\mathbb{E}X_l = \mathbb{E}X_0$ *and* $Cov(X_l, X_{k+l}) = Cov(X_0, X_k)$*, for each* $l, k \in \mathbb{Z}$*.*
We shall denote by m *the common mean of* X_n *and by* $r(k) = Cov(X_0, X_k)$ *the covariance of such a process.*

In other words $(X_n)_{n\in\mathbb{Z}}$ is strictly stationary if for each $k, l \in \mathbb{N}$ and each function continuous and bounded $h : \mathbb{R}^{k+1} \to \mathbb{R}$:

$$\mathbb{E}h(X_l, \ldots, X_{l+k}) = \mathbb{E}h(X_0, \ldots, X_k).$$

Under second moment assumptions strict stationarity implies second order stationarity (set $k = 1$ and h a second degree polynomial.

Under the Gaussian assumption we will see that both notions coincide. Anyway this is not true in general.

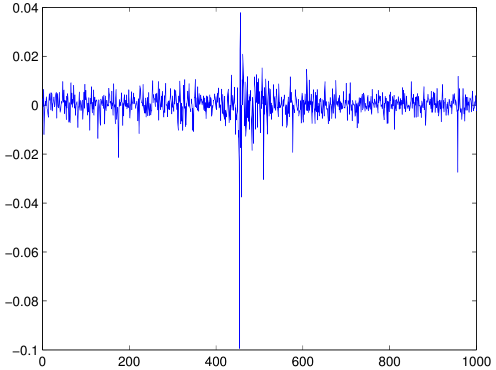


Figure 4.3: SP500 data.

Example 4.1.1. *We consider an independent identically distributed sequence* $(\xi_n)_{n\in\in\mathbb{Z}}$*:*

1. *The iid sequence* $(\xi_n)_{n\in\in\mathbb{Z}}$ *is always strictly stationary, anyway if those variables do not admit finite second order moments, is a strictly stationary but not a second order stationary sequence.*

2. *Assume now that* $\mathbb{E}\xi_0 = 0$ *the sequence* $X_n = \xi_n\xi_{n-1}$ *is thus centered and orthogonal but not independent sequence, indeed if those variables admit 4-th order moments:*

$$\begin{aligned} Cov(X_n^2, X_{n-1}^2) &= \mathbb{E}\xi_n^2\xi_{n-1}^4\xi_{n-2}^2 - \mathbb{E}\xi_n^2\xi_{n-1}^2\mathbb{E}\xi_{n-1}^2\xi_{n-2}^2 \\ &= \left(\mathbb{E}\xi_0^2\right)^2 Var\,\xi_0^2. \end{aligned}$$

   does not vanish if ξ_0^2 *is not a.s. constant.*

3. *A simple modification of this example yields a second order sta-*

tionary sequence which is not strictly stationary:

$$X_n = \xi_n \left(\sqrt{1 - \frac{1}{n}} \cdot \xi_{n-1} + \frac{1}{\sqrt{n}} \cdot \xi_{n-2} \right).$$

If $\mathbb{E}\xi_n^2 = 1$ then $\mathbb{E}X_n X_m = 0$ or 1 if $n \neq m$ or $n = m$. Non stationarity relies on the calculation of $\mathbb{E}X_n X_{n-1} X_{n-2}$ which depends on n.

4. *Write more generally $X_n = \xi_n V_n$ for a sequence such that V_n is independent of ξ_n (as before where $V_n = c_n \xi_{n-1} + s_n \xi_{n-2}$ for constants such that $c_n^2 + s_n^2 = 1$).*
   This sequence is always centered and orthogonal if $\mathbb{E}V_n^2 < \infty$.
   Also using independence $\mathbb{E}X_n^2 X_{n-1} = \mathbb{E}V_n V_{n-1}^2 \xi_{n-1}^2$.
   If now the sequence V_n is independent of the sequence ξ_n we may take an analogue example $V_n = c_n \zeta_{n-1} + s_n \zeta_{n-2}$ for a sequence ζ_n independent of ξ_n in order to finish the previous calculation which we propose as an exercise.

5. *Write now $V_n^2 = c_n \xi_{n-1}^2 + s_n \xi_{n-2}^2$. If $a = \mathbb{E}\xi_0^4 < \infty$ then:*

$$\begin{aligned} \mathbb{E}X_n^4 &= a\mathbb{E}V_n^4 = a\mathbb{E}(c_n \xi_{n-1}^2 + s_n \xi_{n-2}^2)^2 \\ &= a(a(c_n^2 + s_n^2) + 2s_n c_n) \\ &= a(a + 2s_n c_n). \end{aligned}$$

   is not a constant in general. And again we have a second order stationary sequence which is not strictly stationary.

6. *Clearly real data as those for Standard and poor in Figure 4.3 are not stationary! Large pics are September 11...*

Remark 4.1.1. *Stationarity effects are rather mathematical notions. For instance daily insurance real data in (Figure 4.2) nonstationarity is not evident while it is clear for longer observations (Figure 4.6); anyway a more careful analysis of the previous daily data proves that even daly data are not stationary see Figure 4.4 for longer term and 4.5 for short time observations. which look a bit like the previous one. A scaling effect is also to be taken into account but this does not enter our present scope.*

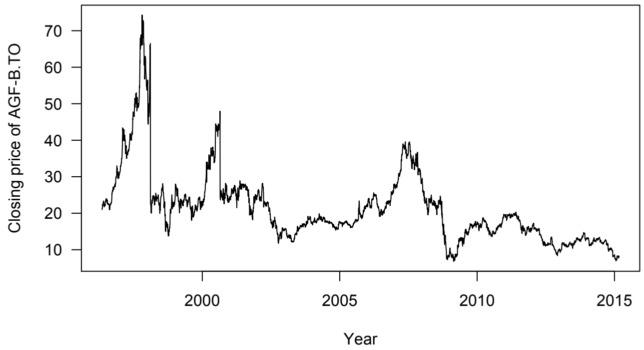


Figure 4.4: AGF stock values.

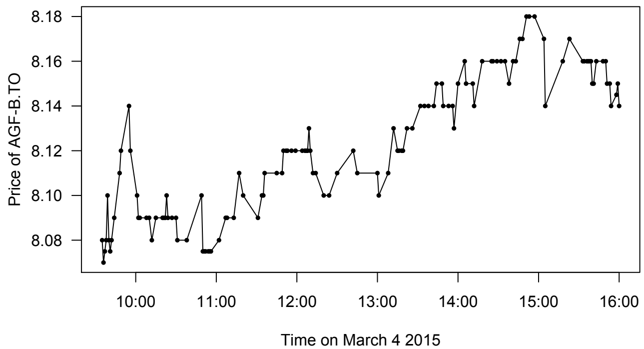


Figure 4.5: Daily AGF stock values.

4.2 Spectral representation

This is easy and important to quote the following property of covariances. Fix $n \in \mathbb{N}^*$. Let $c_l \in \mathbb{C}$ for all $|l| \le n$, then setting $c = (c_l)_{|l|\le n}$ and $\Sigma_n = (r_{|i-j|})_{|i|,|j|\le n}$ we obtain:

$$c^t \Sigma_n \overline{c} = \sum_{|i|,|j|\le n} c_i \overline{c_j} r_{|i-j|} = \mathbb{E}\Big| \sum_{|i|\le n} c_i X_i \Big|^2 \ge 0. \qquad (4.1)$$

Theorem 4.2.1 (Herglotz). *If a sequence $(r_n)_{n\in\mathbb{Z}}$ satisfies (4.1) then there exists a non-decreasing function G (essentially unique) with $G(-\pi) = 0$ and*

$$r_k = \int_{-\pi}^{\pi} e^{ik\lambda} dG(\lambda).$$

Notation. Integral $dG(\lambda)$ is considered in the meaning of Stieljes: define the measure μ with

$$\mu([-\pi,\lambda]) = G(\lambda), \qquad \forall\lambda \in [-\pi,\pi].$$

If $h : [-\pi,\pi] \to \mathbb{R}$ is continuous:

$$\int_{-\pi}^{\pi} h(\lambda)dG(\lambda) = \int_{-\pi}^{\pi} h(\lambda)\mu(d\lambda).$$

Proof. Set

$$g_n(\lambda) = \frac{1}{2\pi n}\sum_{s=0}^{n-1}\sum_{t=0}^{n-1} r_{t-s}e^{-i(t-s)\lambda} = \frac{1}{2\pi}\sum_{j=-(n-1)}^{n-1}\left(1-\frac{j}{n}\right) r_j e^{-ij\lambda},$$

and $G_n(\lambda) = \int_{-\pi}^{\lambda} g_n(u)\,du$ then relation (4.1) implies $g_n(u) \geq 0$ hence G_n is continuous, non-decreasing and $G_n(\pi) = r_0$.
From a compactness argument, some subsequence $G_{n'}$ of G_n is convergent ([1]). Note that $dG_n(\lambda) = g_n(\lambda)d\lambda$ then

$$\left(1-\frac{k}{n}\right) r_k = \int_{-\pi}^{\pi} e^{ik\lambda} dG_n(\lambda).$$

An integration by parts yields

$$r_k = (-1)^k r_0 - ik\int_{-\pi}^{\pi} e^{ik\lambda} dG_n(\lambda)\,d\lambda,$$

[1] Use a triangular scheme, by successive extraction of convergent subsequences. Choose a dense sequence $(\lambda_k)_k$ in $[-\pi,\pi]$.
Here $\phi_{k+1}(n)$ is a subsequence of $\phi_k(n)$ such that $G_{\phi_{k+1}(n)}(\lambda_{k+1})$ converges as $n \to \infty$. We then set $G_{\phi(n)} = G_{\phi_n}(n)$. ∎

and implies the uniqueness of G.
Existence of G follow from the fact that it is the only possible limit of such a convergent subsequence $G_{n'}$. ∎

Definition 4.2.1. *The spectral measure of the second order stationary process $(X_n)_{n\in\mathbb{Z}}$ (defined from G) is such that for each $\lambda \in [-\pi, \pi]$:*

$$\mu_X([-\pi, \lambda]) = G(\lambda).$$

If G is derivable, the spectral density of the process $(X_n)_{n\in\mathbb{Z}}$ is the derivative $g = G'$.

Example 4.2.1.

- *For an orthogonal sequence (i.e $\mathbb{E}X_k X_l = 0$ for $k \neq l$, as in Examples 4.1.1 2., 3., 4. and 5. with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$: $G(\lambda) = 1/2 + \lambda/2\pi$, the measure associated is Lebesgue on $[-\pi, \pi]$.*

- *The random phase model admits complex values. Given constants $a_1, b_1, \ldots, a_k, b_k \in \mathbb{R}$ and independent uniform random variables $U_1, \ldots, U_k$ on $[-\pi, \pi]$ this model is defined through the relation*

  $$X_n = \sum_{j=1}^{k} a_j e^{i(nb_j + U_j)},$$

  one computes

  $$Cov(X_s, X_t) = \mathbb{E}X_s\overline{X_t} = r_{s-t} = \sum_{j=1}^{k} |a_j|^2 e^{i(s-t)b_j}.$$

  This model is associated with a stepwise constant function G.

- *Let $(\xi_n)_{n\in\mathbb{Z}}$ be a centered and independent identically distributed sequence such that $\mathbb{E}\xi_n^2 = 1$, let $a \in \mathbb{R}$, the moving average model $MA(1)$ is defined as*

  $$X_n = \xi_n + a\xi_{n-1}.$$

  Here, $r_0 = 1 + a^2$, $r_1 = r_{-1} = a$, and $r_k = 0$ if $k \neq -1, 0, 1$.

With the proof of Herglotz theorem we derive

$$\begin{aligned} g(\lambda) &= \frac{1}{2\pi}\left(r_0 + 2r_1\cos\lambda\right) \\ &= \frac{1}{2\pi}\left(1 + a^2 + 2a\cos\lambda\right) \\ &= \frac{1}{2\pi}\left((1 + a\cos\lambda)^2 + a^2\sin^2\lambda\right) \\ &\geq 0 \end{aligned}$$

Notation. For a function $g : [-\pi, \pi] \to \mathbb{C}$ denote $g(I) = g(v) - g(u)$ if $I = (u, v)$ is an interval. If $g : [-\pi, \pi] \to \mathbb{R}$ is nondecreasing, we thus identify g and the associated nonnegative measure.

Definition 4.2.2 (Random measure)**.** *A random measure is defined with a random function $Z : \Omega \times [-\pi, \pi] \to \mathbb{C}$, $(\omega, \lambda) \mapsto Z(\omega, \lambda)$, nondecreasing for each $\omega \in \Omega$, with $\mathbb{E}|Z(\lambda)|^2 < \infty$ and such that there exists a nondecreasing function $H : [-\pi, \pi] \to \mathbb{R}^+$ with,*

- $\mathbb{E}Z(\lambda) = 0$ *for* $\lambda \in [-\pi, \pi]$,

- $\mathbb{E}Z(I)\overline{Z(J)} = H(I \cap J)$ *for all the intervals* $I, J \subset [-\pi, \pi]$.

Let $g : [-\pi, \pi] \to \mathbb{C}$ be measurable and $\int_{-\pi}^{\pi} |g(\lambda)|^2 dH(\lambda) < \infty$, we define a stochastic integral with respect to a deterministic function

$$\int g(\lambda) dZ(\lambda),$$

in two steps:

- If g is a step function, $g(\lambda) = g_s$ for $\lambda_{s-1} < \lambda \leq \lambda_s$ (with $-\pi = \lambda_0 \leq \lambda \leq \lambda_S = \pi$) $0 < s \leq S$, set

  $$\int g(\lambda) dZ(\lambda) = \sum_{s=1}^{S} g_s Z([\lambda_{s-1}, \lambda_s]).$$

  Notice that

  $$\mathbb{E}\left|\int_{-\pi}^{\pi} g(\lambda) dZ(\lambda)\right|^2 \quad = \quad \sum_{s,t} g_s \overline{g_t} \mathbb{E}Z([\lambda_{s-1}, \lambda_s])\overline{Z([\lambda_{t-1}, \lambda_t])}$$

$$
\begin{aligned}
&= \sum_s |g_s|^2 \mathbb{E}|Z([\lambda_{s-1}, \lambda_s])|^2 \\
&= \sum_s |g_s|^2 H([\lambda_{s-1}, \lambda_s]) \\
&= \int_{-\pi}^{\pi} g^2(\lambda) dH(\lambda).
\end{aligned}
$$

- Else approximate g by a sequence of step functions g_n with

$$\int_{-\pi}^{\pi} |g(\lambda) - g_n(\lambda)|^2 dH(\lambda) \to_{n\to\infty} 0.$$

The sequence $Y_n = \int g_n(\lambda) dZ(\lambda)$ is such that if $n > m$,

$$\mathbb{E}|Y_n - Y_m|^2 = \int_{-\pi}^{\pi} |g_n(\lambda) - g_m(\lambda)|^2 dH(\lambda) \to_{n\to\infty} 0.$$

This sequence is thus Cauchy. Thus it converges in $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ and its limit defines the considered integral.

Example 4.2.2. *Simple examples are provided from processes with independent increments.*

- *A natural example of such a random measure is the Brownian measure. Namely, denote $W([a, b]) = W(b) - W(a)$ then this random measure is defined with the Lebesgue measure as a control spectral measure λ.*

- *Another random measure of interest is Poisson process on the real line, see Definition 7.3.3.*

Theorem 4.2.2 (Spectral representation of stationary sequences)**.** *Let $(X_n)_{n\in\mathbb{Z}}$ be a centered second order stationary random process then there exists a random spectral measure Z such that*

$$X_n = \int e^{in\lambda} dZ(\lambda),$$

and this random measure is associated to the spectral measure of the process.

Relevant random spectral measures are reported as Examples 4.3.1.

Proof. The spectral function G of the process X_n is nondecreasing, hence its discontinuities are at most denumerable set denoted D_G ([2]). If $I = (a, b)$ is an interval with $a, b \notin D_G$, set

$$Z_n(I) = \frac{1}{2\pi} \sum_{|j| \leq n} X_j \int_a^b e^{-iju} du,$$

then the sequence $(Z_n(I))_{n \geq 1}$ is Cauchy in $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ since for $n > m$,

$$\mathbb{E}|Z_n(I) - Z_m(I)|^2 = \frac{1}{4\pi^2} \mathbb{E} \left| \sum_{m < |j| \leq n} X_j \int_a^b e^{-iju} du \right|^2 = \int_{-\pi}^{\pi} |h_n - h_m|^2 dG.$$

Denote now by h_n, the truncated Fourier series of the indicator function $\mathbb{1}_I$:

$$h_n(\lambda) = \frac{1}{2\pi} \sum_{|j| \leq n} \int_a^b e^{-ij(u-\lambda)} du.$$

Write $Z(I)$ for the limit in $\mathbb{L}^2$ of $Z_n(I)$, then $\mathbb{E}Z(I) = 0$ because $\mathbb{E}X_n = 0$ and with immediate notations

$$\mathbb{E}Z(I)\overline{Z(J)} = \lim_n \mathbb{E}Z_n(I)\overline{Z_n(J)} = \lim_n \int_{-\pi}^{\pi} h_{I,n}\overline{h_{J,n}} dG = G(I \cap J)$$

in case the extremities of I, J are not in D_G. Because the the set of continuity points is dense, taking limits allows to consider extremities

[2]Recall that monotonic functions admit limits on the left and on the right at each point thus non-empty open intervals $(f(x-), f(x+))$ are disjointed in $\mathbb{R}$. Choose a rational number in each of them to derive this result. ∎

of this interval only in D_G. To conclude, quote that

$$\begin{aligned}
\mathbb{E}X_n\overline{Z_n(I)} &= \frac{1}{2\pi}\sum_{|j|\le n} r_{n-j}\int_a^b e^{iju}du \\
&= \int_{-\pi}^{\pi}\frac{dv}{2\pi}\int_a^b \sum_{|j|\le n} e^{ij(u-v)}dG(u) \\
&= \int_a^b e^{inv}dG(v).
\end{aligned}$$

Hence for step functions f:

$$\mathbb{E}X_n\overline{\int f(\lambda)dZ(\lambda)} = \int_{-\pi}^{\pi} e^{in\lambda}\overline{f(\lambda)}dG(\lambda).$$

Figure 4.6: March 4, 2015's AGF data and local autocovariances.

This extends to continuous functions f by considering limits. If $f(\lambda) = e^{in\lambda}$ then

$$\mathbb{E}\left|X_n - \int e^{in\lambda}dZ(\lambda)\right|^2 = r_0 - 2r_0 + r_0 = 0. \blacksquare$$

Example 4.2.3. *Examples of spectral densities may be found in Example 4.2.1. Besides measures with independent increment (Example 4.2.2), some relevant examples are reported as Examples 4.3.1. In Figure 4.7 we plot empirical autocovariances. For non stationary processes a local empirical covariance may be plotted too as in Figure 4.6.*

4.3 Range and spectral density

Here we denote $(X_n)_{n\in\mathbb{Z}}$ a centered second order stationary process. Assume that

$$\sum_{k=0}^{\infty} r_k^2 < \infty,$$

then the spectral density

$$g(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_k e^{-ik\lambda},$$

is defined in $\mathbb{L}^2([-\pi,\pi])$ and

$$r_k = \int_{-\pi}^{\pi} e^{ik\lambda} g(\lambda) d\lambda.$$

Here the spectral measure G of the process is absolutely continuous with derivative $g \in \mathbb{L}^2$.

Definition 4.3.1. *If a centered second order stationary process* (X_n) *satisfies*

$$\sum_{k=0}^{\infty} r_k^2 < \infty \text{ and } \sum_{k=0}^{\infty} |r_k| = \infty.$$

it is long range dependent (LRD).
If

$$\sum_{k=0}^{\infty} |r_k| < \infty,$$

it is short range dependent (SRD). In this case the spectral density g is uniformly continuous and

$$\|g\|_\infty \leq \frac{1}{2\pi} \sum_{k=0}^{\infty} |r_k|.$$

Example 4.3.1. *Some examples follow:*

- *If $r_k \sim k^{-\alpha}$ for $\frac{1}{2} < \alpha < 1$ the sequence is LRD and there exists $\beta > 0$ with $g(\lambda) \sim c\lambda^{-\beta}$ as $\lambda \to 0$.*

- *If the spectral density*

$$g(\lambda) = \frac{\sigma^2}{2\pi},$$

then the sequence

$$\xi_n = \int_{-\pi}^{\pi} e^{in\lambda} Z(d\lambda),$$

is a second order white noise with variance σ^2:

$$\mathbb{E}\xi_n\xi_m = 0 \text{ or } \sigma^2, \qquad \text{according that } n \neq m \text{ or } m = n.$$

This is the case if:

$$Z([0,\lambda]) = \frac{\sigma^2}{2\pi} W(\lambda),$$

with W the Brownian motion. Here Gaussianness of the white noise also implies its independence and it is an independent identically distributed sequence (strict white noise).

If $\lambda \mapsto Z([0,\lambda])$ admits independent increments, the sequence ξ_n is again a strict white noise.
A weak white noise is associated with random spectral measures with orthogonal increments.

- *If*

$$X_n = \sum_{k=-\infty}^{\infty} c_k \xi_{n-k}, \qquad \sum_{k=-\infty}^{\infty} c_k^2 < \infty,$$

then the spectral density g_X of X writes

$$g_X(\lambda) = \left| \sum_{k=-\infty}^{\infty} c_k e^{-ik\lambda} \right|^2 g_\xi(\lambda).$$

To prove it compute X's covariance.
Moreover

$$Z_X(d\lambda) = \left(\sum_{k=-\infty}^{\infty} c_k e^{ik\lambda} \right) Z_\xi(d\lambda),$$

where Z_ξ denotes the random spectral measure associated to ξ. E.g. autoregressive models, AR(p), may also be defined for non independent inputs,

$$X_n = \sum_{k=1}^{p} a_k X_{n-k} + \xi_n.$$

In case the sequence (ξ_n) is a white noise with variance 1, they are such that

$$g_X(\lambda) = \frac{1}{2\pi} \left| 1 - \sum_{k=1}^{p} a_k e^{-ik\lambda} \right|^{-2}$$

Now the spectral density g_X is continuous if the roots of the polynomial

$$P(z) = z^p - \sum_{k=1}^{p} a_k z^{p-k}$$

are outside the complex unit disk. This holds e.g. if $\sum_{k=1}^{p} |a_k| < 1$.

The previous heredity formulas extend to $\mathbb{L}^2-$stationary sequences ξ_n:

Proposition 4.3.1. *Let (X_n) be a centered second order stationary sequence and c_n be a real sequence :*

$$Y_n = \sum_{k=-\infty}^{\infty} c_k X_{n-k}, \quad h(\lambda) = \sum_{k=-\infty}^{\infty} c_k e^{ik\lambda}, \quad \textit{with} \quad \sum_{k=-\infty}^{\infty} c_k^2 < \infty, .$$

Then the sequence Y_n is also centered second order stationary sequence and

$$g_Y(\lambda) = |h(\lambda)|^2 g_X(\lambda), \quad Z_Y(d\lambda) = h(\lambda) Z_X(d\lambda).$$

Proof. The first claim follows from the bilinearity properties of covariance :

$$\mathrm{Cov}\,(Y_0, Y_k) = \sum_{m=-\infty}^{\infty} \left(\sum_{j=-\infty}^{\infty} c_j c_{j-m} \right) r_{k+m}.$$

The second claim is just algebra. ∎

4.3.1 Limit variance

This definition of the range of a process is justified as follows in case X_n is centered, indeed :

$$\mathbb{E}\,|X_1 + \cdots + X_n|^2 = \sum_{s=1}^{n}\sum_{t=1}^{n} \mathbb{E} X_s X_t = \sum_{s=1}^{n}\sum_{t=1}^{n} r_{t-s}$$

Thus:

$$\mathbb{E}\,|X_1 + \cdots + X_n|^2 = \sum_{|k|<n}^{n} (n - |k|) r_k. \tag{4.2}$$

According to the previous section one derives:

Proposition 4.3.2. *If X_n is SRDthen*

$$\mathbb{E}\,|X_1 + \cdots + X_n|^2 \sim n g(0).$$

Proof. Quote that the result is a variant of Cesaro lemma.
This will be enough to prove that:

$$\sum_{|k|<n} |k| r_k = o(n).$$

For each $\epsilon > 0$ there exists K such that $|r_k| < \epsilon$ for $|k| > K$.
Split the expression

$$\sum_{|k|<n} |k||r_k| \le \sum_{|k|<K} |k||r_k| + \epsilon n.$$

∎

Recall that in case $\mathbb{E}X_0 = 0$

$$g(0) = \sum_{k=-\infty}^{\infty} \mathbb{E}X_0 X_k.$$

The previous quantity is thus of a specific importance.
According to the independent case a first possibility is to fit each term of the sum above approximated for a convenient sequence $m = m_n$ by

$$\sigma_m^2 = \sum_{k=-m}^{m} \mathbb{E}X_0 X_k.$$

Then an empirical estimate of this expression writes

$$\widehat{\sigma}_n^2 = \sum_{k=-m}^{m} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}X_i X_{i+k}. \tag{4.3}$$

or alternatively if one only has a sample $X_1, \ldots, X_n$

$$\widehat{\sigma}_n^2 = \sum_{k=-m_n}^{m_n} \frac{1}{n} \sum_{i=1\vee k}^{n\wedge(n+k)} \mathbb{E}X_i X_{i+k}, \tag{4.4}$$

quote that all the terms in the previous sum do not have the same number of elements.
Namely the $k-$element of the sum is over $n - |k|$ terms which makes this estimate biased.
A variant of the previous estimate which is unbiased writes now as

$$\widehat{\sigma}_n^2 = \sum_{k=-m_n}^{m_n} \frac{1}{n-|k|} \sum_{i=1\vee k}^{n\wedge(n+k)} \mathbb{E}X_i X_{i+k}. \tag{4.5}$$

The previous estimate may also be seen as a non parametric estimate of the spectral density a the origin which also justifies the introduction of a smoothing parameter even though one only aims at estimating a real parameter.

4.3.2 Cramer Wold representation

In fact the second order stationary processes write as infinite order moving average of a weak white noise under a weak assumption, the proof of following results may be found in the volume [Azencott and Dacunha-Castelle, 1987]:

Theorem 4.3.1 (Crámer Wald)**.** *Let* $(X_n)_{n\in\mathbb{Z}}$ *be a second ordered stationary sequence with a derivable spectral measure* G *such that* $g = G'$ *satisfies*

$$\int \log g(x)\, dx > -\infty.$$

Then there exists a unique orthogonal sequence ξ_n *second order stationary (weak white noise) with* $\mathbb{E}\xi_0^2 = 1$ *and a sequence* $(c_n)_{n\in\mathbb{N}}$ *with* $\sum_{n=0}^{\infty} c_n^2 < \infty$, $c_0 \geq 0$ *such that*

$$X_n = \mathbb{E}X_0 + \sum_{k=0}^{\infty} c_k \xi_{n-k}. \tag{4.6}$$

Theorem 4.3.2 (Wold decomposition)**.** *Let* $(Z_n)_{n\in\mathbb{Z}}$ *be a second ordered stationary sequence then there exists* X_t, V_t *with* $Z_t = V_t + X_t$ *such that* (X_t) *writes as in eqn. (4.6) and* V_t *is measurable wrt to* $\sigma(\epsilon_u, u \leq s)$ *for each* $s \leq t$*.*

The first part of the representation of Z_t is as before while the second part V_t is something new.
That part is called the deterministic part of Z_t because V_t is perfectly predictable based on past observations X_s for $s \leq t$.
A parameter of a main interest for stationary time series is thus the spectral density.

4.4 Spectral estimation

This section is a very short survey of the question addressed in several nice volumes: see [Azencott and Dacunha-Castelle, 1987], and [Giraitis et al., 2012] for a complete study of the LRD case, see also [Brockwell and Davis, 1991] for parametric setting and for non parametric setting in [Rosenblatt, 1991].
Our aim is more to make explicit how probabilistic limit theory can be used for the development of statistical methods for time series analysis that to provide a course of time series analysis since really nice textbooks are already available. Anyway the present viewpoint allows to present many tools usually not considered directly by statisticians.

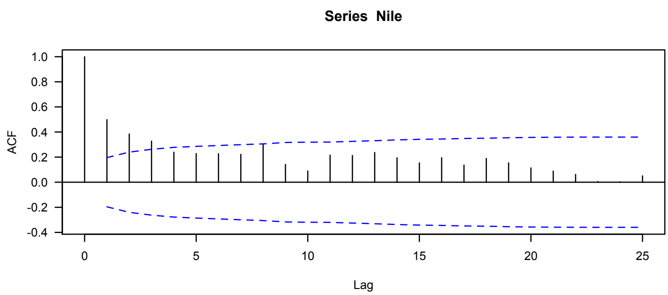


Figure 4.7: Autocorrelation for Nile at Ashwan 1871-1970.

Definition 4.4.1. *For a centered and second order stationary* $(X_t)_{t\in\mathbb{Z}}$ *define the periodogram:*

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{k=1}^{n} X_k e^{-ik\lambda}\right|^2 = \frac{1}{2\pi}\sum_{|\ell|<n} \widehat{r}_n(\ell)e^{-ik\lambda}.$$

for each $n \geq 1$ *and* $\lambda \in \mathbb{R}$, *where*

$$\widehat{r}_n(\ell) = \frac{1}{n}\sum_{k=1\vee(1+\ell)}^{n\wedge(n+\ell)} X_k X_{k+\ell}.$$

Example 4.4.1. *An example of very classical real data the is annual flow of the river Nile at Ashwan 1871-1970 in Figures 4.1 and 4.7 shows the fitted covariances. A rapid decay of covariances is observed from the covariogram.*

Remark 4.4.1. *Quote the last sum is over* $(n-|\ell|)-$*terms thus the estimator* $\widehat{r}_n(\ell)$ *of the covariance* $r(\ell) = \mathbb{E}X_0X_\ell$ *is biased for* $\ell \neq 0$, *which means that we do not necessarily* $\mathbb{E}\widehat{r}_n(\ell) = r(\ell)$.
Remark thus that in case $\sum_\ell ||r(\ell)| < \infty$ *then the spectral density of the process* f *is continuous and that* $\mathbb{E}I_n(\lambda) = f(\lambda)$.
Unfortunately the variance of this estimate of f *does not converge to 0.*
Thus $I_n(\lambda)$ *is not a reasonable estimator of* $f(\lambda$.

Anyway the integrated statistics

$$J_n(g) = \int_0^{2\pi} g(\lambda)I_n(\lambda)\,d\lambda,$$

admit smoother behaviors and usually converge to

$$J(g) = \int_0^{2\pi} g(\lambda)f(\lambda)\,d\lambda.$$

They even may be proved to satisfy a central limit theorem.

The previous feature may be be used in two directions rapidly described in the two following subsections.

4.4.1 Functional spectral estimation

First, we use a kernel method to consider $g \sim \delta_u$ thus for a convenient window width $h = h_n$ and a kernel K we may consider the estimate

$$\widehat{f}_n(\lambda) = I_n \star K_h(\lambda) = \frac{1}{h}\int_0^{2\pi} I_n(\mu)K\left(\frac{\lambda-\mu}{h}\right)\,d\mu.$$

It allows reasonable spectral density estimators.
If now one replaces the smoothing function $\frac{1}{h}K(\frac{\cdot}{h})$ by the Dirichlet kernel

$$D_m(u) = \sum_{k=-m}^{m} e^{iku} = \frac{\sin\left((2m+1)\frac{u}{2}\right)}{\sin\left(\frac{u}{2}\right)},$$

with order $m = m_n = 1/h_n$ the previous estimates writes

$$\widetilde{f}_n(\lambda) = I_n \star D_{m_n}(\lambda) = \int_0^{2\pi} I_n(\mu) D_{m_n}(\lambda - \mu)\, d\mu.$$

almost fits the above mentioned estimate (4.3) of $f(0)$. In fact it writes in such a way that $\widetilde{f}_n(0)$ writes as in eqn. (4.4), contrary to eqn. (4.5) this gives a biased estimate:

$$\widetilde{f}_n(0) = \sum_{k=-m_n}^{m_n} \frac{1}{n} \sum_{i=1\vee k}^{n\wedge(n+k)} \mathbb{E} X_i X_{i+k}.$$

Remark 4.4.2. *Asymptotic properties of such estimates may be derived under specific assumptions on the time series.*
One may prove them by approximating the spectral density by its Fourier expansion. Then standard empirical arguments allow to derive asymptotic properties of such estimates as for the simple empirical means considered in section 3.1 for independent sequences.
Further improvements of inequalities for dependent samples are thus needed to complete the program.
Quote that the case of the kernel estimate is in fact analogue since regularity conditions of a spectral density are tightly related to the quality of their approximation by trigonometric polynomials; this point is proved by using Jackson polynomial approach see in [Lorentz, 1966] or in [Doukhan and Sifre, 2001].

4.4.2 Whittle spectral estimation

Assume that the time series is in a parametric set of models. Maybe ARMA or others: see hereafter. Then the distribution of the whole process $X = (X_t)_{t\in\mathbb{Z}}$ may depend on a parameter θ the spectral density is defined in a family $(f_\theta)_{\theta\in\Theta}$ (for some $\Theta \subset \mathbb{R}^d$) and a suitable estimator, named Whittle estimator, is the value $\widehat{\theta}$ minimizing the contrast, as defined in Section 3.2:

$$\begin{aligned} U_n(\theta) &= \int_0^{2\pi} \left(\log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)} \right) d\lambda \\ &= \int_0^{2\pi} \log f_\theta(\lambda)\, d\lambda + J_n\left(\frac{1}{f_\theta}\right). \end{aligned}$$

Here again central limit theorems extending those for independent sequences allow to expand pointwise the previous expression. An additional argument such as for example a uniform results (see e.g. § 3.1) is then necessary to make the Taylor expansion valid after integration.

4.5 Parametric estimation

Remark also that parameters based on the spectral density maybe be estimated also from other contrast estimators.
Usually there is no close expression for the density $p_\theta(x_1, \ldots, x_n)$ of a sample $(X_1, \ldots, X_n)$ but MLE $\widehat{\theta}$ estimates are defined through the relation:

$$\widehat{\theta} \in \ \mathrm{Argmax}_{\theta\in\Theta} p_\theta(X_1, \ldots, X_n).$$

An interesting special case is that of an homogeneous Markov chain with transitions

$$P_\theta(x, A) = \mathbb{P}_\theta(X_1 \in A | X_0 = x),$$

admitting a density $\pi_\theta(x, y)$ and an invariant measure with density $\nu_\theta(x)$, then:

$$p_\theta(x_1, \ldots, x_n) = \nu_\theta(x_1)\pi_\theta(x_1, x_2)\cdots\pi_\theta(x_{n-1}, x_n).$$

For instance the non linear auto regressive processes

$$X_t = r_\theta(X_{t-1}) + \xi_t,$$

are this way in case ξ_0 admits a density g_θ, and then

$$\pi_\theta(x, y) = g_\theta(y - r_\theta(x)).$$

The MLE of Markov chains written

$$X_t = \xi_t \sigma_\theta(X_{t-1}),$$

with iid centered innovations (ξ_t) writes with

$$\pi_\theta(x, y) = \frac{1}{\sigma_\theta(x)} \cdot g\left(\frac{x}{\sigma_\theta(x)}\right),$$

in case ξ_t admits a density g. Instead of considering p_θ one better considers the minimization of

$$q_\theta(x_1, \dots, x_n) = \pi_\theta(x_1, x_2) \cdots \pi_\theta(x_{n-1}, x_n).$$

Now usually such maximization problems are numerically unstable the QMLE is the minimization of the previous expression but with simply $\xi_0 \sim \mathcal{N}(0,1)$ a normal distribution.
Now the MLE maximizes $\theta \mapsto L_\theta(X_1, \dots, X_n)$. Anyway even in this simplest case of Gaussian inputs f_θ does not usually admit a close form.
This expression, simpler to be minimized writes:

$$L_\theta(X_1, \dots, X_n) = \sum_{t=2}^{n} \frac{X_t^2}{\sigma_\theta^2(X_{t-1})} + \log \sigma_\theta^2(X_{t-1}^2).$$

This estimator is considered in the most general situations in the monograph [Straumann, 2005].

Remark 4.5.1. *A last related remark is that for Gaussian processes with a fixed variance $Var\, X_t \equiv \sigma^2$ the least squares coincide with MLE because of the quadratic expression of a Gaussian density.*

4.6 Subsampling

Besides model based bootstrap techniques in Section 11.3 this section is aimed at explicating the specific features of resampling under dependence.
Namely assume that a limit theorem holds for a sequence

$$t_m(X_1, \dots, X_m) \to_{m \to \infty} T.$$

This is not unusual that the distribution of T is not accessible. Anyway as before a test of goodness-of-fit is based on quantiles of the limiting distribution T. In case one wants more generally fit the distribution of the convergent statistics

$$T_m = t_m(X_1, \dots, X_m), \qquad \text{for some} \qquad m = m_n \ll n.$$

A way to proceed is to consider families of $m-$samples $(X_{i_1},\ldots,X_{i_m})$ with $(i_1,\ldots,i_m)\in E_{m,n}$ and $i_1\le\cdots\le i_m$, then the expression of T_m's distribution is provided from the value of $K(g)=\mathbb{E}g(T_m)$ which is obtained from the empirical method as

$$\widehat{K}(g)=\frac{1}{\text{Card }E_{m,n}}\sum_{(i_1,\ldots,i_m)\in E_{m,n}} g\left(t_m(X_{i_1},\ldots,X_{i_m})\right).$$

In order that the distribution of $t_m(X_{i_1},\ldots,X_{i_m})$ is the same as for T_m a natural assumption is to assume that the distribution of $(X_{i_1},\ldots,X_{i_m})$, is the same as for $(X_1,\ldots,X_m)$.
For iid samples the set $E_{m,n}$ may admits the huge cardinality $\frac{n!}{(n-m)!}\sim n^m$; e.g. select $E_{m,n}$ as the set all the ordered $m-$tuples among $\{1,\ldots,n\}$.
Unfortunately not all $m-$tuples admit the same distribution when independence is omitted.
Two choices of sets may be considered to support this distributional equality,

$$E_{m,n}=\big\{(i+1,\ldots,i+m);0\le i\le n-m\big\},$$

satisfies $\text{Card}\,E_{m,n}=n-m+1$ and gives overlapping samples,

$$E_{m,n}=\Big\{\big((i-1)m+1,\ldots,im\big);1\le i\le n/m\Big\},$$

satisfies $\text{Card}\,E_{m,n}=n/m+1$ for n a multiple of m and gives non-overlapping samples. Again asymptotic consistency of such expressions still relies of moment and exponential inequalities.

Some words of asymptotic.
Set respectively for each of those following schemes:

$$\begin{aligned} g_{i,m} &\equiv g(t_m(X_{(m(i-1)+1},\ldots,X_{(i+1)m})) && (4.7)\\ g_{i,m} &\equiv g(t_m(X_{i+1},\ldots,X_{i+m})) && (4.8)\end{aligned}$$

and the set $E_{m,n}$ is indexed by an integer $i=1,\ldots N\sim n-m$ or n/m.
In order to prove the convergence of such expressions, a simple way

is to calculate the variance of such expressions and from Cesaro trick to derive that

$$\widehat{K}(g) \to_{n\to\infty} \mathbb{E}g(T) \qquad \text{(in probability)}.$$

Eqn. (4.2) entails

$$\operatorname{Var}\widehat{K}(g) \le \frac{1}{\text{Card } E_{m,n}} \sum_{i\in E_{m,n}} |\operatorname{Cov}(g_{0,m}, g_{i,m})|.$$

Usually $g(x) = \mathbb{1}_{(x\le u)}$ so that using Exercise 24 the limit in probability

$$\sup_u |K_{n,m}(u) - \mathbb{P}(T \le u)| \to_{n\to\infty} 0, \quad K_{n,m}(u) = \widehat{K}(g),$$

holds uniformly with respect to u by using Exercise 9 as in the proof of Glivenko-Cantelli Theorem 3.1.1.

Remark 4.6.1. *Such uniform convergences are taken into account to consider non convergent cases, in [Doukhan et al., 2011] we consider extreme value theory.*

Remark 4.6.2. *In order to prove almost sure convergence of such expressions, higher order moments need to be accurately bounded, [Doukhan et al., 2011].*

Part II

Models of time series

Chapter 5

Gaussian chaos

Gaussian distributions play a special role in the field of probability since they appear as limit distributions from the CLT. See e.g. Theorem 2.1.1 and its dependent counterpart, Lemma 11.5.1.
Gaussian random variables are pretty natural and functions of such Gaussian random variables also mechanically appear. A precise study of the Gaussian distributions is deferred to Appendix § A.2.
Gaussian processes admit also an essential property which led them to be extensively used in statistics for which descriptive solutions are a real tool:
$\mathbb{L}^2-$*properties of Gaussian processes are equivalent to their distributional properties.*
Gaussian linear spaces spanned by multivariate random variables of Gaussian random processes are thus also natural tools from this geometric viewpoint...
Since those lectures are devoted to non-linear modeling, the last step in this Gaussian setting is the use of non-linear functions of Gaussian processes. We consider here $\mathbb{L}^2$ functionals in order to work inside a Hilbert setting.
The Gaussian chaos is simply the set of such $\mathbb{L}^2-$integrable functionals of such Gaussian random variables. In case of a 1-dimensional Gaussian random variable the associated chaos admits a simple orthonormal basis: Hermite polynomials. We develop here some aspects of the Gaussian chaos including some recent developments of

the 4th order moment method which exhibit a rigidity property of the chaos useful to derive simply CLTs in this setting. A nice feature is that for such models all calculations seem to be possible.

The organization of the chapter follows. discretely indexed Gaussian random processes (time series) and the central Brownian motion as well as the attractive fractional Brownian motion are first considered. Now we will get enough tools to get to the convergence of functionals of Gaussian processes.

Limits will involve the Brownian motion as well as the fractional Brownian motion and expansion in the convenient chaoses will make it possible to express limits. Thus expressing a functional in the Brownian chaos is a natural way to express its limit. We thus definitely need to work in Gaussian chaoses (including the 1-d Hermite polynomials).

The method of moments is considered rapidly and in order to calculate all the moments of a random variable belonging to some chaos the Mehler and the diagram formula will be explained.

Final sections will introduce the so called 4-order moment method which proves that in order that a sequence Z_n of random variables belonging to some chaos converges to the Normal standard distribution, this is enough to prove that only $\lim_n \mathbb{E}Z_n = 0, \lim_n \mathbb{E}Z_n^2 = 1$ and $\lim_n \mathbb{E}Z_n^4 = 3$.

5.1 Gaussian processes

Definition 5.1.1. *A Gaussian process (or a Gaussian family) $Y = (Y_t)_{t\in\mathbb{T}}$ is a a collection of random variables defined on a same probability space such that each finite subset defines a Gaussian random vector.*

Remark 5.1.1. *Alternatively, if $(u_t)_{t\in\mathbb{T}}$ is a family of real numbers such that $u_t \equiv 0$ excepted for finitely many t then $\sum_{t\in\mathbb{T}} u_tY_t$ is a Gaussian random variable.*
Or for each finite subset of $T \subset \mathbb{T}$, and if $(u_t)_{t\in T}$ is a finite family of real number then $\sum_{t\in\mathbb{T}} u_tY_t$ is a Gaussian random vector.

We begin with the existence of finite dimensional Gaussian random variables which will be an important support for proving the existence

of Gaussian processes defined later.
If Σ is a $d \times d$ symmetric positive definite, we just proved before the existence of a symmetric positive definite matrix R with $R^2 = \Sigma$.
For $Z = (Z_1, \ldots, Z_d)^t$ independent identically distributed standard normal random variables and for each $m \in \mathbb{R}^d$:

$$Y = m + RZ \sim \mathcal{N}_d(m, \Sigma).$$

As an application:

Proposition 5.1.1. *If a sequence of real numbers $(r_k)_k$ satisfies $r_{-n} = r_n$ for all $n \geq 0$ and*

$$\sum_{i,j=1}^{n} u_i u_j r_{i-j} \geq 0,$$

for all $u_1, \ldots, u_n \in \mathbb{R}$, then there exists a stationary Gaussian process with covariance $r_k = \mathbb{E}X_0 X_k$.

Proof. For each $d \in \mathbb{N}^*$, the law $\mathcal{N}_d(0, \Sigma_d)$ is well defined with $\Sigma_d = (r_{i-j})_{1 \leq i,j \leq d}$.
The Kolmogorov consistency theorem thus asserts the existence of such a process. Recall that this theorem asserts that on may define a distribution on a product set $E^{\mathbb{T}}$ is the projections on finite subset $F \subset \mathbb{T}$ exist (we denote them P_F and are coherent in the sense that for $F' \subset F$ the projections satisfy $P_F \circ \pi_{F,F'}^{-1} = P_{F'}$ where $\pi_{F,F'} : E^F \to E^{F'}$ denotes the projection. ∎

The previous result holds in fact under general conditions.

Theorem 5.1.1. *Let $\Gamma : \mathbb{T}^2 \to \mathbb{R}$ be such that the matrix*

$$(\Gamma(t_i, t_j))_{1 \leq i,j \leq n}$$

satisfies (4.1) for all possible choices $t_i \in \mathbb{T}$, then there exists a Gaussian process with covariance Γ.

An essential example for the study of dependence is detailed below.

5.1.1 Fractional Brownian motion

Definition 5.1.2. *The fractional Brownian motion (fBm see Taqqu paper in [Doukhan et al., 2002b]) with exponent $H \in (0,1]$ is a centered Gaussian process $(Z_t)_{t\in\mathbb{R}}$ with covariance $\Gamma(s,t) = Cov(Z_s, Z_t)$ defined as*

$$\Gamma(s,t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}). \tag{5.1}$$

Proposition 5.1.2. *The function Γ in (5.1) for $s,t \in \mathbb{R}$ is indeed the covariance of a centered Gaussian process $(B_H(t))_{t\in[0,1]}$.*

Proof. From Theorem 5.1.1 we need to prove that for all $0 \le t_1 < \cdots < t_n \le 1$, and $u_1, \ldots, u_n \in \mathbb{C}$

$$A = \sum_{i,j=1}^{n} \Gamma(t_i, t_j) u_i \overline{u_j} \ge 0.$$

- *Step 1.* Set $t_0 = 0$, $u_0 = -\sum_{i=1}^{n} u_i$ then

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} |t_i|^{2H} u_i \overline{u_j} &= -\sum_{i=0}^{n} |t_i|^{2H} u_i \overline{u_0} \\ &= -\sum_{i=0}^{n} |t_i - t_0|^{2H} u_i \overline{u_0}\end{aligned}$$

  Analogously

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |t_j|^{2H} u_i \overline{u_j} = -\sum_{j=0}^{n} |t_j - t_0|^{2H} u_0 \overline{u_j}$$

  hence

$$A = -\sum_{i=0}^{n}\sum_{j=0}^{n} |t_i - t_j|^{2H} u_i \overline{u_j}.$$

- *Step 2.* For $\epsilon > 0$ set

$$B_\epsilon = \sum_{i,j=0}^{n} e^{-\epsilon|t_i - t_j|^{2H}} u_i \overline{u_j}.$$

Then Taylor formula simply implies

$$B_\epsilon \sim \epsilon A, \qquad \epsilon \downarrow 0.$$

- *Step 3.* For each $\epsilon > 0$ and $H \in (0,1]$, there exists a real random variable ξ with

$$\phi_\xi(t) = \mathbb{E}e^{it\xi} = e^{-\epsilon|t|^{2H}}$$

(the law is $2H-$stable); this may be derived from Fourier inversion.
Then

$$B_\epsilon = \mathbb{E}\left|\sum_{j=0}^{n} u_j e^{itj\xi}\right|^2 \geq 0. \qquad \blacksquare$$

Remark 5.1.2. *This process may be defined on* $\mathbb{R}$.
The case $H = \frac{1}{2}$ *yields the Brownian motion* $W = B_{\frac{1}{2}}$ *defined on* $\mathbb{R}^+$.
In this case:

$$\Gamma(s,t) = s \wedge t.$$

Lemma 5.1.1. *Let* $0 \leq h < H$ *then with probability 1, there exist constants* $c, C > 0$ *with*

$$|B_H(s) - B_H(t)| \leq C|t-s|^h \qquad \textit{if} \quad 0 \leq s,t \leq 1, |s-t| < c.$$

Proof. This is a consequence of Kolmogorov-Chentsov 10.1.1 and of the calculation

$$\begin{aligned} 2\mathbb{E}(B_H(s) - B_H(t))^2 & \\ &= |s|^{2H} + |t|^{2H} - \left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right) \\ &= |s-t|^{2H}. \qquad \blacksquare \end{aligned}$$

Remark 5.1.3. *Regularity properties of the fBm are clear from Figures 5.1, 5.3 representing its trajectories respectively for* $H = 0.3$ *and 0.9. while their derivatives are provided in 5.2 5.4: clearly bigger is H and larger is the regularity. From the previous result in the latter case derivatives almost exist while a "real" white noise effect appears for* $H = 0.3$. *Compare with real Gaussian noise in Figure A.2.*

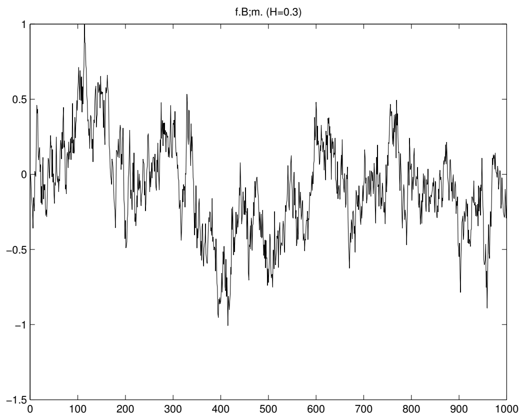


Figure 5.1: fBm $H = 0.3$.

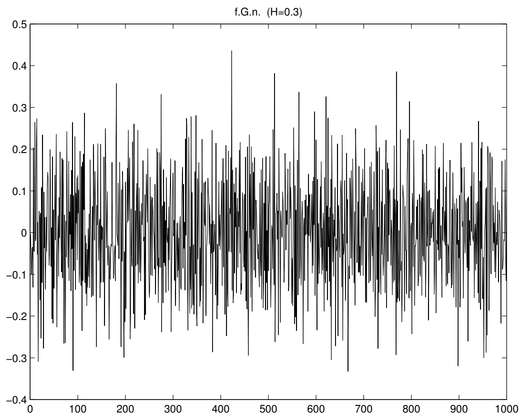


Figure 5.2: Fractional noise $H = 0.3$.

.

Definition 5.1.3. *The process* $(Z(t))_{t\in\mathbb{R}}$ *is* $H-$*self-similar if for all*

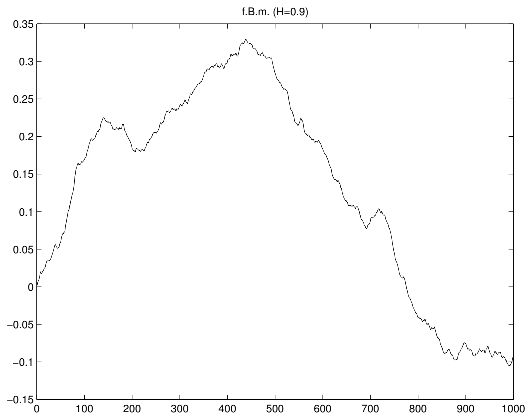


Figure 5.3: fBm $H = 0.9$.

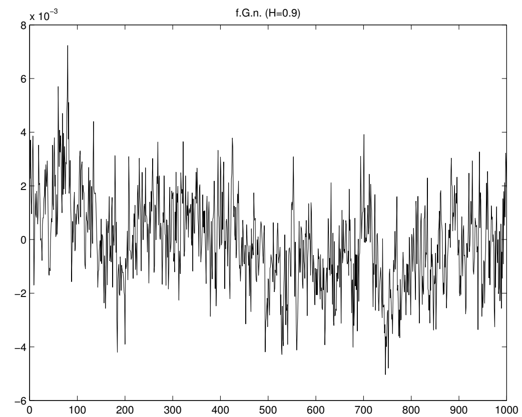


Figure 5.4: Fractional noise $H = 0.9$.

$a > 0$

$$(Z(at))_{t\in\mathbb{R}} = (a^H Z(t))_{t\in\mathbb{R}}, \qquad \textit{in distribution.}$$

We leave the forthcoming point as exercises to a reader:

Proposition 5.1.3. *The condition of $H-$self-similarity is equivalent to the stationarity of the process*

$$(Y(t))_{t>0}, \qquad Y(t) = e^{-tH} Z(e^t)$$

when it is indexed by $\mathbb{R}^+$. For this only check that finite dimensional repartitions of both processes coincide.

Remark 5.1.4. *As $(Y(t))_{t>0}$ is easily proved to be a Gaussian process, strict stationarity and $\mathbb{L}^2-$ (or weak–stationarity) are equivalent: this makes more clear the previous result.*

Remark 5.1.5. *Also quote that:*

1. *If Z is self-similar then $Z(0) = 0$.*
2. *If Z is self-similar and its increments $(Z(t+s) - Z(t))_{t\in\mathbb{R}}$ are stationary for each s then: $\mathbb{E}Z(t) = 0$ if $H \neq 1$ because $\mathbb{E}Z(2t) = 2^H \mathbb{E}Z(t)$ and*

   $$\mathbb{E}(Z(2t) - Z(t)) = \mathbb{E}(Z(t) - Z(0)) = \mathbb{E}Z(t)$$

   implies $(2^H - 2)\mathbb{E}Z(t) = 0$.
3. *If increments of Z are stationary we obtain the equality in distribution $\mathcal{L}(Z(-t)) = -\mathcal{L}(Z(t)$ which follows from the equality of distributions $Z(0) - Z(-t)$ and $Z(t) - Z(0)$.*
4. *From 3) and self-similarity: $\mathbb{E}Z^2(t) = |t|^{2H}$.*
5. *$H \leq 1$ because*

   $$\mathbb{E}|Z(2)| = 2^H \mathbb{E}|Z(1)| \leq \mathbb{E}|Z(2) - Z(1)| + \mathbb{E}|Z(1)| = 2\mathbb{E}|Z(1)|$$

   thus $2^H \leq 2$.
6. *For $H = 1$, $\mathbb{E}Z(s)Z(t) = \sigma^2 st$ thus $\mathbb{E}(Z(t) - tZ(1))^2 = 0$ and the process is degenerated $Z(t) = tZ(1)$.*

Thus

Proposition 5.1.4. *B_H is Gaussian centered and $H-$self-similar with stationary increments.*

Those self similarity properties are not alway present within real data as prove Figures 4.6 and 4.2 which exhibits that stock values of AGF are not at all self similar.

5.2 Gaussian chaos

Linear combinations of Gaussian random variables have been investigated in the above sections. In order to get out of this Gaussian world a first question is set as follows:

how to proceed for multiplying Gaussian random variables?

or equivalently

Do polynomials wrt Gaussian random variables admit a special structure?

Thus polynomials of Gaussian random variables are needed and in order to consider any asymptotic one needs a closed topological vector space. The simplest topology for probability theory is the Euclidean topology of the Hilbert space $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ of the set of classes of squared integrable random variables. This space indeed contains any polynomial expression of a Gaussian vector.
The Gaussian chaos will be shown to be convenient for deriving expressions of any moment expression and thus yields limit theory in this chaos, through Mehler formula and through the diagram formula respectively.
Anyway the diagram formula is extremely complicated to use and we present the so-called 4-th order moment method.
The 4-th moment method is a very powerful technique allowing to prove that sequences of elements in the Gaussian chaos admit a Gaussian asymptotic behavior based on the convergence of the 2 first even order moments.
Namely any element Z in the Gaussian chaos such that $\mathbb{E}Z = 0$ and $\mathbb{E}Z^2 = 1$ is Gaussian if and only if $\mathbb{E}Z^4 = 3$ (and thus belongs to the 1rst order chaos).
This method needs an integral representation of elements of the chaos which we first explain.

Definition 5.2.1. *Let $Y = (Y_t)_{t\in\mathbb{T}}$ be a Gaussian process defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ the Gaussian chaos $\mathcal{C}(Y)$ associated to Y is the smallest complete vector sub-space $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ containing Y_t, $\forall t \in \mathbb{T}$) as well as the constant 1 and which is stable under products (in quicker terms: it is the closure in $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ of the algebra generated by Y).*

Remark 5.2.1. *$\mathcal{C}(Y)$'s elements are thus $\mathbb{L}^2-$limits of polynomials:*

$$Z = \sum_{d=1}^{D} \sum_{t_1 \in T'} \cdots \sum_{t_d \in T'} a^{(d)}_{t_1,\ldots,t_d} Y_{t_1} \cdots Y_{t_d}$$

for some finite subset $T' \subset T$.
This space is thus a Hilbert space. In order to get easy calculations in this space, a basis is first provided in case $\mathbb{T} = \{t_0\}$ is a singleton. Further subsections allow calculations of second order moments and of higher order moments respectively.
Contrarily to what is usual in ergodic theory the denomination of chaoses has nothing to do with some erratic behavior; its origin lies in the tough expression of polynomials with many unknown variables and with the fact that the annulus of such polynomials does not have any of the standard property of such spaces, such as principality or Noether properties; the first of which characterizes ideal sub-rings as generated from products with a fixed polynomial: this property is essential for factorization.

Example 5.2.1 (Hermite expansions)**.**

- *An interesting example of such random variables concerns the case of singletons $T = \{0\}$ is*

  $$Z = g(Y_0), \qquad Y_0 \sim \mathcal{N}(0,1).$$

  If $Z \in L^2$ then we will prove that such expansions exist

  $$Z = \sum_{k=0} \frac{g_k}{k!} H_k(Y_0), \qquad g_k = \mathbb{E} H_k(N) g(N) \quad N \sim \mathcal{N}(0,1),$$

  with H_k some polynomial to be defined below, called Hermite polynomials, see remark 5.2.4.
  Thus Z is also a $\mathbb{L}^2-$limit of polynomials in Y_0.

- *A second case is more suitable for time series analysis $T = \mathbb{Z}$ and $(Y_t)_{t\in\mathbb{Z}}$ is a stationary time serie with $Y_0 \sim \mathcal{N}(0,1)$: on may consider partial sums processes*

  $$Z = g(Y_1) + \cdots + g(Y_n), \qquad \mathbb{E} g^2(Y_0) < \infty.$$

It will be proved that such expressions are again $\mathbb{L}^2-$limits of polynomials and thus belong to the chaos.
A difficult question is to determine the asymptotic behavior of such partial sums. It will be addressed below.

The aim of those notes is to provide a reader with the tools necessary to such considerations.

5.2.1 Hermite polynomials

Definition 5.2.2 (Hermite polynomials)**.** *Let $k \geq 0$ be an arbitrary integer. We set*

$$H_k(x) = \frac{(-1)^k}{\varphi(x)} \frac{d^k \varphi(x)}{dx^k}.$$

Then H_k is a $k-$th degree polynomial with leading term 1.

Thoses polynomials are graphically represented in Figure 5.5.
This last point follows from the relation

$$H_k(x)\varphi(x) = (-1)^k \varphi^{(k)}(x).$$

Through derivation

$$H_k'(x)\varphi(x) + H_k(x)\varphi'(x) = (-1)^k \varphi^{(k+1)}(x)$$

and using $\varphi'(x) = -x\varphi(x)$ we get

$$(H_k'(x) - xH_k(x))\varphi(x) = (-1)^k \varphi^{(k+1)}(x)$$

hence

$$H_{k+1}(x) = xH_k(x) - H_k'(x)$$

Thus $d^\circ H_{k+1} = d^\circ H_k + 1$ admits the same leading coefficient thus $H_0(x) = 1$ concludes.
For example

$$\begin{aligned}
H_0(x) &= 1 \\
H_1(x) &= x \\
H_2(x) &= x^2 - 1 \\
H_3(x) &= x^3 - 3x \\
H_4(x) &= x^4 - 6x^2 + 3 \\
H_5(x) &= x^5 - 10x^3 - 9x
\end{aligned}$$

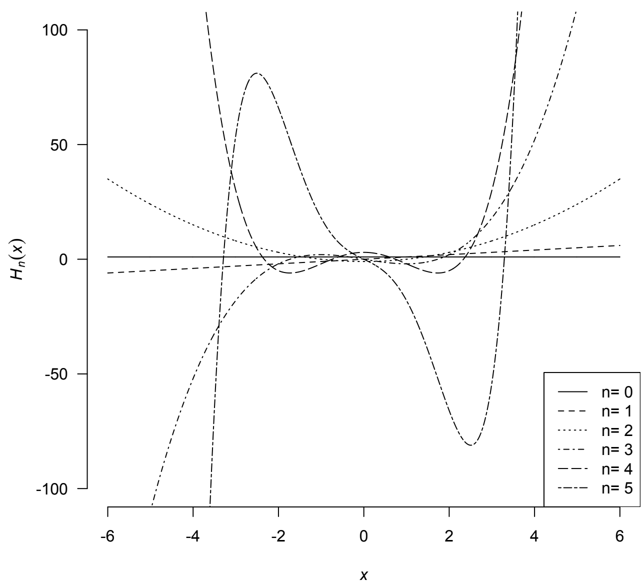


Figure 5.5: Graphs of some Hermite polynomials.

Hermite polynomials $(H_k)_{k\geq 0}$ form an orthogonal system with respect to the Gaussian measure since k integrations by parts yield for $k \geq l$:

$$\begin{aligned}\mathbb{E}H_k(N)H_l(N) &= \int_{-\infty}^{\infty} H_k(x)H_l(x)\varphi(x)dx \\ &= (-1)^k \int_{-\infty}^{\infty} \frac{d^k\varphi(x)}{dx^k} H_l(x)dx \\ &= \int_{-\infty}^{\infty} \frac{d^k H_l(x)}{dx^k}\varphi(x)dx\end{aligned}$$

this expression vanishes if $k > l$. In case $k = l$ this yields

$$\frac{d^k H_k(x)}{dx^k} = k! \qquad \text{hence} \qquad \mathbb{E}H_k^2(N) = k!$$

This system is also complete; we admit this result proved eg. in [Choquet, 1973].
Hence any measurable function g with $\mathbb{E}|g(N)|^2 < \infty$ admits the $\mathbb{L}^2$ representation:

$$g(x) = \sum_{k=0}^{\infty} \frac{g_k}{k!} H_k(x), \ g_k = \mathbb{E}g(N)H_k(N), \ \mathbb{E}|g(N)|^2 = \sum_{k=0}^{\infty} \frac{|g_k|^2}{k!}.$$

Definition 5.2.3. *Assume that $g \in \mathbb{L}^2(\varphi)$ is not the null-function. Define as before $g_k = \mathbb{E}g(N)H_k(N)$.*
The Hermite rank of the function g ($\neq 0$) is the smallest index $k \geq 0$ such $g_k \neq 0$.
We denote m or $m(g)$ his this is Hermite rank.

Proposition 5.2.1. *This orthonormal basis in $\mathbb{L}^2(\varphi(x)dx)$ also satisfies:*

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} H_k(x) = e^{zx - z^2/2}. \tag{5.2}$$

This equality is thus only an equality in the Hilbert space $\mathbb{L}^2(\varphi(x)dx)$. The previous series converges (normally) in $\mathbb{L}^2(\varphi(x)dx)$ because:

$$\mathbb{E}\left(\frac{z^k}{k!} H_k(N) \overline{\frac{z^l}{l!} H_l(N)}\right) = \begin{cases} 0, & \text{if} \quad k \neq l \\ \frac{|z|^{2k}}{k!}, & \text{if} \quad k = l \end{cases}$$

We shall need the following Lemma:

Lemma 5.2.1. $H'_k = kH_{k-1}$.

Proof of Lemma 5.2.1. As $d^\circ(H'_k - kH_{k-1}) < k-1$, this will follow from the relation

$$\int (H'_k(x) - kH_{k-1}(x))H_l(x)\varphi(x)dx = 0 \quad \text{for all} \quad l < k.$$

First

$$k \int H_{k-1}(x)H_l(x)\varphi(x)dx = \begin{cases} 0, & \text{if } l < k-1 \\ k(k-1)! = k!, & \text{if } l = k-1 \end{cases}$$

An integration by parts implies

$$\begin{aligned}\int H_k'(x)H_l(x)\varphi(x)dx &= (-1)^l\int H_k'(x)\varphi^{(l)}(x)dx\\ &= (-1)^{l+1}\int H_k(x)\varphi^{(l+1)}(x)dx\\ &= \int H_k(x)H_{l+1}(x)\varphi(x)dx\end{aligned}$$

This expression vanishes if $l < k-1$.
If now $l = k-1$ we get the same value, $k!$, as for the other quantity which implies $H_k' = kH_{k-1}$. ∎

Remark 5.2.2. *An alternative and more elementary proof of the previous relation begins with the identity:*

$$\varphi'(x) = x\varphi(x).$$

From the definition $\varphi^{(k)}(x) = (-1)^k\varphi(x)$ *hence the previous expression rewrites*

$$H_{k+1}(x) = xH_k(x) - H_k'(x).$$

Derive k times this relation with Leibniz formula, then

$$\varphi^{(k+1)}(x) = -x\varphi^{(k)}(x) - k\varphi^{(k-1)}(x)$$

thus

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x).$$

The formula follows from comparing the two previous expressions of H_{k+1}.

Now the function

$$x \mapsto g_z(x) = e^{zx-z^2/2}$$

belongs to $\mathbb{L}^2(\varphi)$: it admits an Hermite expansion

$$g_z = \sum_k \frac{g_{z,k}}{k!}H_k,$$

this function satisfies:

$$
\begin{aligned}
g_{z,k} &= \mathbb{E}g_z(N)H_k(N)\\
&= \int_{-\infty}^{\infty} H_k(x)e^{zx-z^2/2}\varphi(x)dx\\
&= \int_{-\infty}^{\infty} H_k(x)e^{-(z-x)^2/2}\frac{dx}{\sqrt{2\pi}}\\
&= \int_{-\infty}^{\infty} H_k(t+z)\varphi(t)dt, \text{ from the change in variable } t=x-z\\
&= \sum_{l=0}^{k}\frac{z^l}{l!}\int_{-\infty}^{\infty} H_k^{(l)}(t)\varphi(t)dt, \quad \text{with a Taylor expansion}\\
&= \sum_{l=0}^{k} C_k^l z^l \int_{-\infty}^{\infty} H_{k-l}(t)\varphi(t)dt, \quad \text{because } H_k^{(l)} = \frac{k!}{(k-l)!}H_{k-l}\\
&= z^k
\end{aligned}
$$

We thus get the $\mathbb{L}^2-$expansion:

$$
\sum_{k=0}^{\infty}\frac{z^k}{k!}H_k(N) = e^{zN-z^2/2} \qquad \text{in} \qquad \mathbb{L}^2(\Omega,\mathcal{A},\mathbb{P}). \tag{5.3}
$$

This $\mathbb{L}^2(\varphi)-$convergence also implies the $x-$a.s. convergence of the series

$$
g(x,z) = \sum_{k=0}^{\infty}\frac{z^k}{k!}H_k(x) \qquad \forall z\in\mathbb{C}.
$$

Remark 5.2.3. *If one would knows how to prove that this series $x\mapsto g(x,z)$ converges in a more accurate space and if is derivable for each z, then it would be easy to deduce*

$$
\frac{\partial}{\partial x}g(x,z) = zg(x,z)
$$

and the function $x\mapsto e^{zx-z^2/2}$ satisfies the same partial differential equation.
In both cases $\mathbb{E}g(N,z)=1$ implies (5.2) for all $x\in\mathbb{R}$, $z\in\mathbb{C}$.
A alternative proof of (5.3) would follow: unfortunately we do not know such a result.

Exercise 15 (Orthogonal polynomials)**.** *Assume that* $-\infty \le a < b \le +\infty$. *Let more generally* $p : (a,b) \to \mathbb{R}^+$ *be a measurable function such that* $\lambda(\{x \in (a,b);\ p(x) = 0\}) = 0$.

$$f, g \mapsto (f,g)_p = \int_a^b f(x)g(x)\,p(x)dx$$

defines a scalar product on the pre-Hilbert space $\mathbb{L}^2(p)$ *of classes (wrt to a.s. equality) of measurable functions with*

$$\int_a^b f^2(x)\,p(x)dx < \infty.$$

1. *Suppose that the polynomials* $x \mapsto 1, \dots, x^N$ *satisfy this relation. Analogously to Hermite polynomials one may define recursively a sequence of orthogonal polynomials such that* $P_0 \equiv 1$, $\dots$, $P_n(x) - x^n$ *is orthogonal to* $1, x, \dots, x^{n-1}$, *for* $0 \le n \le N \le +\infty$.

2. *There exist sequences* $a_n \in \mathbb{R}$, $b_n > 0$ *such that*

   $$P_n(x) - (x + a_n)P_{n-1}(x) + b_n P_{n-2}(x), \quad \forall x \in (a,b),\ 2 \le n \le N$$

3. *Roots of orthogonal polynomials. In case* (a,b) *is any closed, open or semi-open interval of* $\mathbb{R}$, *then each orthogonal polynomial admits* n *distinct roots.*

   From now on we consider examples with $N = \infty$.

4. *If* $(a,b) = [-1,1]$ *and* $p(x) = (1-x)^u(1+x)^v$ *we get Jacobi polynomials for* $u, v > -1$. *In case* $u = v = 1$ *one obtains Legendre polynomials and* $u = v = \frac{1}{2}$ *yields Tchebichev polynomials. Prove that* (P_n) *is a complete system.*

5. *If* $(a,b) = [0,+\infty)$ *and* $p(x) = e^{-x}$ *we get Legendre polynomials. Analogously to Hermite case prove that*

   $$P_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}\left(x^n e^{-x}\right).$$

Proofs.

1. Schmidt orthogonalization technique.
   Quote that the fact that the system of orthogonal polynomial is a Hilbertian basis is quite different and needs the proof that

$$f \in \mathbb{L}^2(p): \quad \Big\{\forall n \geq 0,\ (f, P_n)_p = 0\Big\} \Rightarrow f = 0$$

2. As the degree of $P_n(x) - xP_{n-1}(x)$ is $< n-1$ one may write its expansion

$$P_n(x) - xP_{n-1}(x) = c_0 P_0(x) + \cdots + c_{n-2}P_{n-2}(x)$$

   Now $(xP_{n-1}, P_k)_p = (P_{n-1}, xP_k)_p = c_k(P_k, P_k)_p \geq 0$. For $k < n-2$ this entails $c_k = 0$ and if $k+1 = n-1$ this is > 0.

3. Let $x_1 < \cdots < x_k$ the real roots of P_n with a change of sign. Set $Q(x) = (x - x_1)\cdots(x - x_k)$ then $P_n(x)Q(x)p(x) > 0$ (a.s.) which excludes the relation $(P_n, Q)_p = 0$ which holds by construction in case $k < n$.

4. Properties of such polynomials may be found in [Szegö, 1959] or in [Sansone, 1959], and Weierstrass theorem (Exercise 6 for a glance and [Doukhan and Sifre, 2001] for more appropriate comments) entails that those systems are complete.

5. Prove that the leading coefficient of RHS is 1 and that the corresponding system is orthogonal. To this aim again use integrations by parts and due to the fact that integrated terms all vanish we get for $n > k$:

$$\begin{aligned}(P_n, P_k)_p &= \frac{1}{n!}\int_0^\infty P_k(x)\frac{d^n}{dx^n}\big(x^n e^{-x}\big)dx \\ &= \frac{(-1)^n}{n!}\int_0^\infty P_k^{(n)}(x)e^{-x}dx.\end{aligned}$$

∎

5.2.2 Second order moments

Lemma 5.2.2 (Mehler formula)**.** *Let* $Y = (Y_1, Y_2)$ *be a Gaussian random vector with law* $\mathcal{N}_2\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right)$, *then*

$$Cov\ (H_k(Y_1), H_l(Y_2)) = \begin{cases} 0, & \text{if } k \neq l \\ k!r^k, & \text{if } k = l \end{cases}$$

Proof. If $t_1, t_2 \in \mathbb{R}$ set

$$\sigma^2 = \mathrm{Var}\,(t_1Y_1 + t_2Y_2) = t_1^2 + t_2^2 + 2rt_1t_2,$$

then

$$t_1Y_1 + t_2Y_2 \sim \sigma N$$

The relation (A.5) implies thus

$$\mathbb{E}\exp\left(t_1Y_1 + t_2Y_2 - \frac{t_1^2 + t_2^2}{2}\right) = e^{rt_1t_2}$$

From the $\mathbb{L}^2-$identity (5.3) we may exchange integrals and sums from dominated convergence

$$\begin{aligned} \mathbb{E}e^{t_1Y_1+t_2Y_2-\frac{t_1^2+t_2^2}{2}} &= e^{rt_1t_2} \\ &= \sum_{k,l=0}^{\infty} \frac{t_1^k}{k!}\frac{t_2^l}{l!}\,\mathbb{E}H_k(Y_1)H_l(Y_2) \end{aligned}$$

Identify the previous expansion with respect to powers of t_1 and t_2 yields the conclusion since $\mathbb{E}H_k(Y_1) \neq 0$ only for the case $k = 0$. ∎

Remark 5.2.4. *Let* $g : \mathbb{R} \to \mathbb{C}$ *be measurable and* $\mathbb{E}|g(N)|^2 < \infty$ *in the setting of Example 5.2.1. Then*

$$g = \sum_k \frac{g_k}{k!}H_k, \qquad g_k = \mathbb{E}H_k(N)\overline{g(N)}$$

and

$$\begin{aligned} \mathbb{E}g(Y_1)\overline{g(Y_2)} &= \sum_{k=0}^{\infty} \frac{|g_k|^2}{k!}r^k \\ Cov\,(g(Y_1), g(Y_2)) &= \sum_{k=1}^{\infty} \frac{|g_k|^2}{k!}r^k \end{aligned}$$

For $(Y_n)_{n\in\mathbb{Z}}$ a stationary Gaussian process with $\mathbb{E}Y_0 = 0,\ Var Y_0 = 1$ and $r_n = \mathbb{E}Y_0Y_n$.
If $\mathbb{E}g(Y_0) = 0$ (this means that the Hermite rank is such that $m(g) \geq 1$):

$$\begin{aligned}
\mathbb{E}\left|\sum_{j=1}^{n} g(Y_j)\right|^2 &= \sum_{s,t=1}^{n} \mathbb{E}g(Y_s)\overline{g(Y_t)} \\
&= n\sum_{|l|<n}^{n}\left(1-\frac{|l|}{n}\right)\mathbb{E}g(Y_0)\overline{g(Y_l)} \\
&= n\sum_{|l|<n}^{n}\left(1-\frac{|l|}{n}\right)\sum_{k=m(g)}^{\infty}\frac{|g_k|^2}{k!}r_l^k \\
&= n\sum_{k=m(g)}^{\infty}\frac{|g_k|^2}{k!}\sum_{|l|<n}^{n}\left(1-\frac{|l|}{n}\right)r_l^k \qquad (5.4)
\end{aligned}$$

Thus in case $\sum_l |r_l| < \infty$, each series $R_k = \sum_l r_l^k$ converges (for $k \geq 1$) because $|r_l| \leq r_0 = 1$ and

$$\mathbb{E}\left|\sum_{j=1}^{n} g(Y_j)\right|^2 \sim n\sum_{k=m(g)}^{\infty}\frac{R_k|g_k|^2}{k!} = \mathcal{O}(n)$$

If only

$$S \equiv \sum_l |r_l|^{m(g)} < \infty$$

with $m(g)$ the Hermite rank in Definition 5.2.3 then the previous claim still holds true; indeed all series R_k are then convergent for $k \geq m(g)$.
Also remark that Cauchy-Schwarz inequality implies $|r(\ell)| \leq 1 \equiv \mathbb{E}Y_0^2$ thus

$$|r(\ell)|^k \leq |r(\ell)|^{m(g)} \quad \text{if} \quad k \geq m(g).$$

Moreover $|R_k| \leq S$ which proves that the previous expansion (5.4) is indeed convergent.

Example 5.2.2. *For statistics the empirical cumulative distribution is of a first importance*

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1\!\!1_{\{Y_k \le x\}}.$$

$F_n(x)$ is an unbiased estimator of the cumulative function because a simple calculation yields $\mathbb{E}F_n(x) = F(x)$.
The expression of its variance relies on the previous identity for $g(u) = 1\!\!1_{\{u \le x\}}$.
Here

$$\begin{aligned}
g_k &= \mathbb{E}H_k(N)\, 1\!\!1_{\{N \le x\}} \\
&= \int_{-\infty}^{x} H_k(u)\varphi(u)du \\
&= (-1)^k \int_{-\infty}^{x} \varphi^{(k)}(u)du \\
&= \begin{cases} \Phi(x), & (a\ primitive\ of\ \varphi) \quad for \quad k = 0, \\ \varphi(x)H_{k-1}(x), & \qquad\qquad\qquad\quad if \quad k \neq 0. \end{cases}
\end{aligned}$$

Hence

$$Var\ F_n(x) = \frac{1}{n}\sum_{k=m(g)}^{\infty} \frac{|\varphi^{(k-1)}(x)|^2}{k!} \sum_{|l|<n}^{n} \left(1 - \frac{|l|}{n}\right) r_l^k.$$

This expression is

$$Var\ F_n(x) = \mathcal{O}(\frac{1}{n}), \ as\ n \to \infty, \qquad if \quad \sum_l |r_l| < \infty.$$

If now $\sum_l |r_l| = \infty$ its order of magnitude is

$$Var\ F_n(x) = \mathcal{O}\left(\frac{1}{n}\sum_{|l|<n}\left(1 - \frac{|l|}{n}\right) r_l\right)$$

which is more that $\frac{1}{n}$.
Anyway from Ćesaro lemma this expression converges 0 if the sequence r_l does converge to 0.

5.2.3 Higher order moments

The technique used to derive Mehler formula *suggests* an extension for an arbitrary number of factors $H_{l_j}(Y_j)$.
Thus let $Y = (Y_1, \ldots, Y_p) \sim \mathcal{N}_p(0, R)$ for a symmetric matrix $R = (r_{i,j})_{1\le i,j\le p}$ with diagonal entries $r_{i,i} = 1$.
If $(t_1, \ldots, t_p) \in \mathbb{R}^p$ we derive

$$\mathrm{Var}\left(\sum_{j=1}^{p} t_j Y_j\right) = \sum_{j=1}^{p} t_j + 2\rho, \qquad \rho = \sum_{1\le i<j\le p} r_{i,j} t_i t_j.$$

Relation (A.5) proves

$$e^{\rho} = \mathbb{E}\exp\left(\sum_{j=1}^{p}\left(t_j Y_j - \frac{t_j^2}{2}\right)\right).$$

In case an $\mathbb{L}^p-$analogue of (5.3) would exist then

$$\exp\left(\sum_{1\le i<j\le p} r_{i,j} t_i t_j\right) \quad = \quad \mathbb{E}\sum_{l_1=0}^{\infty}\cdots\sum_{l_p=0}^{\infty}\frac{t_1^{l_1}}{l_1!}\cdots\frac{t_p^{l_p}}{l_p!}\mathbb{E}\prod_{j=1}^{p} H_{l_j}(Y_j).$$

Any argument allowing the inversion of sums and integrals would provide the identification of such moments.
Anyways such convergences are not simple and our technique to derive such moments will rely on an argument from [Slepian, 1972].
The characteristic function of the random vector $Y = (Y_1, \ldots, Y_k)$ writes

$$\phi_Y(s) = e^{-\frac{1}{2}s^t\Sigma s},$$

if this is a centered Gaussian vector and its covariance Σ. Then an alternative representation of its density function follows from Fourier inversion.
Assume Σ to be invertible will imply the convergence of the forthcoming integrals:

$$f(y, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{is^t y} e^{-\frac{1}{2}s^t\Sigma s} ds.$$

If $\Sigma = (r_{i,j})_{1\le i,j\le k}$ with $r_{i,i} = 1$ we thus get the heat equation from derivations:

$$\frac{\partial f(y,\Sigma)}{\partial r_{i,j}} = \frac{\partial^2 f(y,\Sigma)}{\partial y_i \partial y_j}, \qquad \text{if} \qquad i \neq j.$$

The function $f(y,\Sigma)$ is analytic wrt the (multidimensional) variable Σ. This will allow the expansion below.
Let $n = (n_{i,j})_{1\le i<j\le k}$ be such that $n_{i,j} \in \mathbb{N}$ for each couple $1 \le i < j \le k$ we denote

$$r^n = \prod_{i<j} r_{i,j}^{n_{i,j}}, \qquad n! = \prod_{i<j} n_{i,j}!.$$

Also set

$$n_{i,j} = n_{j,i}, \quad \text{if } i > j, \quad \text{and} \quad s_{n,i} = \sum_{j\neq i} n_{i,j}$$

then denoting by

$$f(y, I_k) = \prod_{i=1}^{k} \varphi(y_i),$$

we get

$$\begin{aligned}
f(y,\Sigma) &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \frac{\partial^{(\sum_{i<j} n_{i,j})} f(y,I_k)}{\prod_{i<j} \partial r_{i,j}^{n_{i,j}}} \\
&= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \frac{\partial^{s_{n,i}} f(y,I_k)}{\prod_{i<j} \partial y_i^{n_{i,j}} \partial y_j^{n_{i,j}}} \\
&= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^{k} \frac{\partial^{s_{n,i}} \varphi(y_i)}{\partial y_i^{s_{n,i}}} \\
&= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^{k} \varphi^{(s_{n,i})}(y_i) \\
f(y,\Sigma) &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^{k} H_{s_{n,i}}(y_i) \cdot \phi(y) \qquad (5.5)
\end{aligned}$$

where

$$\phi(y)=\prod_{i=1}^k \varphi(y_i)$$

denotes the density function of a random vector $\mathcal{N}_k(0,I_k)$ and the previous sums extend to all integer multi-indices

$$n=(n_{i,j})_{1\le i<j\le k}.$$

Indeed $s_{n,i}$ is the number of apparitions for y_i in the second identity. Relation (5.5) thus implies

$$\mathbb{E}\prod_{i=1}^k H_{s_i}(Y_i)=\sum_n \frac{r^n}{n!}\prod_{i=1}^k\int_{-\infty}^{\infty} H_{s_{n,i}}(y_i)H_{s_i}(y_i)\varphi(y_i)dy_i,$$

and orthogonality of the Hermite polynomials implies

$$\mathbb{E}\prod_{i=1}^k H_{s_i}(Y_i)=s_1!\cdots s_k!\sum_{n\in N(s_1,\dots,s_k)}\frac{r^n}{n!},$$

for sums

$$\sum_{n\in N(s_1,\dots,s_k)}$$

extended to such multi-indices n with

$$s_{n,i}=s_i \qquad \text{if} \qquad 1\le i\le k.$$

If $k=2$ the sum in n is a simple sum on the set of integers $\mathbb{N}$ because $i<j$ implies $i=1$ and $j=2$.
Thus

$$\sum_{n\in N(s_1,s_2)}$$

corresponds the only value $n_{1,2}=s_1=s_2$: this is again Mehler formula.
For $k\ge 2$ the previous formula is called the diagram formula.
The $n_{i,j}$'s correspond to partitions of the array such that

$$\begin{array}{llll} x_1 & \dots & x_1 & \text{appears } s_1 \text{ times},\\ x_2 & \dots & x_2 & \text{appears } s_2 \text{ times},\\ \dots & \dots & \dots & \qquad\dots\\ x_k & \dots & x_k & \text{appears } s_k \text{ times}. \end{array}$$

Precisely the first line the arrays may be divided into $k-1$ parts with respective sizes $n_{1,2}, \ldots, n_{1,k}$.
The number of such multi-indices is also the number of arrays satisfying the constraints $s_{n,i} = s_i$. ∎

Remark 5.2.5. *Various uses of this formula are known.*
[Breuer and Major, 1983] prove that if a stationary Gaussian process satisfies $Y_0 \sim \mathcal{N}(0,1)$ and

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} g(Y_k) \xrightarrow{in\ law} \mathcal{N}(0, \sigma^2)$$

if

$$\sum_k |r_k|^m < \infty, \qquad r(k) = \mathbb{E}Y_0 Y_k$$

and $m = (g)$ denotes the Hermite rank of g.
For this those authors prove the convergence of moments of S_n to the Gaussian ones with the diagram formula.
Another application is Arcones inequality for vector valued processes, see Taqqu in [Doukhan et al., 2002b].
This inequality is extended in [Soulier, 2001] and further by Bardet and Surgailis.
Other developments are also reported in [Rosenblatt, 1985].

The forthcoming 4-th order moment approach allows an impressive simplification of the necessary calculations.
This is the justification of the two forthcoming subsections.

5.2.4 Integral representation of the Brownian chaos

Consider a square integrable function $f : \mathbb{R}^+ \to \mathbb{R}$.
We already defined Wiener integrals

$$I_1(f) = \int_0^\infty f(t) dW(t)$$

as centered Gaussian random variables, in the corresponding Gaussian closed space generated by the Brownian process $(W(s))_{s \geq 0}$, with

variance
$$\|f\|_2 = \left(\int_0^\infty f^2(t)dt\right)^{\frac{1}{2}}.$$

We aim at defining it first for step functions and noticing that for such functions $f \mapsto I_1(f)$ is an isometry $\mathbb{L}^2(\mathbb{R}^+) \to \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$,

$$\|f\|_2 = \left(\int_0^\infty f^2(t)dt\right)^{\frac{1}{2}} = \left(\mathbb{E}I_1(f)^2\right)^{\frac{1}{2}} = \|I_1(f)\|_{\mathbb{L}^2(\Omega,\mathcal{A},\mathbb{P})}.$$

We extend it by using a topological argument (density).
A first simple extension is to define stochastic integrals on the real line.
Consider two independent Brownian motions W_- and W_+ a way to define the Brownian motion on the line is to set $W(t) = W_+(t)$ if $t \geq 0$ and $W(t) = W_-(-t)$ if $t < 0$.
Wiener integral is straightforwardly extended on $(-\infty, \infty)$.
There exist two different ways to define

$$I_k(h) = \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty h(t_1, \ldots, t_k)\, dW(t_1) \cdots dW(t_k)$$

We denote by $\mathcal{H}_k$ the set of symmetric functions $h \in L^2(\mathbb{R}^k)$, *i.e.* such that for any arbitrary bijection (permutation) $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$:
$$h(t_{\pi(1)}, \ldots, t_{\pi(k)}) = h(t_1, \ldots, t_k).$$

We will better use the symmetrized version of a function $h \in L^2(\mathbb{R}^k)$ by setting:

$$\mathrm{Sym}(h)(t_1, \ldots, t_k) = \frac{1}{k!} \sum_\pi h(t_{\pi(1)}, \ldots, t_{\pi(k)}).$$

Those spaces are naturally equipped with their natural Hilbert norms

$$\|h\|^2_{\mathcal{H}_k} = \int_{\mathbb{R}^k} h^2(t)dt_1 \cdots dt_k,$$

and then the triangle inequality entails

$$\|\mathrm{Sym}(h)\|_{\mathcal{H}_k} \leq \|h\|_{\mathcal{H}_k}.$$

We defer a reader to [Major, 1981] for more precise statements and we only provide a very fast overview.
Questions of convergence are extremely specific and technically difficult in this frame as quoted in the forthcoming Chapter concerned with dependence.

- A first way is to simply set it by recursion but in this case stochastic integrals to be considered are non anticipative.
  An alternative way to proceed is to consider integrals over sets
  $$\{(t_1,\ldots,t_k)\in\mathbb{R}^k,\quad t_1\le\cdots\le t_k\}$$
  then if the function is invariant through permutations one defines
  $$I_k(h)=k!\int_{-\infty}^{\infty}\int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_{k-1}}h(t_1,\ldots,t_k)\,dW(t_1)\cdots dW(t_k)$$
- Assume now that h is a symmetric function with
  $$h(\pm t_1,\ldots,\pm t_k)=h(t_1,\ldots,t_k).$$
  An alternative construction in [Major, 1981] is based again on an approximation by step functions.
  First if $A_1,\ldots,A_k\subset\mathbb{R}^+$ are closed intervals, set $\Delta_j=A_j\cup(-A_j)$ and $\Delta=\Delta_1\times\cdots\times\Delta_k$, then we define
  $$I_k(\mathbb{I}_\Delta)=L_1\cdots L_k,\quad\text{with}\quad L_j=W_+(A_j)-W_-(A_j).$$
  If $A_1,\ldots,A_k$ are pairwise disjoint then those random variables are independent.
  This definition is extended by linearity to functions constant on such intervals Δ.
  A uniform continuity argument is thus used to define such multiple integrals for $h\in\mathcal{H}_k$. Namely this integral is an isometry over simple functions it thus extends to the closure $\mathcal{H}_k$ of this set.

Exercise 16. *Prove that:*
$$H_k(I_1(f))=I_k(f^{\otimes k}),$$

with I_k the $k-th$ Ito -Wiener integral and

$$f^{\otimes k}(t_1, \ldots, t_k) = f(t_1) \cdots f(t_k).$$

For example this formula is just Ito formula for $k = 2$.

Hint. As for the construction of multiple Ito integrals, first proceed with simple indicator functions, and then extend it linearly to piecewise constant functions. Conclude with the previous extension argument. ∎

5.2.5 The fourth order moment method

The method of diagrams is now less used because of the recents developments by Peccati and coauthors as eg. [Nourdin et al., 2011] ([1]). In order to simplify expressions we consider the chaos generated by

$$\{W(t), t \geq 0\}.$$

From now on we restrict to functions on the interval $[0, 1]$ and we keep using the same notations as above. For $f \in \mathcal{H}_k$ and $g \in \mathcal{H}_m$, for $1 \leq p \leq k \wedge m$, define with [Nourdin et al., 2011]

$$\begin{aligned} f \otimes_p g(t_1, \ldots, t_{m+k-2p}) \\ &= \int_{\mathbb{R}^p} f(t_1, \ldots, t_{k-p}, s_1, \ldots, s_p) \\ &\quad \times g(t_{k-p+1}, \ldots, t_{k+m-2p}, s_1, \ldots, s_p)\, ds_1 \cdots ds_p \end{aligned}$$

For example if $m = 0$ or k we have respectively:

$$f \otimes_0 g = f \otimes g, \qquad f \otimes_k g = \int_{\mathbb{R}^k} f(s)g(s)\, ds.$$

Ito formula writes in this case as the forthcoming formula and the other 2 formulas are also useful:

$$I_k(f)I_m(g) = \sum_{p=0}^{k\wedge m} p! \binom{k}{p} \binom{m}{p} I_{k+m-2p}(f \otimes_p g).$$

[1] Many thanks to Ivan Nourdin for his essential and friendly help.

$$\frac{(k+m)!}{k!m!}\|\mathrm{Sym}(f\otimes g)\|^2_{\mathcal{H}_{k+m}} = \|f\|^2_{\mathcal{H}_k}\|g\|^2_{\mathcal{H}_m} + \sum_{q=1}^{k\wedge m}\binom{k}{q}\binom{m}{q}\|f\otimes_q g)\|^2_{\mathcal{H}_{k+m-2q}}.$$

Now the 4-th order moments may also be calculated:

$$\mathbb{E}I_k^4(f) = 3k!^2\|f\|^4_{\mathcal{H}_k} + \frac{3}{k}\sum_{p=1}^{k-1} p\cdot p!\binom{k}{p}^4\big(2(k-p)\big)!\|\mathrm{Sym}(f\otimes_p f)\|^2_{\mathcal{H}_{2(k-p)}}.$$

In particular, observe from the above representation that

$$\lim_{n\to\infty}\big(\mathbb{E}I_k^4(f_n) - 3(\mathbb{E}I_k^2(f_n))^2\big) = 0,$$

is equivalent to

$$\lim_{n\to\infty}\|\mathrm{Sym}(f_n\otimes_p f_n)\|^2_{\mathcal{H}_{2(k-p)}} = 0, \qquad \forall p\in\{1,\dots,k-1\}.$$

We now present the most impressive rigidity result from this **Nualart-Peccati-Tudor** theory.

Theorem 5.2.1. *Assume that a sequence* $f_n\in\mathcal{H}_k$ *satisfies*

$$\lim_n \|f_n\|_{\mathcal{H}_k} = 1,$$

then

$$I_k(f_n)\overset{in\ law}{\longrightarrow}_{n\to\infty}\mathcal{N}(0,1) \iff \lim_n \mathbb{E}I_k^4(f_n) = 3.$$

Proof (thanks to Ivan Nourdin). In fact this will be enough to prove the result if, only $\mathbb{E}I_k^2(f_n)=1$ and $\lim_n \mathbb{E}I_k^4(f_n)=3$.
In order to prove the result two additional tools will be needed

1. $\mathbb{E}I_k(f)\psi(I_k(f)) = kE\psi'(f)\int_{-\infty}^{\infty} I_{k-1}^2(f(\cdot,t))\,dt$ for each function $\psi:\mathbb{R}\to\mathbb{R}$ in C_b^1.

2. $$\begin{aligned}&\mathrm{Var}\int_{-\infty}^{\infty} I_{k-1}^2(f(\cdot,t))\,dt\\ &\quad = \frac{1}{k^4}\sum_{p=1}^{k-1} p(p!)^2\binom{k}{p}^4\big(2(k-p)\big)!\|\mathrm{Sym}(f\otimes_p f\|^2_{\mathcal{H}_{2(k-p)}}\end{aligned}$$

This entails in particular

$$\lim_{n\to\infty} \mathrm{Var} \int_{-\infty}^{\infty} I_{k-1}^2(f_n(\cdot,t))\,dt = 0$$
$$\Leftrightarrow \lim_{n\to\infty} \|\mathrm{Sym}(f_n \otimes_p f_n)\|^2_{\mathcal{H}_{2(k-p)}} = 0, (1 \le p < k)$$
$$\Leftrightarrow \lim_{n\to\infty} \mathbb{E} I_k^4(f_n) = 3.$$

Now set

$$\psi_n(t) = e^{\frac{t^2}{2}}\ \mathbb{E}(\exp(itI_k(f_n))),$$

then

$$\begin{aligned}
\psi_n'(t) &= t\psi_n(t) + ie^{\frac{t^2}{2}}\mathbb{E}(I_k(f_n)\exp(itI_k(f_n))) \\
&= te^{\frac{t^2}{2}}\mathbb{E}\left(1 - \int_{-\infty}^{\infty} I_{k-1}^2 f_n(\cdot,t)\ dt\right) I_k(f_n)\exp(itI_k(f_n))
\end{aligned}$$

Then :

$$|\psi_n'(t)| \le te^{\frac{t^2}{2}}\mathbb{E}\left|1 - \int_{-\infty}^{\infty} I_{k-1}^2 f_n(\cdot,t)\ dt\right|.$$

Quote that

$$\mathbb{E}\int_{-\infty}^{\infty} I_{k-1}^2 f_n(\cdot,t)dt = 1,$$

then from Cauchy-Schwarz inequality we need to control the variance of

$$\int_{-\infty}^{\infty} I_{k-1}^2(f_n(\cdot,t))\ dt$$

which tends to 0 from the above equivalence.
We thus just proved that the sequence of the characteristic functions of $I_k(f_n)$ converge to that of a standard Gaussian. ∎

Chapter 6

Linear processes

Stationary sequences generated through independent identically distributed $(\xi_n)_{n\in\mathbb{Z}}$ are considered hereafter.
Such models are natural in signal theory since they appear through linear filtering of a white noise ([1]).

6.1 Stationary linear models

Definition 6.1.1. *Let $(c_n)_{n\in\mathbb{Z}}$ a sequence of real numbers, and $(\xi_n)_{n\in\mathbb{Z}}$ be an iid sequence. we define linear processes (when it makes sense) as:*

$$X_n = \sum_{k=-\infty}^{\infty} c_k \xi_{n-k}. \tag{6.1}$$

Lemma 6.1.1. *Relation*

$$\sum_k |c_k|^m < \infty$$

implies that the previous series converge if $\mathbb{E}|\xi_0|^m < \infty$ for some $m \in (0,1]$ (if $\mathbb{E}|\xi_0| < \infty$ then $m=1$), then this series converges in probability.

[1] Usually this is only a $\mathbb{L}^2-$stationary white noise sequence and not an independent identically distributed sequence.

Proof. From Markov inequality we derive:

$$\mathbb{P}\left(\sum_k |c_{n-k}||\xi_k| > A\right)$$
$$\leq \frac{1}{A^m}\mathbb{E}\left(\sum_k |c_k||\xi_{n-k}|\right)^m$$
$$\leq \frac{1}{A^m}\mathbb{E}|\xi_|^m \sum_k |c_k|^m.$$

Use the following exercise:

Exercise 17. *Let $a, b \geq 1$*

1. *Prove relation $(a+b)^m \leq a^m + b^m$ if $0 \leq m \leq 1$.*
2. *Prove relation $(a+b)^m \leq 2^{m-1}(a^m + b^m)$ if $m \geq 1$.*

Hints. Use the fact that $g'(t) = m(1+t)^{m-1} \leq 1$ for $t > 0$, if $g(t) = (1+t)^m$ for $m \leq 1$.
The function $h(x) = x^m$ is convex in case $m \geq 1$ and the inequality follows with convexity inequality with equal weights

$$\lambda = \mu = \frac{1}{2}: \qquad h\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\Big(h(a) + h(b)\Big).$$

The sequence (ξ_n) considered is zero mean in case $m \geq 1$ and we assume excepted if we explicitly mention it that this is an independent sequence in order to serive strict stationarity assumptions.
If $\mathbb{E}\xi_0^2 < \infty$ $(m = 2)$ and $\mathbb{E}\xi_0 = 0$, the weaker condition for stationarity and existence in $\mathbb{L}^2$ holds

$$\sum_k |c_k|^2 < \infty.$$

Definition 6.1.2. *If $c_k = 0$ for $k < 0$ the process (6.1) is causal.*

Assume here that (ξ_n) is a $\mathbb{L}^2$ white noise. This process admits the covariance:

$$r_k = \operatorname{Cov}(X_0, X_k) = \sum_l c_l c_{l+k} = c \star \widetilde{c}_k, \tag{6.2}$$

denoting $\widetilde{c} = (\widetilde{c}_k)_{k\in\mathbb{Z}}$ with $\widetilde{c}_k = c_{-k}$.

Remark that
$$\sum_k |r_k| \le \left(\sum_k |c_k|\right)^2,$$
thus this series converges in case
$$\sum_k |c_k|. < \infty$$

We thus obtain:

Proposition 6.1.1. *Let (X_t) be a linear process defined from eqn. (6.1) (with iid inputs ξ_n) then the above series converge a.s., this process is stationary and in $\mathbb{L}^m$ in case either*
$$\mathbb{E}|\xi_0|^m < \infty, \qquad \sum_k |c_k|^m < \infty, \qquad 0 < m \le 1,$$
or it is causal and,
$$\mathbb{E}|\xi_0|^2 < \infty, \qquad \sum_k |c_k|^2 < \infty, \qquad m = 2.$$
In the latter case the covariance of the process writes as in eqn. (6.2). The series of covariances converges if
$$\sum_{k=0}^{\infty} |c_k| < \infty.$$

Definition 6.1.3. *The Backward or shift operatorindexShift operator B ([2]) is defined for sequences*
$$x = (x_n)_{n\in\mathbb{Z}} \mapsto Bx : \qquad (Bx)_n = x_{n-1}, \quad \forall n \in \mathbb{Z}.$$

Remark 6.1.1. *Moreover notational conventions are $Bx = (Bx_n)_{n\in\mathbb{Z}}$ or equivalently $Bx_n = x_{n-1}$, eg. for any discrete time stochastic process we set:*
$$BX_t = X_{t-1}, \qquad t \in \mathbb{Z}.$$

[2] In the econometric literature this operator is also denoted by L, as the Lag-operator.

Using the backward operator B the previous causal models also write

$$X = g(B)\xi, \qquad g(z) = \sum_{k=0}^{\infty} c_k z^k, \qquad |z| < 1.$$

We now rapidly describe some such very simple models of a constant use in statistics.
Clearly this chapter has no statistical ambition but we shall simply rephrase some of the currently used models.

Remark 6.1.2 (centering)**.** *In case $(X_t)_{t\in\mathbb{Z}}$ is not a centered process, $m \equiv \mathbb{E}X_0$ may be estimated empirically by*

$$\widehat{m} = \frac{1}{2n+1}\sum_{k=-n}^{n} X_k$$

the estimation is consistent from the ergodic theorem (Corollary 9.1.3) in case the process $(X_t)_{t\in\mathbb{Z}}$ is indeed ergodic.
Assume now that the process is observed on the period $\{1,\ldots,N\}$ and there exists a continuous function and a centered stationary linear process such that

$$X_t = m\left(\frac{t}{N}\right) + Y_t, \qquad t = 1,2,\ldots,N.$$

In this case a local mean may be used the function m is fitted by

$$\widehat{m}(x) = \frac{1}{2n+1}\sum_{k=-n}^{n} X_{[Nx]+k},$$

and for $n \equiv n(N)$ such that $\lim_{N\to\infty} N/n(N) = 0$ this estimation is consistent.
More generally, a smoothing technique as in eqn. (3.4) may also be used.

6.2 ARMA$(p,q)-$processes

Relation

$$X_t - \sum_{j=1}^{p} a_j X_{t-j} = \xi_t - \sum_{k=1}^{q} b_k \xi_{t-k}, \tag{6.3}$$

is written

$$\alpha(B)X_t = \beta(B)\xi_t$$

Trajectories of such ARMA models are reported in Figure 6.1.

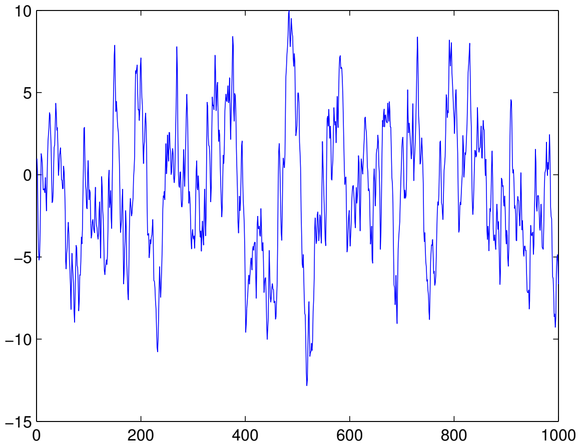


Figure 6.1: ARMA(1,1)–trajectory.

Definition 6.2.1 (ARMA-processes)**.** *Consider the recursion (6.3). If the roots $r_1, \ldots, r_p$ of the polynomial α are such that*

$$|r_1| > 1, \ldots, |r_p| > 1$$

Then:

$$\forall k \in \mathbb{Z} : |r_k| \leq c\rho^{|k|} \qquad \textit{for some} \qquad 0 \leq \rho < 1.$$

Moreover, the series (6.3) converge in $\mathbb{L}^p$ in case inputs $\xi_j \in \mathbb{L}^p$ for some $p > 0$.

For the proposed ARMA(1,1)–process note that it looks really non-deterministic and its covariances decay extremely fast.

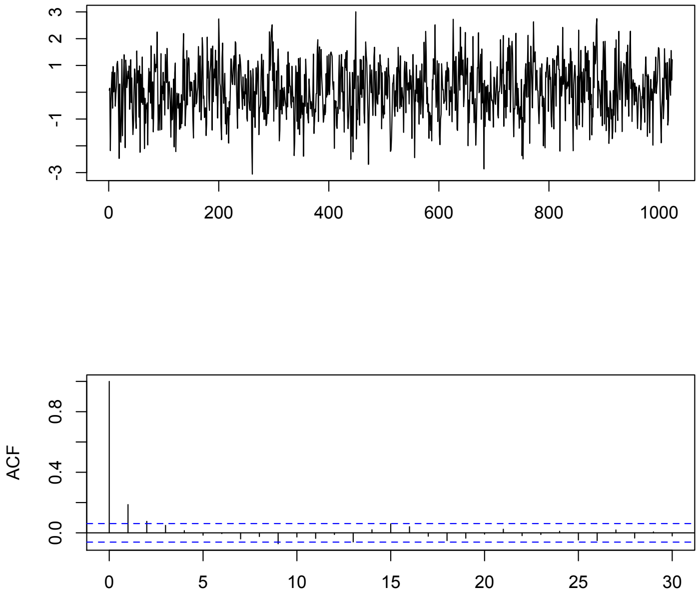


Figure 6.2: An ARMA(1,1)–model

Sketch of the proof. A solution of eqn. (6.3) is written:

$$X_t = \sum_{j=0}^{\infty} c_j \xi_{t-j}$$

where c_j's are defined from

$$\sum_{j=0}^{\infty} c_j z^j = \frac{\beta(z)}{\alpha(z)},$$

with

$$\alpha(z) = 1 - a_1 z - \cdots - a_p z^p = \left(1 - \frac{z}{r_1}\right) \cdots \left(1 - \frac{z}{r_p}\right).$$

If the roots $r_1, \ldots, r_p$ of the polynomial α are such that

$$|r_1| > 1, \ldots, |r_p| > 1$$

then the function $1/\alpha$ is analytic if

$$|z| < \min\{|r_1|, \ldots, |r_p|\}$$

and thus on a neighborhood of the closed complex unit disk. For example

$$\left(1 - \frac{z}{r_1}\right)^{-1} = \sum_{l=0}^{\infty} r_1^{-l} z^l.$$

Moreover the analycity of the function β/α on some disk $D(0, 1+\epsilon)$ implies $|c_k| \le Ce^{-\gamma k}$.

6.3 Yule-Walker equations

This section aims at providing a very fact approach to Yule-Walker equations yielding parametric estimation for ARMA models, we defer a reader to the remarkable textbook of Peter Brockwell and Richard Davis (2006), [Brockwell and Davis, 1991].
For simplicity we restrict to AR(p) models where $(\xi_t)_{t\in\mathbb{Z}}$ denotes an iid sequence centered and with $\sigma^2 = \mathbb{E}\xi_0^2$, as before

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \xi_t \tag{6.4}$$

we again assume that

$$\alpha(z) = 1 - a_1 z - \cdots - a_p z^p = \prod_{j=1}^{p} \left(1 - \frac{z}{r_j}\right)$$

admits roots $|r_j| > 1$ for $j = 1, \ldots, p$. Then we just proved that a MA(∞)–expansion indeed holds:

$$X_t = \sum_{j=0}^{\infty} c_j \xi_{t-j}$$

Parameters of interest in this model are $\theta = (a, \sigma^2)$ with $a^t = (a_1, \ldots, a_p)$. In case the inputs are iid Gaussian $\mathcal{N}(0, \sigma^2)$ those are the only parameters. We aim at estimating those parameters.
Multiply equation (6.4) by X_{t-j} for $0 \le j \le p$ then taking expectations entails

$$\begin{array}{rcl} R_p a &=& \mathbf{r}_p, \qquad R_p = (r_{i-j})_{1 \le i,j \le p}, \quad \mathbf{r}'_p = (r_0, \ldots, r_p) \\ \sigma^2 &=& r_0 - a'\mathbf{r}_p \end{array}$$

By plugging-in estimates $\widehat{r}_j$ of covariances r_j as in eqn. (9.2) provides us with empirical estimates of the parameters.
First this is easy to define $\widehat{R}_p = (\widehat{r}_{i-j})_{1 \le i,j \le p}$ and $\widehat{\mathbf{r}}_p$ and thus

$$\begin{array}{rcl} \widehat{R}_p \widehat{a} &=& \widehat{\mathbf{r}}_p \\ \widehat{\sigma}^2 &=& \widehat{r}_0 - \widehat{a}' \widehat{\mathbf{r}}_p \end{array}$$

Remark 6.3.1 (ARMA case)**.** *Those equations also extend for ARMA models, but besides the previous estimates CLT results are optimal under AR–modeling, see again [Brockwell and Davis, 1991] Chapter 8.*

Remark 6.3.2 (nonlinear models)**.** *Extensions to the case of weak-white noise are used; for example nonlinear models such as ARCH models are such white noises and a linear process with such input may also be considered. In the forthcoming chapter we describe some elementary versions of this idea.*

Remark 6.3.3 (Durbin-Levinson algorithm)**.**
From such estimation a plug-in one-step ahead prediction of the process writes:

$$\widehat{X}_t = \widehat{a}_1 X_{t-1} + \cdots + \widehat{a}_p X_{t-p},$$

once the parameters have been estimated from the data $X_0, \ldots, X_{t-1}$. 2-steps ahead predictions are similar by replacing now X_t by $\widehat{X}_t$ in the previous relation and:

$$\widehat{X}_{t+1} = \widehat{a}_1 \widehat{X}_t + \widehat{a}_2 X_{t-1} + \cdots + \widehat{a}_p X_{t-p+1}.$$

Now we may replace the covariances by their empirical counterparts (see [Brockwell and Davis, 1991]. § 8.2).

6.4 FARIMA$(0, d, 0)-$processes

Set $\Delta = I - B$ with B the Backward operator. The operator Δ allows to rewrite the previous models but it also helps to define some new models. We aim at solving the formal equation

$$\Delta^d X_t = \xi_t.$$

In case $d = 1$ the equation writes $X_t - X_{t-1} = \xi_t$ thus

$$X_t = X_0 + \xi_1 + \cdots + \xi_t,$$

which is a random walk if $X_0 = 0$.
If $d = 2$ the relation still writes

$$\Delta^2 X_t = \Delta(\Delta X_t) = \xi_t,$$

which leads to a recursive definition with initial condition 0 for the solution of equation

$$\Delta^d X_t = \xi_t \qquad \text{for} \quad d \in \mathbb{N}.$$

If $d \in -\mathbb{N}$ the relation writes

$$X_t = \Delta^{-d}\xi_t = \sum_{j=0}^{-d} C_{-d}^j \xi_{t-j}.$$

More generally the relation $X_t = (I - B)^{-d}\xi_t$ is interpreted as an expansion for $|z| < 1$ of the function

$$(1-z)^{-d} = \sum_{j=0}^{\infty} b_j z^j,$$

then:

$$b_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \frac{1}{\Gamma(d)} \prod_{k=1}^{j} \frac{k-1+d}{k}. \tag{6.5}$$

Stirling formula (see eg. [Doukhan and Sifre, 2001]):

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \qquad n \to \infty,$$

implies indeed $b_j \sim \frac{1}{\Gamma(d)} j^{d-1}$ as $j \to \infty$.
For $-\frac{1}{2} < d < \frac{1}{2}$, define the operators $\Delta^{\pm d}$. In order to Δ^d out of this range use relations $\Delta^{d+1} = \Delta\Delta^d$ and

$$\Delta^{d-1} X_t = \xi_t \Rightarrow \Delta^d X_t = \Delta\xi_t = \xi_t - \xi_{t-1}.$$

The evolution of trajectories of such FARIMA$(0, d, 0)$ is designed in Figure 6.3.
Clearly the smallest values of $d = .01$ yields a white noise behavior and the trajectories look more and more regular as $d < .5$ becomes larger.
The corresponding covariograms (listing covariance estimates for such models) is clear from Figure 6.4 confirms the impression provided by trajectories of such FARIMA models. The evolution of covariances which are those of white noise again for $d = .01$ and then seem more and more cyclical for larger values of d.

Definition 6.4.1. *FARIMA*$(0, d, 0)$ *are linear causal processes given by coefficients given from eqn. (6.5).*
Hence

$$\sum_j b_j^2 < \infty, \qquad \textit{if} \qquad d < \frac{1}{2},$$

and the series

$$X_t = \sum_{j=0}^{\infty} b_j \xi_{t-j} \qquad \textit{converge in } \mathbb{L}^2.$$

Moreover:

$$\begin{aligned} r(0) &= \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}, \qquad\qquad \sigma^2 = \mathbb{E}\xi_0^2, \\ r(k) &= \sigma^2 \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k-d+1)} \sim \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} |k|^{2d-1} \quad \textit{as } |k| \to \infty. \end{aligned}$$

Thus

$$\sum_k |r(k)| = \infty \iff d \in \left(0, \frac{1}{2}\right).$$

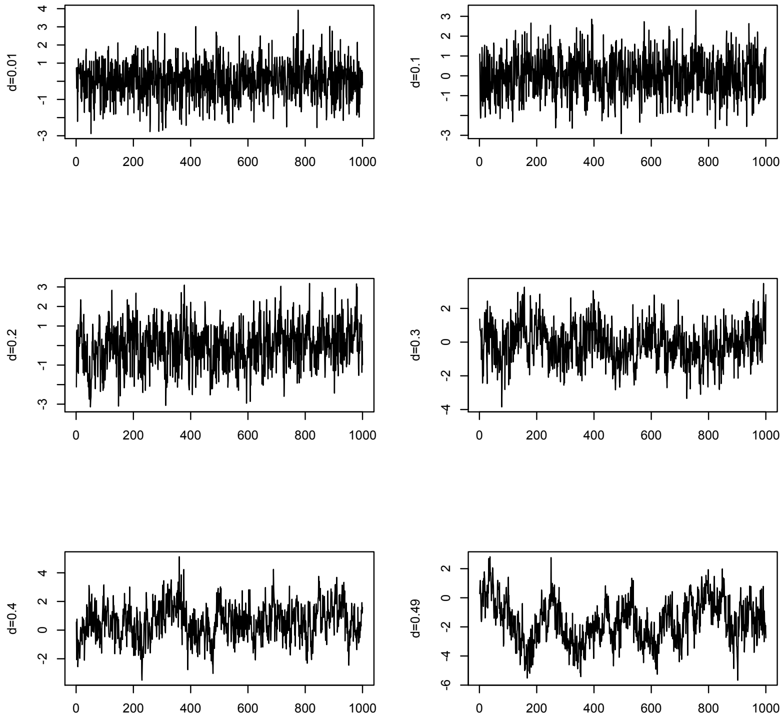


Figure 6.3: Trajectories of FARIMA$(0, d, 0)$.

Remark 6.4.1. *The Hurst coefficient* $H = d + \frac{1}{2}$ *is designed to parameterize those models. They were introduced to models river flooding.*

Set Z for the random spectral measure associated to the white noise ξ_t.

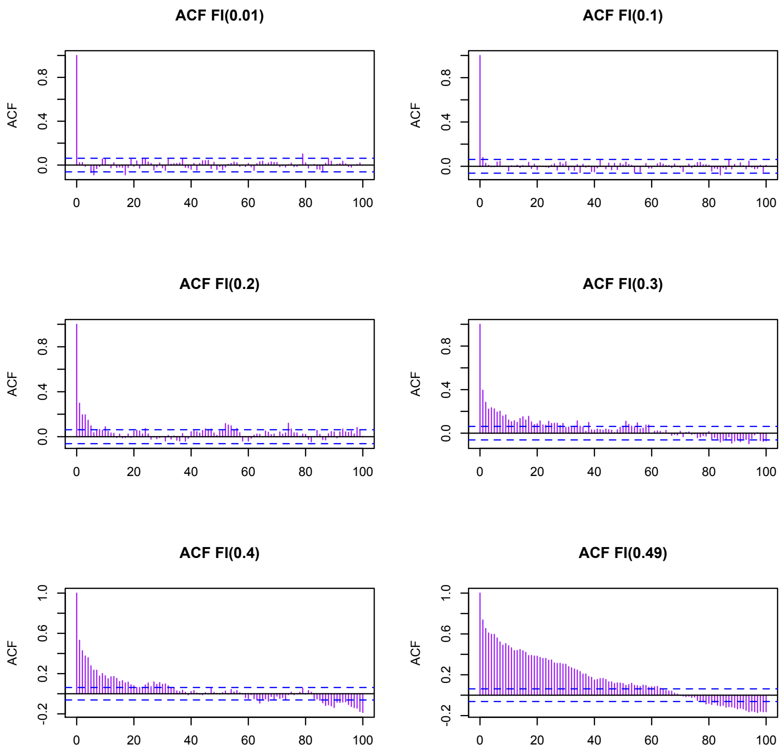


Figure 6.4: Covariograms of FARIMA$(0, d, 0)$.

Then

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}(1 - e^{-i\lambda})^{-d} Z(d\lambda)$$

and

$$g_X(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{-i\lambda}|^{-2d} = \frac{\sigma^2}{2\pi}\left(4\sin^2\frac{\lambda}{2}\right)^{-2d}.$$

Remark 6.4.2 (Simulation)**.** *Such integral representations are used*

to simulate such models. For the case of Gaussian inputs the previous spectral process is Gaussian with independent increments which makes the previous simulation trick possible by providing independent random variables with a given distribution, see Remark A.1.4.
This idea extends to each process with independent increments as Poisson unit process.
Another possibility to simulate such time series is clearly to truncate the corresponding series. The problem is that simulation is approximative in this case.

6.5 FARIMA(p, d, q)−processes

The models FARIMA(p, d, q) fit the equation

$$\alpha(B)(I - B)^d X_t = \beta(B)\xi_t.$$

If $d < \frac{1}{2}$ the process is causal and well defined in case the roots of α are not inside the unit disk.
It is invertible if $d > -\frac{1}{2}$ and the roots of α are out of the unit disk. Indeed in this case $\xi_t = \gamma(B)X_t$ for a function γ analytic on the unit disk

$$D(0,1) = \{z \in \mathbb{C}/|z| < 1\}.$$

Let again Z denote the random spectral measure associated to the white noise ξ_t then

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \left(1 - e^{-i\lambda}\right)^{-d} \frac{\beta(e^{i\lambda})}{\alpha(e^{i\lambda})} Z(d\lambda).$$

Thus

$$g_X(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{-i\lambda}|^{-2d} \left|\frac{\beta(e^{i\lambda})}{\alpha(e^{i\lambda})}\right|^2.$$

6.6 Extensions

Clearly for any meromorphic function $\gamma : \mathbb{C} \to \mathbb{C}$ without singularities on $D(0,1)$ with finitely many singularities on the unit circle we mayo define a process

$$X_t = \gamma(B)\xi_t.$$

In case $1/\gamma$ satisfies the same assumptions then the process is reversible (zeros replacing singularities).
Singularities $\neq 1$ on the unit circle circle are called periodic long range singularities.
Now let $(c_{k,n})_{k,n\in\mathbb{Z}}$ be a sequence of real numbers, analogously to eqn. (6.1) we may define non stationary linear processes from the relation

$$X_n = \sum_{k=-\infty}^{\infty} c_{k,n}\xi_{n-k}.$$

Analogue existence results may be derived in this case.

Chapter 7

Non-linear processes

This Chapter aims at describing stationary sequences generated from independent identically distributed samples $(\xi_n)_{n\in\mathbb{Z}}$.
Many usual models of statistics will be proved to be this way.
This organisation follows the order from natural extensions of linearity to more general settings.
From linear processes it is natural to build polynomial models or their limits. Then we consider more general Bernoulli shift models in order to define recurrence equations besides the standard Markov setting.

7.1 Discret chaos

This section aims at introducing some basing tools for algebraic extensions of linear to polynomial models.

7.1.1 Volterra expansions

Set $X_n^{(0)} = c^{(0)}$ some constant and consider arrays $(c_j^{(k)})_{j\in\mathbb{Z}^k}$ of constants and a sequence of arrays of independent identically distributed random variables $\left(\left(\xi_n^{(k,j)}\right)_{1\le j\le k}\right)_{n\in\mathbb{Z}}$.

In case this makes sense set:

$$X_n^{(k)} = \sum_{j_1<j_2<\cdots<j_k} c_{j_1,\ldots,j_k}^{(k)} \xi_{n-j_1}^{(k,1)} \cdots \xi_{n-j_k}^{(k,k)}$$

then a Volterra expansion writes as:

$$X_n = \sum_{k=0}^{\infty} X_n^{(k)}.$$

Remark 7.1.1. *According to the previous Gaussian Chapter 5, such stationary models also write in the chaos generated from*

$$\left(\left(\xi_n^{(k,j)}\right)_{1\le j\le k}\right)_{n\in\mathbb{Z}}.$$

Anyway we prefer to keep on with the more standard expression of Volterra expansions.

Example 7.1.1. *In order to understand why the previous definition involves arrays of independent identically distributed random variables $((\xi_n^{(k,j)})_{1\le j\le k})_{n\in\mathbb{Z}}$, it seems to be better to consider the simplest example of second degree polynomials*

$$X_n = \sum_{i,j=-\infty}^{\infty} a_{i,j}\xi_i\xi_j,$$

the previous expansion holds if we note

$$\begin{aligned}
X_n^{(2)} &= \sum_{i<j}(a_{i,j}+a_{j,i})\xi_{n-i}\xi_{n-j},\\
X_n^{(1)} &= \sum_{i} a_{i,i}(\xi_{n-i}^2-\sigma^2), \qquad \sigma^2=\mathbb{E}\xi_0^2\\
X_n^{(0)} &= \sum_{i} a_{i,i}\sigma^2.
\end{aligned}$$

For Volterra models with higher order Appell polynomials $A_s(\xi_n)$ replace $\xi_n^2-\sigma^2$ in order to take into account the repetitions in diagonal terms.

Exercise 18. *Suppose (without loss of generality) that* $\mathbb{E}\left|\xi_n^{(k,j)}\right|^2 = 1$, *then*

$$\begin{array}{rcll} \mathbb{E}X_0^{(k)}X_n^{(l)} & = & 0, & \textit{if } k \neq l, \\ \mathbb{E}X_0^{(k)}X_0^{(k)} & = & \displaystyle\sum_{j_1<j_2<\cdots<j_k} \left|c_{j_1,\ldots,j_k}^{(k)}\right|^2, & \\ \mathbb{E}X_0^{(k)}X_n^{(k)} & = & \displaystyle\sum_{j_1<j_2<\cdots<j_k} c_{j_1,\ldots,j_k}^{(k)} c_{n+j_1,\ldots,n+j_k}^{(k)}. & \end{array}$$

Those calculations yield explicit expressions for the covariance of the process $(X_n)_{n\in\mathbb{Z}}$ *from a simple summation in case*

$$\sum_{j_1<j_2<\cdots<j_k} \left|c_{j_1,\ldots,j_k}^{(k)}\right|^2 < \infty.$$

7.1.2 Appell polynomials

Analogously for the special case of the Gaussian laws which yields the construction of Hermite chaos, one may define orthogonal polynomials associated to a fixed distribution on the real line $\mathbb{R}$.
Let ξ_0 a real valued random variable with finite moments up to some order $m > 0$.
Appell polynomials $A_0, \ldots, A_m$ are defined recursively with $A_0(x) = 1$ and

$$A_k'(x) = kA_{k-1}(x), \qquad \sum_{j=0}^{k} \mathbb{E}\xi_0^j \cdot A_j(0) = 0, \qquad 1 \leq k \leq m.$$

Hence

$$\begin{array}{rcl} A_0(x) & = & 1 \\ A_1(x) & = & x - \mathbb{E}\xi_0 \\ A_2(x) & = & x^2 - 2\mathbb{E}\xi_0 x + 2(E\xi_0)^2 - \mathbb{E}\xi_0^2 \\ \cdots & \cdots & \cdots \\ A_k(x) & = & x^k + \cdots \end{array}$$

If the Laplace transform of ξ_0's distribution is analytic around 0, this entails

$$\sum_{k=0}^{\infty}\frac{z^k}{k!}A_k(x)\mathbb{E}e^{z\xi_0}=e^{zx}.$$

Let P be a polynomial with $d=d^\circ P$: it may be uniquely written as

$$P(x)=\sum_{k=0}^{d}\frac{c_k}{k!}A_k(x).$$

(reasoning on the degree allows to derive uniqueness).
If the cumulative distribution function F of ξ_0's distribution ($F(x)=\mathbb{P}(\xi_0\le x)$) is regular enough we denote by $f=F'$ the density of this law. Then

$$c_k=\mathbb{E}P^{(k)}(\xi_0)=(-1)^k\int_{-\infty}^{\infty}P(x)f^{(k)}(x)dx.$$

An important property of those Appell polynomials writes

$$\mathbb{E}A_k(\xi_0)P(\xi_0)=0,\qquad\text{if}\quad d^\circ P<k.$$

Set $g=fP$ then

$$\mathbb{E}A_k(\xi_0)P(\xi_0)=\int_{-\infty}^{\infty}A_k(x)g(x)dx.$$

Since the function g admits k derivatives then k integrations by parts prove this identity. Set $g_l(x)=f^{(l)}/f$ thus ([1])

$$\mathbb{E}A_k(\xi_0)g_l(\xi_0)=\begin{cases}1, & \text{if}\quad k=l\\ 0, & \text{if}\quad k\neq l.\end{cases}$$

Remark 7.1.2. *Extensions to more general functions is much more complicated that the previous Gaussian theory! To be in order to consider non-polynomial functions [Kazmin, 1969] assumes that the function*

$$x\mapsto A(z)=1/\mathbb{E}e^{z\xi}$$

[1] The proof is quite analogue to that for the Gaussian chaos.

is analytic and it does not vanish on the open disk

$$D_\sigma \equiv \{z \in \mathbb{C}/|z| < \sigma\}.$$

Then each function $g \in E_\tau$ (set of analytic function on a disk D_τ) admits a representation

$$g(z) = \sum_{n=0}^{\infty} \frac{c_n}{n!} A_n(z), \quad \limsup_{n\to\infty} |c_n|^{1/n} < \tau$$

for series which converge uniformly over compact subsets of the disk D_τ.
Conversely for a sequence such that

$$\limsup_{n\to\infty} |c_n|^{1/n} < \tau,$$

the function g defined this way is proved to be analytic on D_τ.
Under those assumption the series defining g is convergent and

$$c_n = \mathbb{E} g^{(n)}(\xi,)$$

thus this proves uniqueness of the expansion of analytic functions.
Those results are far from representing all the $\mathbb{L}^2$ functions as in the Gaussian case.

Multivariate Appell polynomials
If now $\xi = (\xi_1, \ldots, \xi_k) \in \mathbb{R}^k$ is a vector valued random variable this is easy to define analogously $A_{n_1,\ldots,n_k}(x_1, \ldots, x_k)$ through relations

$$\begin{aligned} \frac{\partial}{\partial x_i} A_{n_1,\ldots,n_k}(x_1, \ldots, x_k) &= n_i A_{n_1,\ldots,n_k}(x_1, \ldots, x_k), \quad 1 \le i \le k \\ \mathbb{E} A_{n_1,\ldots,n_k}(\xi) &= 1 \text{ if } n_1 + \cdots + n_k = 0 \text{ and } 0 \text{ else.} \end{aligned}$$

If random variables $\xi_1, \ldots, \xi_k$ are independent and admit respective distributions $\nu_1, \ldots, \nu_k$, then

$$A_{n_1,\ldots,n_k}(x_1, \ldots, x_k) = A_{n_1}^{(\nu_1)}(x_1) \cdots A_{n_k}^{(\nu_k)}(x_k).$$

7.2 Memory models

The section aims at considering some models which Volterra expansions may be explicitly determined.
As usual those memory models will be excited by iid innovations with values in the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. They are solution of some recursion:

$$X_t = M(X_{t-1}, X_{t-2}, \dots, \xi_t).$$

for iid inputs an some explicit function $\mathbb{R}^{\mathbb{N}} \times \mathbb{R} \to \mathbb{R}$.
In some cases more complicated spaces may by used for innovations and for the model but this section is essentially restricted to real values for simplicity.

7.2.1 Bilinear models

For simplicity we first

Proposition 7.2.1. *Consider the Markov bilinear model*

$$X_n = (a + b\xi_n)X_{n-1} + \xi_n, \tag{7.1}$$

Assume that for some $p \geq 1$,

$$\alpha^p = \mathbb{E}|a + b\xi_0|^p < 1.$$

Then there exists stationary solution of this Markov recursion, this solution is in $\mathbb{L}^p$ and writes:

$$X_n = \sum_{k=0}^{\infty} \xi_{n-k} \prod_{j=0}^{k-1} (a + b\xi_{n-j}).$$

Proof. It is simple to check that the previous series in normally convergent since independence entails

$$\|\xi_{n-k} \prod_{j=0}^{k-1} (a + b\xi_{n-j})\|_p = \|\xi_0\|_p \|a + b\xi_{n-j}\|_p^k.$$

In order to check the result write

$$X_n = \sum_{k=0}^{m} \xi_{n-k} \prod_{j=0}^{k-1} (a + b\xi_{n-j}) + X_{n-m} \prod_{j=0}^{m-1} (a + b\xi_{n-j}).$$

Then previous remark implies that the main term in this equality is a convergent as $m \uparrow \infty$, and its L^p-norm is thus bounded above by some $A > 0$.
Now this also entails $(1-\alpha)\|X_0\|_p \le A$. ∎

Bilinear models (7.1) behave quite analogously to some white noise. Their covariances present some bumps and then rapidly decay. In Figure 7.1 we present in fact empirical covariances, the convergence of those expressions is considered later on: see Remark 9.1.3 and Examples 9.1.3 include the current model. For such models a recursion is also available for the sequence of covariances.

Exercise 19. *Assume that $\mathbb{E}\xi_0 = 0$, $\mathbb{E}\xi_0^2 = 1$ and consider the $\mathbb{L}^2-$strictly stationary solution (X_t) of eqn. (7.1). Set $M = \mathbb{E}X_0^2$ and $C = Cov(X_0, X_1)$.*

1. *Prove that*

$$\mathbb{E}X_0 = 0, \quad M = \frac{1}{1-(a^2+b^2)}, \quad C = \frac{a}{1-(a^2+b^2)}$$

2. *From empirical estimates of the previous expressions*

$$\widehat{M} = \frac{1}{n}\sum_{k=1}^n X_k^2, \qquad \widehat{C} = \frac{1}{n-1}\sum_{k=2}^n X_k X_{k-1},$$

deduce that the following estimates of parameters a, b in the model are consistent:

$$\widehat{a} = \frac{\widehat{M}}{\widehat{C}}, \qquad \widehat{b} = \sqrt{\frac{\widehat{M}^2 - \widehat{C}^2 - \widehat{M}}{\widehat{M}}}$$

Hints.

1. From independence of ξ_t with X_{t-1} and eqn. (7.1): $\mathbb{E}X_1 = a\mathbb{E}\xi_0$, $\mathbb{E}X_1^2 = \mathbb{E}(a+b\xi_0)^2\mathbb{E}X_0^2 + \mathbb{E}\xi_0^2$ hence $M(1-(a^2+b^2) = 1$, moreover $C = \mathbb{E}X_0X_1 = aM$.

2. The previous relations are rewritten accurately:

$$C = aM, \qquad M^2(1-(a^2+b^2)) = M^2(1-b^2) - C^2 = M$$

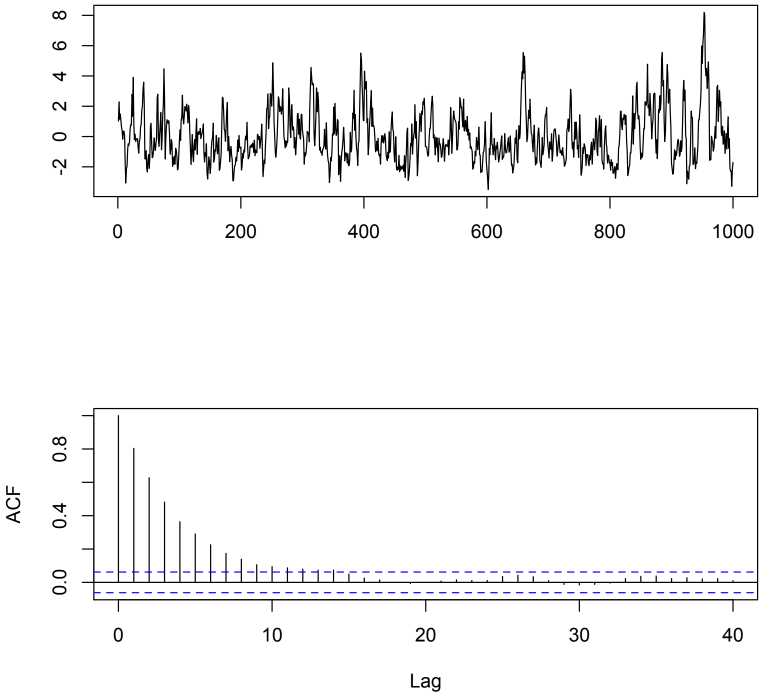


Figure 7.1: Bilinear model.

hence $Mb^2 = M^2 - C^2 - M$, thus

$$a = \frac{M}{C}, \qquad b = \sqrt{\frac{M^2 - C^2 - M}{M}}$$

See results in § 7.3.3 for a formal justification and the consistency of those estimators, namely the ergodic theorem applies to prove a.s. consistency of those estimates (Corollary 9.1.3) and a $\sqrt{n}$−CLT also applies to get asymptotic confidence bounds for those estimates. The $\Delta-$method applies to transfer those properties to the proposed plug-in estimates.

One variant for this model (7.1) writes

$$X_n = h(\xi_n)X_{n-1} + \xi_n$$

and, in case $\mathbb{E}|h(\xi_0)| < 1$, a stationary solution still writes

$$X_n = \sum_{k=0}^{\infty} \xi_{n-k} \prod_{j<k} h(\xi_{n-j}).$$

Notice that analogue expressions may be provided under more complicated assumptions for models like

$$X_n = H_n X_{n-1} + \xi_n$$

for some adapted and stationary sequence H_n eg.

$$\begin{aligned}
X_n &= \xi_n + H_n X_{n-1} \\
&= \xi_n + H_n(\xi_{n-1} + H_{n-1}X_{n-2}) = \xi_n + \xi_{n-1}H_n + X_{n-2}H_nH_{n-1} \\
&= \xi_n + \xi_{n-1}H_n + (\xi_{n-2} + H_{n-2}X_{n-3})H_nH_{n-1} \\
&= \xi_n + \xi_{n-1}H_n + \xi_{n-2}H_nH_{n-1} + H_{n-2}X_{n-3}H_nH_{n-1} \\
&= \cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

If $\sum_{k=0}^{\infty} \|\xi_{n-k} \prod_{j<k} H_{n-j}\|_p < \infty$ then the following series is convergent in $\mathbb{L}^p$, and it defines a solution of the previous recursion:

$$X_n = \sum_{k=0}^{\infty} \xi_{n-k} \prod_{j<k} H_{n-j}$$

In the case when $H_n = (\xi_n, \xi_{n-1}, \ldots, \xi_{n-r+1})$ then H_n X_{n-1} are not independent anymore which needs additional moment conditions,

$$\begin{aligned}
\|\xi_{n-k} \prod_{j<k} H_{n-j}\|_p &= \|\xi_{n-k} \prod_{k-r<j<k} H_{n-j}\|_p \|\prod_{j=0}^{k-r} H_{n-j}\|_p \\
&\le \|\xi_{n-k} \prod_{k-r<j<k} H_{n-j}\|_p \|H_0\|_{pr}^{(\ell-1)r}, \text{ if } k = \ell r
\end{aligned}$$

indeed this follows from Hölder inequality if $k = \ell r$ that

$$\prod_{j=0}^{(\ell-1)r} H_{n-j},$$

writes as the product of r products of products of $(\ell-1)$ products of independent sequences. Assumptions

$$\|H_0\|_{pr} < 1, \qquad \|H_0 \cdots H_{r-1}\xi_r\|_p < \infty,$$

together ensure the $\mathbb{L}^p-$convergence of the previous series.
The last relation holds if

$$\|\xi_0\|_{qr} < \infty \text{ and } \|H_0\|_{q'pr} < \infty,$$

for $q, q' \in [1, +\infty]$ with $\dfrac{1}{q} + \dfrac{1}{q'} = 1$.

7.2.2 LARCH$(\infty)-$models

Theorem 7.2.1. *Consider the general non-Markov model solution of the recurrence equation of the LARCH$(\infty)-$equation:*

$$X_n = \left(b_0 + \sum_{j=1}^{\infty} b_j X_{n-j}\right)\xi_n.$$

A $\mathbb{L}^p-$valued strictly stationary solution of this recursion writes

$$\begin{aligned} X_n &= b_0 \sum_{k=1}^{\infty}\sum_{l_1=1}^{\infty}\cdots\sum_{l_k=1}^{\infty} b_{l_1}\cdots b_{l_k}\xi_{n-l_1}\xi_{n-l_1-l_2}\cdots\xi_{n-(l_1+\cdots+l_k)} \\ &= b_0\sum_{k=1}^{\infty}\sum_{0<j_1<\cdots<j_k=1}^{\infty} b_{j_1}b_{j_2-j_1}\cdots b_{j_k-j_{k-1}}\xi_{n-j_1}\xi_{n-j_2}\cdots\xi_{n-j_k} \end{aligned}$$

Under condition

$$\|\xi_0\|_p \sum_{k=1}^{\infty} |b_k| < 1.$$

Hints. Indeed, this is easy to derive from independence of each factor that: $\|\xi_{n-j_1}\xi_{n-j_2}\cdots\xi_{n-j_k}\|_p = \|\xi_0\|_p^k$. ∎

If now the variables ξ_n are centered and admit a finite variance the previous representation still holds in $\mathbb{L}^2$ if

$$\mathbb{E}\xi_0^2 \sum_{k=1}^{\infty} b_k^2 < 1.$$

This assumption allows long rand dependent behaviors as proved in Giraitis *et al.* [Giraitis et al., 2012].
A vector valued variant of this model as well as a random field variants have both been developed.
Usual ARCH–models $(Y_n)_{n\in\mathbb{Z}}$ are such that squares $X_n = Y_n^2$ satisfy the previous equation.
They are defined through a sequence of nonnegative real numbers (b_j) with $b_j = 0$ if j is large enough or a centered sequence of independent identically distributed random variables (ξ_j)

$$Y_n = \sqrt{b_0 + \sum_{j=1}^{J} b_j Y_{n-j}^2} \cdot \xi_n.$$

In this case the vector valued model $Y_n = (X_n, \ldots, X_{n-J+1})$ is a Markov process will values in $\mathbb{R}^J$. Quote that the general model is not $J-$Markov for any $J > 0$.

7.3 Stable Markov chains

Proposition 7.6 of [Kallenberg, 1997] proves that any Markov chain (homogeneous in time) (X_t) with values in $\mathbb{R}^d$ for some $d \geq 1$ may be represented as the solution of a recursion or iterative random model or autoregressive models assuming the Condition 1 below:

$$X_t = M(X_{t-1}, \xi_t). \tag{7.2}$$

Condition 1. $(\xi_t)_{t\in\mathbb{Z}}$ *an independent identically distributed sequence with values in a measurable space* $(E, \mathcal{E})$ *for a measurable function* M

is a (measurable) kernel

$$M : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \times (E, \mathcal{E}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$$

For several models the innovation space has to be specified differently. Sometimes it will be $\mathbb{R}^d$ but sometimes a product space (associated to thinning operators) or a point process distribution (associated to Poisson processes) are needed.
The objective of the section is to determine simple conditions for such iterative models to admit a stationary solution. Further we will see that such solutions write as Bernoulli schemes (7.11).

Definition 7.3.1. *Suppose that (ξ_t) is an independent identically distributed with values in a space E, moreover for $d \geq 1$ and for some measurable space $(E, \mathcal{E})$ (we denote by $\|\cdot\|$ any norm on $\mathbb{R}^d$). A stable Markov chain $X_n \in \mathbb{R}^d$ in $\mathbb{L}^p$ $(p \geq 1)$ is a solution of a recursive stochastic equation (7.2) satisfying the conditions 1 and $\forall u, v \in \mathbb{R}^d$:*

$$\exists a \in [0,1), \quad \mathbb{E}\|M(u,\xi_0) - M(v,\xi_0)\|^p \leq a^p \|u - v\|^p, \tag{7.3}$$

$$\exists u_0 \in \mathbb{R}^d, \quad |M(u_0,\xi_0)|^p < \infty. \tag{7.4}$$

Condition 2. *We assume in the previous definition as in [Duflo, 1996] that such a model is contracting; this means that the kernel $M(u,z)$ fits the condition (7.3).*

Condition 3. (Fixed Point) *Suppose also that for some $e \in E$ the function $u \mapsto M(u,e)$ admits a fixed point u_0 (if E is a vector space a simple change allow to suppose $e = 0$).*

Proof. Define $(U_t^{(n)})_{t \in \mathbb{Z}}$ a Markov chain such that
$U_t^{(n)} = u_0$ if $t \leq -n$, and $U_t^{(n)} = M(U_{t-1}^{(n)}, \xi_t)$ if $t > -n$.
Lipschitz condition implies with independence of inputs:

$$\mathbb{E}\left\|U_0^{(n)} - U_0^{(n+1)}\right\|^p \leq a^p \mathbb{E}\left\|U_0^{(n-1)} - U_0^{(n)}\right\|^p.$$

From a recursion

$$\mathbb{E}\left\|U_0^{(n)} - U_0^{(n+1)}\right\|^p \leq a^{np} \mathbb{E}\left\|M(u_0,\zeta_0) - u_0\right\|^p$$

Hence $U_0^{(n)} \to U_0$ $(n \to \infty)$ converges in $\mathbb{L}^p$ to a random variable $U_0 \in \mathbb{L}^p$.
Moreover $U_0^{(n)}$ is measurable wrt the $\sigma-$algebra generated by $\{\xi_t,\ t \le 0\}$ hence this is also the case for U_0. U_0 may thus also be represented as a function $U_0 = H(\xi_0, \xi_{-1}, \ldots)$ of this sequence.
Then the sequence $X_t = H(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots)$ is a stationary solution of the previous recursion.
Now the sequences $(U_t^{(0)})_t$ and $(U_t^{(1)})_t$, satisfy

$$\begin{aligned} U_0^{(0)} &= u_0, \\ U_1^{(0)} &= M(u_0, \xi_1) = H(\xi_1, 0, 0, \ldots), \\ U_2^{(0)} &= M(M(u_0, \xi_1), \xi_2) = H(\xi_2, \xi_1, 0, 0, \ldots) \end{aligned}$$

and from a recursion for each $t > 0$,

$$U_t^{(0)} = V(\xi_t, \xi_{t-1}, \ldots, \xi_1, 0, 0, 0, \ldots).$$

Analogously

$$U_t^{(1)} = H(\xi_t, \xi_{t-1}, \ldots, \xi_1, \xi_0, 0, 0, \ldots).$$

Hence $\gamma_n = \mathbb{E}^{1/p} \left\| U_n^{(0)} - U_n^{(1)} \right\|^p \le a\gamma_{n-1}$ and

$$\gamma_n \le a^n \gamma_0 = a^n \mathbb{E}^{1/p} \left\| M(u_0, \zeta_0) - u_0 \right\|^p \tag{7.5}$$

decays exponentially to 0 since $a < 1$. In fact the assumption that $u \mapsto M(u, e)$ admits a fixed point may simply be replaced by assumption (7.4).
Only set $U_{-n}^{(n)} = M(u_0, \xi_{-n})$. In fact we obtain the following:

Theorem 7.3.1. *Assume that conditions (7.4) and (7.3) hold for some $p \ge 1$.*
The equation (7.2) admits a stationary condition in $\mathbb{L}^p$ such that for each $t \in \mathbb{Z}$, X_t is measurable wrt to the σ-algebra $\mathcal{F}_t = \sigma(\xi_s, s \le t)$.

Example 7.3.1. *[Diaconis and Freedman, 1995] provide nice series of examples for which the previous technique applies. One may also refer to [Doukhan, 1994], [Doukhan and Louhichi, 1999], as well as to the monograph [Dedecker et al., 2007].*

7.3.1 AR-ARCH–models

Proposition 7.3.1. *Let $d = 1$, $E = \mathbb{R}$ and set*

$$M(u,z) = A(u) + B(u)z \tag{7.6}$$

for Lipschitz functions $A(u), B(u), u \in \mathbb{R}$.
If

$$Lip(A) = \sup_{u \neq v} \frac{|A(u) - A(v)|}{|u - v|}$$

then the stability conditions in Definition 7.3.1 hold if $p = 2$ and $\mathbb{E}\xi_t = 0$ with

$$a^2 = (Lip(A))^2 + \mathbb{E}\xi_0^2\,(Lip(B))^2 < 1.$$

and if $p \geq 1$ in case

$$a = Lip(A) + \|\xi_0\|_p Lip(B) < 1.$$

Proofs. Note that Minkowski inequality implies that for $p \geq 1$,

$$\|M(u_0,\xi_0)\|_p \leq \|A(u_0)\| + \|B(u_0)\|\|\xi_0\|_p,$$
and
$$\|M(u,\xi_0) - M(v,\xi_0)\|_p \leq \|A(u) - A(v)\| + \|B(u) - B(v)\|\|\xi_0\|_p,$$

which allows to derive the second point of the Proposition.
If $p = 2$ then

$$\begin{aligned}\mathbb{E}(M(u,\xi_0) - M(v,\xi_0))^2 &= (A(u) - A(v))^2 + (B(u) - B(v))^2\mathbb{E}\xi_0^2 \\ &\quad + 2(A(u) - A(v))(B(u) - B(v))\mathbb{E}\xi_0\end{aligned}$$

and the last rectangle term simply vanishes from $\mathbb{E}\xi_0 = 0$, this allows to improve the previous bound, indeed

$$(\mathrm{Lip}(A))^2 + \mathbb{E}\xi_0^2\,(\mathrm{Lip}(B))^2 \leq (\mathrm{Lip}(A) + \|\xi_0\|_2\mathrm{Lip}(B))^2\,.$$

Those relations yield a simple way to conclude. ∎

Examples of such models follow:

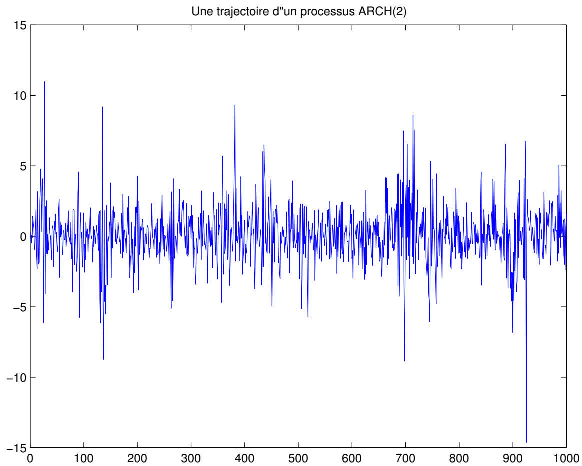


Figure 7.2: ARCH(2)–trajectory.

Example 7.3.2 (Some special cases)**.**

- *Non-linear AR(1)–models ($B \equiv 1$) satisfy the equation*

$$X_n = A(X_{n-1}) + \xi_n.$$

- *Stochastic volatility models ($A \equiv 0$) are solution of the equation*

$$X_n = B(X_{n-1})\xi_n.$$

- *The AR-ARCH(1)–classical models is solution of equation*

$$X_n = \alpha X_{n-1} + \sqrt{\beta + \gamma^2 X_{n-1}^2} \cdot \xi_n.$$

Here $A(u) = \alpha u$ and $B(u) = \sqrt{\beta + \gamma^2 u^2}$ for $\alpha, \beta, \gamma \geq 0$. Lipschitz constant writes $a = \alpha^2 + \mathbb{E}\zeta_0^2 \gamma$ from a direct calculation of the derivatives $A'(u) = \alpha$ and

$$|B'(u)| = \frac{\gamma^2 |u|}{\sqrt{\beta + \gamma^2 u^2}} = \gamma \cdot \frac{\sqrt{\gamma^2 u^2}}{\sqrt{\beta + \gamma^2 u^2}} \leq \gamma.$$

This model is defined conditionally wrt to its past history:

$$X_t \,|\, \mathcal{F}_{t-1} \;\sim\; \mathcal{N}(\alpha X_{t-1}, \beta + \gamma^2 X_{t-1}^2),$$

quote that the above recursion is just the simplest way to get such conditional distributions for Gaussian innovations.

- *ARCH(2)–models are solutions of equations*

$$X_t = \sigma_t \xi_t, \qquad \sigma_t^2 = \alpha^2 + \beta^2 X_{t-1}^2 + \gamma^2 X_{t-2}^2$$

Their trajectories may be seen in Figure 7.2.

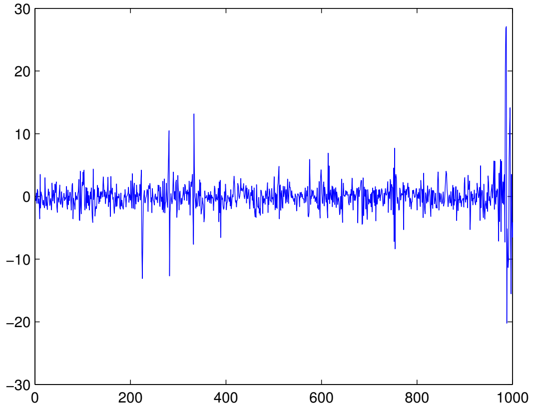


Figure 7.3: A GARCH(1,1)–trajectory.

- *GARCH(1,1)–models are solutions of equations*

$$X_t = \sigma_t \xi_t, \qquad \sigma_t^2 = \alpha^2 + \beta^2 X_{t-1}^2 + \gamma^2 \sigma_{t-1}^2.$$

This is clear through iterations that one may rewrite such models as

$$\sigma_t^2 = \alpha^2 + \sum_{k=1}^{\infty} \beta_k^2 X_{t-1}^2,$$

and such models have also be designed for financial purposes for their clustering aspects and trajectories may be seen in Figure 7.3.

7.3.2 Moments of ARCH(1)-models

We are interested here to check that recursive models without low order moments may be generated from inputs with all finite moments. Consider the simplest ARCH-model

$$X_t = \sqrt{\beta + \gamma^2 X_{t-1}^2} \cdot \xi_t.$$

Check that the function $p \mapsto \|\xi_0\|_p$ is monotonically non-decreasing from Jensen inequality (Proposition A.1.1) applied with $t \mapsto t^r$ for $r \geq 1$. If $|\xi_0|$ is not constant a.s. this function is strictly increasing. E.g. if $|\xi_0| \in \{0, a\}$ then

$$\|\xi_0\|_p = (1 + a^p \mathbb{P}(|\xi_0| = a))^{1/p}.$$

More precisely the forthcoming Lemma will give a precise answer. Hence if $\gamma\|\xi_0\|_2 = 1$ the previous equation admits a strictly stationary solution in $\mathbb{L}^p$ for each $p < 2$. Moreover this solution is not $\mathbb{L}^2-$integrable. Else indeed:

$$\mathbb{E}X_t^2 = (\beta + \gamma^2 \mathbb{E}X_{t-1}^2)\|\xi_t\|_2^2 = \beta\|\xi_t\|_2^2 + \mathbb{E}X_{t-1}^2 \qquad (^2).$$

For the AR-LARCH models with centered inputs the limit condition

$$\alpha^2 + \gamma^2 \mathbb{E}\xi_0^2 = 1$$

analogously implies that any solution of this equation does not have second order moment. Also there exists a $\mathbb{L}^p-$solution of this equation in case p is small enough in case $|\xi_0|$ is not constant. This is simple to see if either α of $\gamma = 0$.

Lemma 7.3.1. *Let $Z \geq 0$ be a non-negative and non a.s. constant random variable such that $\mathbb{E}Z^m < \infty$ for some $m > 0$ then the function $p \mapsto \|Z\|_p$ defined $(0, m] \to \mathbb{R}^+$ is strictly monotonic.*

Proof. The present proof follows from a personal communication with Adam Jakubowski.
With $Z = |\xi_0|^p$ we need to prove that if $p' > p$ and $r = p'/p$ then $\mathbb{E}Z \le \|Z\|_r$.
As in the proof of (A.2) Jensen inequality for $g(u) = u^r$ with $r = p'/p > 1$ we consider an affine minorant $f(u) = au+b$ for the function g with $f(z) = g(z)$ for some z to be defined ($a = rz^{r-1}$ makes $f'(z) = g'(z)$ and $b = (1-r)z^r$ then does $f(z) = g(z)$).
Now if $u \ne z$ then $f(u) < g(u)$ hence $\mathbb{E}f(Z) < \mathbb{E}g(Z)$ because Z is not a.s. a constant.
Let now $z = \mathbb{E}Z$ then $\mathbb{E}f(Z) = (\mathbb{E}Z)^r < \mathbb{E}g(Z) = \mathbb{E}Z^r$.
This is enough to conclude. ∎

7.3.3 Estimation of LARCH(1)–models

This section aims at describing some important features of LARCH(1)–models in order to provide some simple estimators of their parameters as this was already sketched in Exercise 19. Ideas are essentially those from Yule-Walker equations, § 6.3, and the main point is a MA–representation with $\mathbb{L}^2-$weak-white noise inputs.
Besides (ξ_t) is an iid real valued sequence with $\mathbb{E}|\xi_0|^p < \infty$ for some $p > 0$ and

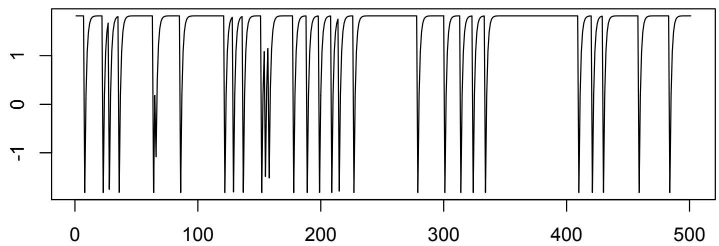
$$Z_t = (\beta + \delta Z_{t-1})\xi_t \tag{7.7}$$

Quote that even though covariances of the model appear to decay quite rapidly, the behavior of trajectory looks quite erratic (Figure 7.4).

Lemma 7.3.2. *Let $p > 0$ a fixed positive number. Then assumption $|\delta| \cdot \|\xi_0\|_p < 1$ implies that a unique stationary solution exists and it is in $\mathbb{L}^p$.*

Proof. $|\delta| \|\xi_0\|_p < 1$ is the contraction constant in this case. Now the solution of the equation is the limit of a polynomial in the innovations and thus writes as a Bernoulli shift in $\mathbb{L}^p$. ∎

A first proposal of estimators was provided in § 4.4.2 (the Whittle estimator of the parameter $\theta = (\beta, \delta)$. It needs explicit expressions



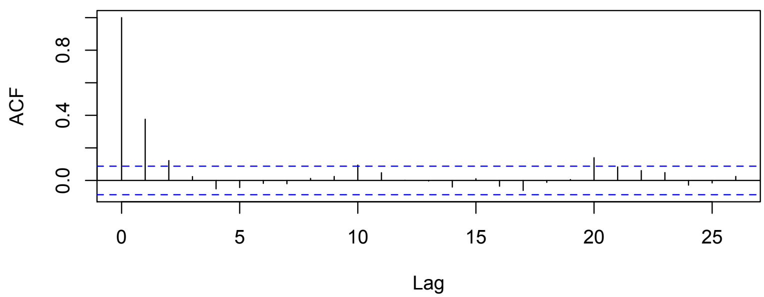


Figure 7.4: A LARCH(1)–model.
Quote that even though covariances of the model appear to decay quite rapidly, the behavior of trajectory looks quite erratic.

of Z's spectral density (or equivalently of all the covariances of Z, which may be quite heavy.
In § 4.5 the QMLE of such Markov chains

$$Z_t = \xi_t \sigma_\theta(X_{t-1})$$

is considered in case $\xi_t \sim \mathcal{N}(0,1)$; here the transition probability density writes

$$\pi_\theta(x,y) = \frac{1}{\sqrt{2\pi(\beta+\delta x)^2}} \exp(-\frac{1}{2}\frac{y^2}{(\beta+\delta x)^2})$$

the QMLE of (Z_t) is now the couple $\theta = (\beta, \delta)$ minimizing the expression

$$L_\theta(Z_1,\ldots,Z_n) = \sum_{t=2}^{n} \frac{Z_t^2}{(\beta+\delta Z_{t-1})^2} + \log(\beta+\delta Z_{t-1})^2$$

This estimator is considered in the most general situations in the monograph [Straumann, 2005].
Anyway in our simple situation we choose a more direct way to estimate the parameters. It will result in simple empirical estimators.

Lemma 7.3.3 (Close expressions of moments)**.** *Let (Z_t) be the stationary solution of eqn. (7.7), we assume that $Z_t \in \mathbb{L}^p$.*

1. *Assume that $p \geq 1$ then $m = \mathbb{E}Z_0 = \beta\mathbb{E}\xi_0/(1-\delta)$.*
2. *Assume that $p \geq 2$ and assume $\mathbb{E}\xi_0 = 1$ and set $\nu = \mathbb{E}\xi_0^2$ then*
$$M := \mathbb{E}Z_0^2 = \frac{\nu\beta^2(1+\delta)}{(1-\delta)(1-\nu\delta^2)}$$
3. *Assume now that $p \geq 3$ and that $\mathbb{E}\xi_0 = E\xi_0^3 = 0$, then:*
$$M = \mathbb{E}Z_0^2 = \frac{\nu\beta}{1-\nu\delta^2}$$
Set $\ell = Cov(Z_0, Z_1^2) = \mathbb{E}Z_0Z_1^2$ the leverage of Z then:
$$\ell = 2\nu\beta\delta M = \frac{2\nu^2\beta^2\delta}{1-\nu\delta^2}.$$
4. *Assume now that $p \geq 3$ and that $\mathbb{E}\xi_0 = 0$ and $\mathbb{E}\xi_0^3 \neq 0$, then:*
$$M = \mathbb{E}Z_0^2 = \frac{\nu\beta}{1-\nu\delta^2}$$
Set $\eta = \mathbb{E}\xi_0^3$ then:
$$P = \mathbb{E}Z_0^3 = \frac{\eta\beta(\beta^2+3\delta^2M)}{1-\eta\delta^3}.$$

Proofs.

1. From $\mathbb{L}^1-$stationarity and independence: $\mathbb{E}Z_0 = E\xi_0(\beta+\delta\mathbb{E}Z_0)$.

2. From independence and then $\mathbb{L}^2-$stationarity:

$$\begin{aligned} \mathbb{E}Z_0^2 &= \nu\mathbb{E}\{(\beta+\delta Z_0)^2\} \\ &= \nu(\beta^2 + 2\beta\delta\mathbb{E}Z_0 + \delta^2\mathbb{E}Z_0^2) \\ &= \nu\left(\beta^2 + \frac{2\beta^2\delta}{1-\delta} + \delta^2\mathbb{E}Z_0^2\right) \\ &= \nu\left(\frac{\beta^2(1+\delta)}{1-\delta} + \delta^2\mathbb{E}Z_0^2\right) \end{aligned}$$

3. Note that $\mathbb{E}\xi_0 = 0$ then as before we get the identity $\mathbb{E}Z_0^2 = \nu(\beta^2 + \delta^2\mathbb{E}Z_0^2)$, thus

$$M = \mathbb{E}Z_0^2 = \frac{\nu\beta}{1-\nu\delta^2}.$$

We have here $\mathbb{E}\xi_0^3 = 0$, hence:

$$\begin{aligned} \ell &= \nu\mathbb{E}Z_0(\beta+\delta Z_0)^2 \\ &= \nu(2\beta\delta\mathbb{E}Z_0^2 + \delta^2\mathbb{E}Z_0^3) \\ &= 2\nu\beta\delta\mathbb{E}Z_0^2 \qquad (7.8) \\ &= \frac{2\nu^2\beta^2\delta}{1-\nu\delta^2} \end{aligned}$$

4. From independence

$$P = \mathbb{E}Z_1^3 = \mathbb{E}\xi_1^3\mathbb{E}(\beta+\delta Z_0)^3$$

Using the binomial formula yields the result.

All the possible cases when moments exist have thus been considered.
■

Lemma 7.3.4. *A.s. consistent estimates of* $m = \mathbb{E}Z_0, M = \mathbb{E}Z_0^2$, $P = \mathbb{E}Z_0^3$ *and* ℓ *are provided if respectively* $p \geq 1, 2, 3$ *by*

$$\widehat{m} = \frac{1}{n}\sum_{k=1}^{n} Z_k, \qquad \widehat{M} = \frac{1}{n}\sum_{k=1}^{n} Z_k^2,$$
$$\widehat{P} = \frac{1}{n}\sum_{k=1}^{n} Z_k^3, \qquad \widehat{\ell} = \frac{1}{n-1}\sum_{k=1}^{n} Z_k Z_{k+1}$$

Proof. Quote that from Proposition 7.2.1, the process Z_t admits an explicit chaotic expansion with respect to the iid sequence (ξ_t), thus it is ergodic from the examples following Corollary 9.1.3. The ergodic theorem (Corollary 9.1.3) proves a.s. convergence of those expressions. ∎

From the above results we derive simple empirical estimators through the Δ−method, they are built from empirical estimators explicited by Lemma 7.3.4:

Corollary 7.3.1. *Assume that* $|\delta|\|\xi_0\|_1 < 1$ *and* $\beta = 1$ *then an a.s. consistent estimator of* δ *writes*

$$\widehat{\delta} = 1 - \frac{1}{\widehat{m}}$$

Corollary 7.3.2. *Assume that* $|\delta|\|\xi_0\|_2 < 1$ *and* $\mathbb{E}\xi_0 = 1, \mathbb{E}\xi_0^2 = \nu$ *then a.s. consistent estimators of* β, δ *write:*

$$\widehat{\delta} = \sqrt{\frac{\nu\widehat{M} - \widehat{m}^2}{\nu(\widehat{M} - \widehat{m}^2)}}, \quad \widehat{\beta} = \left(1 - \sqrt{\frac{\nu\widehat{M} - \widehat{m}^2}{\nu(\widehat{M} - \widehat{m}^2)}}\right)\widehat{m}$$

Remark 7.3.1. *Applying the previous results to the ARCH(1) model*

$$X_t = \sqrt{\beta + \delta X_{t-1}^2} \cdot \zeta_t.$$

is simple since $Z_t = X_t^2$ *is a LARCH(1)–model with innovations* $\xi_t = \zeta_t^2$ *hence* $\mathbb{E}\xi_0 \neq 0$ *and may be chosen equal to 1 and* $\nu = \mathbb{E}\zeta_0^4$.

Proof. $\beta = (1-\delta)m$ thus

$$\frac{m^2(1-\delta^2)}{1-\nu\delta^2} = M$$

and $m^2(1-\delta^2) = (1-\nu\delta^2)M$ implies $\nu\delta^2(\nu M - m^2) = (M-m^2)$, quote that $\mathrm{Var}\, Z_0 = M - m^2 \geq 0$.
Also, even though Cauchy-Schwarz inequality entails $\nu \geq 1$ the above relation entails $M - \nu m^2 = \nu m^2 \mathrm{Var}\, \xi_0 \geq 0$ and the following expression is well defined:

$$\delta = \sqrt{\frac{M-\nu m^2}{\nu(M-m^2)}}, \quad \beta = \left(1 - \sqrt{\frac{M-\nu m^2}{\nu(M-m^2)}}\right) m$$

Now the corresponding $\widehat{\beta}, \widehat{\delta}$ are thus consistent estimators. ∎

Corollary 7.3.3. *Assume that* $|\delta|\|\xi_0\|_3 < 1, \mathbb{E}\xi_0 = 0, \mathbb{E}\xi_0^2 = \nu, , \mathbb{E}\xi_0^3 = 0$ *then a.s. consistent estimators of* β, δ *write:*

$$\widehat{\delta} = -1 + \sqrt{1+\nu\widehat{\ell}}, \qquad \widehat{\beta} = \frac{\widehat{M}}{\nu}\left(2\sqrt{1+\nu\widehat{\ell}} - (1+\nu\widehat{\ell})\right)$$

Remark 7.3.2. *As as special case of the situation 3. in Lemma 7.3.3 quote that for the symmetric innovations with 3 moments, we have indeed* $\mathbb{E}\xi_0 = E\xi_0^3 = 0$.
In the special case $\mathbb{P}(\xi_0 = \pm 1) = \frac{1}{2}$ *of Rademacher distributed inputs and* $\beta = 1$, *[Doukhan et al., 2009] prove that the model is not strong mixing if* $\delta \in \left]\frac{3-\sqrt{5}}{2}, \frac{1}{2}\right]$.
Moreover $\ell\delta^2 + 2\delta - 1 = 0$ *admits the solution* $\delta = -1 + \sqrt{1+\ell}$. *Indeed the other solution of the previous second degree equation is not in* $]-1, 1[$.

Proof. Relations $\beta, \delta > 0$ imply with its existence that $\ell > 0$. Now eqn. (7.8) together with $\nu\beta = M(1-\delta^2)$ entails $\ell(1-\nu\delta^2) = 2\delta M$ thus δ is the positive solution of the second order equation

$$\nu\ell\delta^2 + 2\delta M - \ell = 0$$

hence $\delta = -1 + \sqrt{1 + \nu\ell}$.

$$\beta = \frac{M}{\nu}\left(1 - \left(-1 + \sqrt{1+\nu\ell}\,\right)^2\right) = \frac{M}{\nu}\left(2\sqrt{1+\nu\ell} - (1+\nu\ell)\right).$$

The plug-in empirical estimate takes the same form as above.
Depending on the sign of ℓ which is that of the product $\beta \cdot \delta$ one may choose the other solution of the equation $\delta = -1 - \sqrt{1+\nu\ell} < 0$.

∎

Remark 7.3.3. *Assume that* $|\delta|\|\xi_0\|_3 < 1$, $\mathbb{E}\xi_0 = 0, \mathbb{E}\xi_0^2 = \nu, \mathbb{E}\xi_0^3 = \eta \neq 0$ *and* $\beta, \delta > 0$ *then a.s. consistent estimators of* β, δ *write analogously by solving eqns. 4. in Lemma 7.3.3 and replacing* M, P *by their empirical counterparts* $\widehat{M}, \widehat{P}$.
To this aim simply inject

$$\beta = \frac{M}{\nu(1 - \nu\delta^2)}$$

in the definition of P *and solve the remaining equation wrt* δ

$$P = \frac{\eta\beta(\beta^2 + 3\delta^2 M)}{1 - \eta\delta^3}.$$

Unfortunately the resulting equation appears as a polynomial of degree 3 wrt to the variable δ^2 *hence the solution would result in a pretty complicated form on the Cardan formula which provides the roots of 3rd degree polynomials.*

7.3.4 Branching models

Here $d = 1$ and $E = \mathbb{R}^{D+1}$ for some $D \geq 2$ and we choose again $m = 2$.
Let $\xi_t = \left(\xi_t^{(0)}, \xi_t^{(1)}, \ldots, \xi_t^{(D)}\right)$ be such that

- $\xi_t^{(0)}$ is independent of $\left(\xi_t^{(1)}, \ldots, \xi_t^{(D)}\right)$,
- $\mathbb{E}\xi_t^{(i)}\xi_t^{(j)} = 0$ if $i \neq j$ in case $i, j \geq 1$ and

- $\mathbb{P}(\xi_t^{(0)} \notin \{1, 2, \ldots, D\}) = 0$.

If the functions $M_1, \ldots, M_D$ are Lipschitz on $\mathbb{R}$ satisfying assumptions (7.3) and (7.4) with constants $a_j > 0$ for each $j = 1, \ldots, D$:

$$\begin{array}{rcll} \|M_j(u, \xi_0^{(j)}) - M(v, \xi_0^{(j)})\|_p & \leq & a_j \|u - v\|, & (\forall u, v \in \mathbb{R}^d) \\ \|M_j(u, \xi_0)\|_p & < & \infty. & (\exists u \in \mathbb{R}^d) \end{array}$$

We set

$$M\left(u, \left(z^{(1)}, \ldots, z^{(D)}\right)\right) = \sum_{j=1}^{D} M_j(u, z^{(j)}) \mathbb{I}(z^{(0)} = j),$$

for $(z^{(0)}, \ldots, z^{(D+1)}) \in \mathbb{R}^D$.
The previous contraction assumption writes with the Euclidean norm $\| \cdot \|$ if

$$a = \sum_{j=1}^{D} a_j \mathbb{P}(\xi_0^{(0)} = j) < 1.$$

Now in case $p = 2$ we also improve the result in case $\mathbb{E}\xi_0^{(j)} = 0$ and we denote

$$a^2 = \sum_{j=1}^{D} a_j^2 \, \mathbb{P}(\xi_0^{(0)} = j) < 1.$$

For example in case $M_j(u, z) = A_j(u) + z$ we have $a_j = \mathrm{Lip}\, A_j$:

- if $D = 2$ and $\xi_t^{(1)} \sim b(p)$ are independent and Bernoulli distributed and independent of the centered and independent identically distributed real valued sequence $\xi_t^{(2)} \in \mathbb{L}^2$ and $M_1(u, z) = u + z$, $M_2(u, z) = z$ the previous relation holds if $p < 1$.
  This model is defined through the equation

$$X_n = \begin{cases} X_{n-1} + \xi_n^{(2)}, & \text{if} \quad \xi_n^{(1)} = 1 \\ \xi_n^{(2)}, & \text{if} \quad \xi_n^{(1)} = 0 \end{cases}$$

  Its trajectories are simulated in Figure 7.5.

  Exercise 20. *Identify the parameters* (p, μ), $p = \mathbb{P}(\xi_t^{(1)} = 1)$ *and* $\mu = \mathbb{E}\xi_t^{(2)}$, *in this model.*

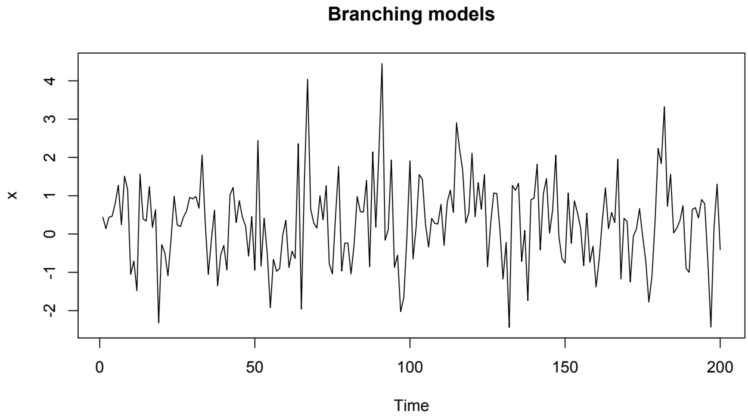


Figure 7.5: A trajectory of a switching model.

Hints. Use a moment method for $m = \mathbb{E}X_0$ and $M = \mathbb{E}X_0^2$. Set $q = 1 - p$ and $\nu = \mathbb{E} = \mathbb{E}(\xi_t^{(2)})^2$ (assumed to be known) then

$$m = q\mu + p(\mu + m) = \mu + pm \Rightarrow m = \frac{\mu}{q}$$

and

$$M = q\nu + p(M + 2\mu m + \nu) = pM + \nu + 2\mu^2 \frac{p}{q}$$

then

$$M = \frac{\nu}{q} + 2\mu^2 \frac{\nu}{q^2}.$$

Thus relation $q = \frac{\mu}{m}$ entails $M = \nu m/\mu + 2\nu m^2$

$$\mu = \frac{\nu m}{M - 2\nu m^2}, \qquad p = 1 - \frac{\nu}{M - 2\nu m^2}$$

expressions respectively fitted by

$$\widehat{\mu} = \frac{\nu \widehat{m}}{\widehat{M} - 2\nu \widehat{m}^2}, \qquad \widehat{p} = 1 - \frac{\widehat{\nu}}{\widehat{M} - 2\nu \widehat{m}^2}$$

The ergodic theorem proves the consistency of such estimates.
$\blacksquare$

- if $D = 3$ and $\xi_0^{(1)} = 1 - \xi_0^{(2)} \sim b(p)$ is again independent of the centered random variable $\xi_0^{(3)} \in \mathbb{L}^2$ we get random regime models if $A_3 \equiv 1$ and the contraction condition writes if $\mathbb{E}\left|\xi_0^{(3)}\right|^2 < \infty$ as
$$a = p\left(\mathrm{Lip}(A_1)\right)^2 + (1-p)\left(\mathrm{Lip}(A_2)\right)^2 < 1$$
This model is defined through the recursion
$$X_n = \begin{cases} A_1(X_{n-1}) + \xi_n^{(3)}, & \text{if} \quad \xi_n^{(1)} = 1 \\ A_2(X_{n-1}) + \xi_n^{(3)}, & \text{if} \quad \xi_n^{(1)} = 0 \end{cases}$$

7.3.5 Integer valued autoregressions

Definition 7.3.2. *Let $\mathbf{P}(a)$ denote a family of integer valued distributions with mean a. The Steutel-van Harn (or Thinning) operator is defined if $x \in \mathbb{N}$ as*

$$a \circ x = \sum_{i=1}^{x} Y_i, \qquad \textit{for } x \geq 1, \textit{ and } 0 \textit{ else.}$$

for a sequence of independent identically distributed random variables with marginal distribution $Y_i \sim \mathbf{P}(a)$. The random variables Y_i are also assumed to be context free, ie. independent of any past history.

For example Galton-Watson processes with immigration (naturally called INAR(1)–process) fits the simple equation

$$X_t = a \circ X_{t-1} + \zeta_t$$

for another independent identically distributed and integer valued sequence (ζ_t) independent of this operator. This model is simulated in Figure 7.6. This means that for an independent identically distributed triangular array $(Y_{t,i})_{t\in\mathbb{Z}, i\in\mathbb{N}}$ we have

$$X_t = \sum_{i=1}^{X_{t-1}} Y_{t,i} + \zeta_t$$

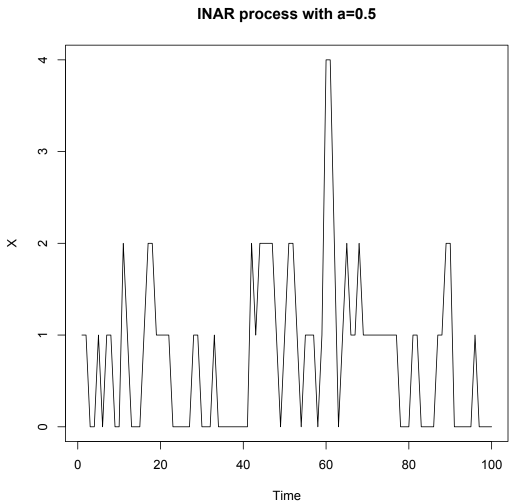


Figure 7.6: INAR(1)–model.

Hence we again write this as a model with independent and identically distributed innovations $(\xi_t)_{t\in\mathbb{Z}}$

$$X_t = M(X_{t-1}, \xi_t) \qquad \text{with} \qquad \xi_t = ((Y_{t,i})_{i\geq 1}, \zeta_t)$$

Here $M(0, \xi_0) = \zeta_0$ hence $\|M(0, \xi_0)\|_p = \|\zeta_0\|_p$.
Now for $y > x$ and $p \geq 1$ we derive

$$M(y, \xi_0) - M(x, \xi_0) = \sum_{x+1}^{y} Y_i$$

thus

$$\|M(y, \xi_0) - M(x, \xi_0)\|_p \leq a|y - x|.$$

Many other integer models write the same idea.
E.g. the bilinear model

$$X_t = a \circ X_{t-1} + b \circ (X_{t-1}\zeta_t) + \zeta_t$$

As an exercise, one may check assumptions on a, b and on ζ_0's distribution such that the assumptions of the previous theorem hold.
Integer valued extensions of AR(p) processes are also easy to define as well as vector valued models.

7.3.6 Generalized Linear Models

Another way to produce attractive classes of integer valued models follows the same lines as for AR-ARCH models. Generalized Linear

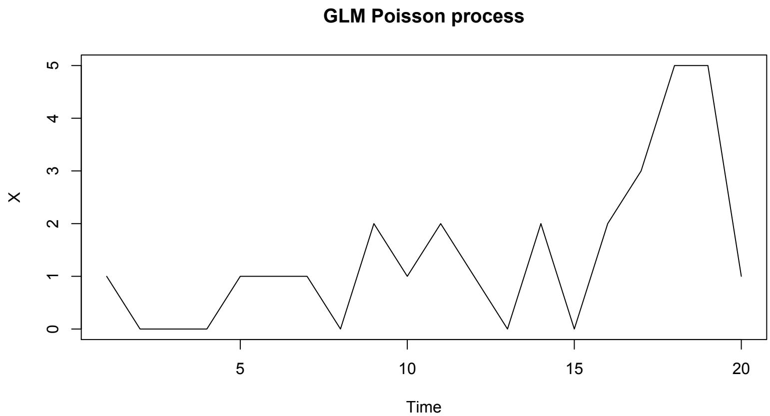


Figure 7.7: A GLM-Poisson trajectory.

Models (GLM) are derived produced from [Kedem and Fokianos, 2002].
Assume that $(V(u))_{u\in\mathbb{U}}$ is a process defined on a Banach space $\mathbb{U}$ equipped with a norm $\|\cdot\|$ and $f : E \times \mathbb{U} \to E$ is a function then

$$X_t|\mathcal{F}_{t-1} \sim V(U_t), \qquad U_t = f(X_{t-1}, U_{t-1})$$

and $\mathcal{F}_{t-1} = \sigma(Z_s/s < t)$ denotes the historical filtration associated to the process $Z_t = (X_t, U_t)$.
The usual way to define ARCH models follows with $\mathbb{U} = \mathbb{R}$, $V = W$

(the Brownian motion) and

$$f(x,u) = \sqrt{\beta + \gamma^2 x^2} \cdot u.$$

Set $\mathcal{P}(\lambda)$ the Poisson distribution with parameter λ Analogously Poisson GLM models (integer valued) are defined as:

$$X_t \,|\, \mathcal{F}_{t-1} \;\sim\; \mathcal{P}(\lambda_t), \qquad \lambda_t = f(X_{t-1}, \lambda_{t-1})$$

Definition 7.3.3 (Poisson processes)**.** *A unit Poisson process is a process $(P(\lambda))_{\lambda\ge 0}$ such that*

- $P(\lambda) \sim \mathcal{P}(\lambda)$ *follows a Poisson distribution with parameter λ,*
- *It satisfies moreover that $P(\lambda) - P(\mu)$ is independent of the sigma-field $\sigma(P(\nu);\ \nu \le \mu)$ if $\lambda > \mu \ge 0$, and*
- *The distribution of $P(\lambda) - P(\mu)$ is $\mathcal{P}(\lambda - \mu)$ for $\lambda > \mu \ge 0$.*

A simple solution of the previous equation writes as a recursive system

$$X_t = P_t(\lambda_t), \qquad \lambda_t = f(X_{t-1}, \lambda_{t-1}) \tag{7.9}$$

for some independent identically distributed sequence P_t of unit Poisson processes.
Note that X_t is not Markov and that either (λ_t) or $Z_t = (X_t, \lambda_t)$ are Markov processes - equivalently iterative systems $X_t = M(X_{t-1}, \xi_t)$.
As an exercise on may check the existence of L^1 solutions of those processes represented with affine function $f(x,\ell) = a + bx + c\ell$ in Figure 7.7.
A main point relies on the fact that for any Poisson process

$$|P(u) - P(v)| \sim P(|u - v|)$$

Consider the bivariate model $Z_t = (X_t, \lambda_t)$ on $\mathbb{R}^+ \times \mathbb{N} \subset \mathbb{R}^2$ equipped with the norm $\|(u,\ell)\| = |u| + \epsilon|\ell|$ for a given parameter $\epsilon > 0$. Now for $Z \in \mathbb{R}^2$ a random vector we get, $\|Z\|_1 = \mathbb{E}\|\mathbb{Z}\|$.
Then this GLM model writes with

$$M((x,\ell); P) = (P(f(x,\ell)), f(x,\ell)).$$

Then it is possible to check assumptions of this Section:

- $M((0,0),P)$ is a vector with a first random coordinate $\mathcal{P}(f(0,0))$ and and a deterministic second coordinate $f(0,0)$: it thus admits moments eg. with order 1.

- $$\begin{aligned} M((x,\ell),P) &- M((x',\ell');P) \\ &= (P(f(x,\ell)) - P(f(x',\ell')), f(x,\ell) - f(x',\ell')), \end{aligned}$$

  thus

  $$\|M((x,\ell),P) - M((x',\ell');P)\|_1 = (1+\epsilon)|f(x,\ell) - f(x',\ell')|.$$

  If the function f is Lipschitz with

  $$|f(x,\ell) - f(x',\ell')| \le a|x-x'| + b|\ell - \ell'|$$

  then relations $(1+\epsilon)a < 1$ and $(1+\epsilon)b < \epsilon$ imply together the relation (7.3).

Then, some cases may be considered:

- the stability holds if $\operatorname{Lip} f < \frac{1}{2}$ (set $\epsilon = 1$).
- if $f(x,\ell) \equiv g(x)$ only depends on x (analogously to ARCH-cases), the stability condition holds if $\operatorname{Lip} g < 1$ ($\epsilon = 0$).
- if $f(x,\ell) \equiv g(\ell)$ only depends on ℓ (analogously to the MA-case), the stability condition holds if $\operatorname{Lip} g < 1$ (with a large ϵ).

Exercise 21. *In equation (7.9) consider the function $f(x,\ell) = a + bx + c\ell$. Assume that coefficients are such that a stationary solution of the equation $(X_t, \lambda_t)_t$ exists in $\mathbb{L}^2$. Then $\mathbb{E}X_0 = \mathbb{E}\lambda_{=\frac{a}{b+c}}0$. Set $\mu = \mathbb{E}X_0$ then $\mu = a + (b+c)\mu$. As before such considerations are useful to fit the model.*

7.3.7 Non-linear AR(d)-models

The (real valued) non linear auto-regressive model with order d writes:

$$X_t = r(X_{t-1}, \ldots, X_{t-d}) + \xi_t, \tag{7.10}$$

The vector valued sequence $U_n = (X_n, X_{n-1}, \ldots, X_{n-d+1})$ writes as a Markov models with values in $\mathbb{R}^d$. Here $E = \mathbb{R}$ and

$$M(u_1, \ldots, u_d, z) \;=\; A(u_1, \ldots, u_d) + (1, 0, \ldots, 0)z,$$

where

$$A(u_1, \ldots, u_d) = \big(r(u_1, \ldots, u_d), u_1, \ldots, u_{d-1}\big).$$

Theorem 7.3.2. *Assume* $\mathbb{E}|\xi_0|^m < \infty$ *and*

$$|r(u_1, \ldots, u_d) - r(v_1, \ldots, v_d)| \;\le\; \sum_{i=1}^{d} a_i |u_i - v_i|$$

for $a_1, \ldots, a_d \ge 0$ *such that*

$$\alpha^d \equiv \sum_{i=1}^{d} a_i < 1.$$

Then equation 7.10 admits a stationary solution and this solution is in $\mathbb{L}^m$*.*

Proof. Define a norm on $\mathbb{R}^d$ by

$$\|(u_1, \ldots, u_d)\| = \max\{|u_1|, \alpha|u_2|, \ldots, \alpha^{d-1}|u_d|\}.$$

For $u = (u_1, \ldots, u_d)$, $v = (v_1, \ldots, v_d) \in \mathbb{R}^d$ set $w_j = |u_j - v_j|$ for $j = 1, \ldots, d$:

$$\begin{aligned}
\|A(u) - A(v)\| &\le \max\left\{\alpha^d \max\{w_1, \ldots, w_d\}, \alpha w_1, \ldots, \alpha^{d-1} w_{d-1}\right\} \\
&\le \alpha \max\left\{\alpha^{d-1} \max\{w_1, \ldots, w_d\}, w_1, \ldots, \alpha^{d-2} w_{d-1}\right\} \\
&\le \max\{w_1, \alpha w_2, \ldots, \alpha^{d-1} w_d\} \equiv \alpha\|u - v\|.
\end{aligned}$$

Duflo condition (7.3) thus holds with $a = \alpha^m < 1$. ∎

7.4 Bernoulli schemes

The following approach to time series modeling is definitely simpler and sharper but it is also less intuitive so that it is set only at the end of the chapter.

7.4.1 Structure and tools

Definition 7.4.1 (unformal definition). *Such models write*

$$X_n = H(\xi^{(n)}), \qquad \textit{with} \qquad \xi^{(n)} = (\xi_{n-t})_{t\in\mathbb{Z}}. \tag{7.11}$$

the function H is thus defined $E^{\mathbb{Z}} \to \mathbb{R}$ and $\xi^{(n)} = (\xi_{n-k})_{k\in\mathbb{Z}}$ is again an independent identically distributed sequence with a shifted time index.
Suppose $\xi = (\xi_k)_{k\in\mathbb{Z}}$ to take values in a measurable space $(E, \mathcal{E})$.
We consider some examples of such situations.
An important special case is that of causal processes. $H : E^{\mathbb{N}} \to \mathbb{R}$ and we write in a simpler formulation

$$X_n = H(\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots).$$

Such a stationary process is said causal since the history of X before the epoch n is included in that of ξ. Mathematically expressed this means

$$\sigma(X_s/s \le n) \subset \sigma(\xi_s/s \le n).$$

Fix $e \in E$ we denote $\widehat{\xi}(n)$ the sequence with current element ξ_j, if $|j| \le n$ and e if $|j| > n$. Let $m \ge 1$ a simple condition to define such models writes

$$\sum_{n=1}^{\infty} \zeta_n < \infty, \tag{7.12}$$

with

$$\zeta_n^p = \mathbb{E}\left| H\left(\widehat{\xi}(n)\right) - H\left(\widehat{\xi}(n-1)\right)\right|^p. \tag{7.13}$$

Due to the completeness of the space $\mathbb{L}^p$ a normally convergent series is convergent and:

Proposition 7.4.1. *Let $p \ge 1$ be such that relation (7.12) holds then the sequence $(X_n)_{n\in\mathbb{Z}}$ defined this way is stationary and $\mathbb{L}^p-$valued.*

Proof. This relation indeed implies the convergence $\mathbb{L}^p$ of the well defined sequence $H\left((\xi_j)_{|j|\le n}\right)$.
To prove the result a bit more is needed and on extends the previous remark to the random variable $Z_n = (X_{n+1}, \ldots, X_{n+s}) \in \mathbb{R}^s$.

This is thus the limit of a sequence of $\mathbb{R}^s-$valued random variable with a distribution independent of n. ∎

Example 7.4.1 (Bernouilli shifts)**.**

- *Let* $H : \mathbb{R}^m \to \mathbb{R}$ *the process* $X_n = H(\xi_n, \ldots, \xi_{n-m+1})$ *is an* $m-$*dependent sequence i.e.* $\sigma\{X_j, j < a\}$ *and* $\sigma\{X_j, j > a+m\}$ *are independent* $\sigma-$*algebras.*

- *Stochastic volatility model. Let* $Y_n = H(\xi_n, \xi_{n-1}, \ldots)$ *be a causal Bernoulli scheme such that the independent identically distributed innovations* $\xi_n \in \mathbb{L}^2$ *are centered.*
  Set
  $$X_n = \xi_n Y_{n-1} = \xi_n H(\xi_{n-1}, \xi_{n-2}, \ldots).$$
  The sequence X_n *is orthogonal and*
  $$Var(X_n|\mathcal{F}_{n-1}) = Y_{n-1}^2.$$
  This property indicates possible rapid changes adapted to model the stock exchange.

- *All the previous sections of the present chapter provide us with a series of examples of this situation.*

The previous definition 7.4.1 is really adapted to deal with the previous chaotic examples for which tails may be bounded.
A more general setting is adapted to prove the existence of a stationary processes.

Definition 7.4.2 (formal definition)**.** *Let* μ *a probability distribution on a measurable space* $(E, \mathcal{E})$*.*
Consider an independent identically distributed sequence $(\xi_n)_{n\in\mathbb{Z}}$ *with marginal law* μ*.*
Set $\nu = \mu^{\otimes\mathbb{Z}}$ *the law of* $(\xi_n)_{n\in\mathbb{Z}}$ *on the space* $(E^{\mathbb{Z}}, \mathcal{E}^{\otimes\mathbb{Z}})$*. Then* $\mathbb{L}^p(\nu)$ *is the space of measurable functions* $\nu-a.s$ *defined on* $E^{\mathbb{Z}}$ *and such that*
$$\mathbb{E}|H\left((\xi_n)_{n\in\mathbb{Z}}\right)|^p < \infty$$
An analogue definition holds with $\nu^+ = \mu^{\otimes\mathbb{Z}}$ *the law of* $(\xi_n)_{n\in\mathbb{N}}$ *on the space* $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}})$*.*

Remark 7.4.1. *The spaces $\mathbb{L}^p(\nu)$ and $\mathbb{L}^p(\nu^+)$ are Banach spaces (complete normed vector spaces) equipped respectively with the norms*

$$\begin{aligned} \|H\|_p &= \left(\mathbb{E}|H\left((\xi_n)_{n\in\mathbb{Z}}\right)|^p\right)^{\frac{1}{p}}, && \textit{for the general case,} \\ &= \left(\mathbb{E}|H\left((\xi_n)_{n\in\mathbb{N}}\right)|^p\right)^{\frac{1}{p}}, && \textit{for the causal case.} \end{aligned}$$

The definition of Bernouilli schemes is as in the unformal definition 7.4.1 and holds non-causal or causal schemes for elements $H \in \mathbb{L}^p(\nu)$ or $\mathbb{L}^p(\nu^+)$ respectively.
Moreover condition (7.12) implies with Proposition 7.4.1, a simple sufficient condition for functions of infinitely many variables to exist in those huge spaces.
The next subsection also proves that those assumptions are relevant to prove short range conditions.

Proof of Theorem 7.3.1. A quite simple and elegant proof relies on the previous notions proves moreover that there exists a unique element $H \in \mathbb{L}^p(\nu^+)$ such that a stationary solution of eqn. (7.2) writes

$$X_t = H(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots).$$

To this end consider the application

$$\Phi : \mathbb{L}^p(\nu^+) \to \mathbb{L}^p(\nu^+), \qquad H \mapsto K,$$

with

$$K(v_0, v_1, \ldots) = M(H(v_1, v_2, \ldots), v_0).$$

Conditions (7.4) and (7.3) allow to prove that prove that $K \in L^p(\nu^+)$ if $H \in L^p(\nu^+)$ (for this condition wrt ξ_0 and use triagular inequality). Consider the fixed point e as an element of $\mathbb{L}^p$:

$$\|K\|_p = \mathbb{E}^{1/p}|M(H(\xi_1,\ldots),\xi_0)|^p \le \mathbb{E}^{1/p}|M(e,\xi_0)|^p + a\|H-e\|_p.$$

Now if $H, H' \in L^p(\nu^+)$ then again conditioning with respect to $\xi_1, \xi_2, \ldots$ implies

$$\|K - K'\|_p \le a\|H - H'\|_p.$$

The classical Banach-Picard fixed point theorem thus implies that Φ admits a unique fixed point $H^\star$.

This theorem also implies that the iterates $H_m = \Phi^m \circ H_0$ converge to this fixed point $H^\star$ of Φ. In other words $X_{n,m} = H_m(\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots)$ converge in $\mathbb{L}^p$ to the stationary solution of the process as $m \uparrow \infty$ for each value of n.
Moreover the convergent rate is geometric:

$$\|X_n - X_{n,m}\|_p = \|H_m - H^\star\|_p \le Ca^m. \qquad \blacksquare$$

7.4.2 Coupling

This section explicits ways to couple such Bernouilli shifts. Decorrelation rates are also deduced. This will allow to derive quantitative laws of large numbers for expressions of a statistical interest.
Those ideas are widely developed later in chapter 9 in order to understand how to derive limit theorems in distribution.

Let $(\xi'_k)_{k\in\mathbb{Z}}$ another independent identically distributed sequence independent of $(\xi_k)_{k\in\mathbb{Z}}$ and with the same distribution.
For $n \ge 0$ set $\widetilde{\xi}(n) = (\widetilde{\xi}(n)_k)_{k\in\mathbb{Z}}$ with

$$\widetilde{\xi}(n)_k = \begin{cases} \xi_k, & \text{if} \quad |k| \le n, \\ \xi'_k, & \text{if} \quad |k| > n. \end{cases}$$

Then we set

$$\delta_n^p = \mathbb{E}\left| H\left(\widetilde{\xi}(n)\right) - H(\xi)\right|^p. \tag{7.14}$$

Definition 7.4.3. *Assume that a Bernoulli shift satisfies* $\lim_n \delta_n^{(p)} = 0$ *with the above definition (7.14) then it will be called* $\mathbb{L}^p-$*dependent.*

Remark 7.4.2. *Replace* $\widetilde{\xi}(n)$ *by*

$$\widehat{\xi}(n)_k = \begin{cases} \xi_k, & \textit{if} \quad |k| \ne n, \\ \xi'_k, & \textit{if} \quad |k| = n. \end{cases}$$

leads to the fruitful physical measure of dependence by Wei Biao Wu. The two previous proposals are **couplings** *in the sense that they leave unchanged the marginal distribution of the Bernoulli shift.*
Another alternative is to set

$$\xi'(n)_k = \begin{cases} \xi_k, & \textit{if} \quad |k| \le n, \\ 0, & \textit{if} \quad |k| > n. \end{cases}$$

which is essentially the same as $\widetilde{\xi}(n)$ and makes easy to define functions of infinitely many variables as limits of functions of finitely many variables in the Banach space $\mathbb{L}^p(\nu^+)$ from the fact that $H(\xi'(n))$ is a Cauchy sequence in case

$$\sum_n \|H(\xi'(n)) - H(\xi'(n-1))\|_p < \infty$$

Now as an introduction to weak dependence conditions in § 11.4 we note that:

Proposition 7.4.2 (Decorrelation)**.** *If the stationary process $(X_n)_{n\in\mathbb{Z}}$ satisfies $\mathbb{E}|X_0|^p < \infty$ for $p \geq 2$ and is as before depending if H is unbounded or bounded yields:*

$$\begin{aligned} |Cov(X_0, X_k)| &\leq 4(\mathbb{E}|X_0|^p)^{1/p}\delta_{[k/2]}, \\ &\leq 4\|H\|_\infty \delta^2_{[k/2]}. \end{aligned}$$

If the Bernoulli scheme is causal the previous inequalities write:

$$\begin{aligned} |Cov(X_0, X_k)| &\leq 2(\mathbb{E}|X_0|^p)^{1/p}\delta_k, \\ &\leq 2\|H\|_\infty \delta^2_k. \end{aligned}$$

Remark 7.4.3. *Such results imply short range dependence of the process X in the sense of definition 4.3.1, in case the above covariances are summable.*

Proof. Use Hölder inequality after the relation:

$$\mathrm{Cov}\,(X_0, X_k) = \mathrm{Cov}\,(X_0 - X_{0,l}, X_k) + \mathrm{Cov}\,(X_{0,l}, X_k - X_{k,l}),$$

which holds if $2l \leq k$ when setting $X_{k,l} = H\left(\widetilde{\xi}(l)^{(k)}\right)$.
Recall that $\widetilde{\xi}(l)^{(k)}$ is the sequence whose $j-$element writes ξ_{k-j} if $|j| \leq l$ and ξ'_{k-j} if $|j| > l$.
If the Bernoulli scheme is causal the relation simplifies since

$$\mathrm{Cov}\,(X_0, X_k) = \mathrm{Cov}\,(X_0, X_k - X_{k,k}).$$

Now factors 4 and 2 arise from the fact that covariances are expectations of a product minus the product of expectations: same bounds are provided for both terms. ∎

An important question is the heredity of such quantities through instantaneous images $Y_k = g(X_k)$. A large number of usual statistics of interest write as functions of the process of interest.
Denote the corresponding expressions by $\delta_{k,Y}$ and $\delta_{k,X}$, then:

Lemma 7.4.1. *Assume that $m \geq 1$ and $g : \mathbb{R} \to \mathbb{R}$ satisfies $\mathrm{Lip}\, g \equiv L < \infty$. Set $Y_k = g(X_k)$ then:*

$$\delta_{k,Y} \leq L\delta_{k,X}.$$

In case the function g is not Lipschitz such relations do not hold in a general setting.
Anyway simple indicators $g_x(u) = 1\!\!\mathrm{I}_{\{u \leq x\}}$ are the convenient functions to derive bounds for the empirical process ([3]).
We obtain:

Lemma 7.4.2. *If $p = 2$ and if there exist constants $c, C > 0$ such that on each interval $\mathbb{P}(X \in [a, b]) \leq C|b - a|^c$, then the process defined by $Y_{x,n} = 1\!\!\mathrm{I}_{\{X_n \leq x\}}$ satisfies:*

$$\delta_{k,Y_x} \leq 2(2C)^{2/(c+2)} \delta_{k,X}^{\frac{c}{c+2}}.$$

Proof. Set

$$g_{x,\epsilon} = \begin{cases} 1, & \text{if} \quad u \leq x - \epsilon, \\ 0, & \text{for} \quad u \geq x, \\ \text{is affine}, & \text{else}. \end{cases}$$

Consider $Y_{x,\epsilon,n} = g_{x,\epsilon}(X_n)$
Then

$$|g_{x,\epsilon}(u) - g_{x,\epsilon}(v)| \leq |u - v|/\epsilon \text{ and } \ \delta_{k,Y_{x,\epsilon}} \leq \delta_{k,X}/\epsilon.$$

Moreover

$$|\delta^2_{k,Y_{x,\epsilon}} - \delta^2_{k,Y_x}| \leq 2\mathbb{P}(X_0 \in [x - \epsilon, x]) \leq 2C\epsilon^c.$$

[3] Those are the simplest discontinuous functions. They are classes of functions with only one singularity. More general functions with finitely many discontinuities may be analogously considered.

So

$$\delta^2_{k,Y_{x,\epsilon}} \leq \delta^2_{k,X}/\epsilon^2 + 2C\epsilon^c.$$

We then conclude with $\epsilon^{c+2} = \delta^2_{k,X}/(2C)$. ∎

Up to a constant the result remains valid for a function g Lipschitz-continuous on intervals.

Remark 7.4.4. *A control for the cumulative empirical distribution follows:*

$$Var\, F_n(x) = \mathcal{O}\left(\frac{1}{n}\right), \qquad \textit{if} \qquad \sum_{k=0}^{\infty} \delta_{k,X}^{c/(c+2)} < \infty.$$

In case $c = 1$, which hold for X_0's distribution with a bounded density, the condition writes

$$\sum_{k=0}^{\infty} \delta_{k,X}^{1/3} < \infty.$$

This holds for example in case the marginal law of X_0 admits a bounded density.

Remark 7.4.5. *Quote that higher order moment inequalities are derived from analogue ideas as in Chapter 12.*

Chapter 8

Associated processes

The notion of association or positive correlation was naturally introduced in two different fields reliability [Esary et al., 1967] and statistical physics [Fortuin et al., 1971] to model a tendency that the coordinates of a vector valued random variable admit analogue behaviors.
We defer a reader to the nice paper [Newman, 1984] for more details.
This notion deserves much attention since it provides a class of random variables for which independence and orthogonality coincide (as for the Gaussian case).
The notion of independence is more related to $\sigma-$algebras but in those two cases it is related to the geometric notion of orthogonality.
Those remarks are of a wide interest for modeling dependence as this is the aim of the forthcoming chapter 9.

8.1 Association

Definition 8.1.1. *A random vector $X \in \mathbb{R}^p$ is associated if for all measurable functions $f, g : \mathbb{R}^p \to \mathbb{R}$, with $\mathbb{E}|f(X)|^2 < \infty$ and $\mathbb{E}|g(X)|^2 < \infty$ such that f, g are coordinatewise non decreasing*

$$Cov(f(X), g(X)) \geq 0$$

Definition 8.1.2. *A random process $(X_t)_{t\in\mathbb{T}}$ is associated is the vector $(X_t)_{t\in F}$ is associated for each finite $F \subset \mathbb{T}$.*

Remark 8.1.1. *Covariances of an associated process are non negative if this process is square integrable.*

Example 8.1.1. *Now we will present a worked example and WE WILL SHOW that it fits weak dependence conditions in the forthcoming chapter.*

A real random variable is always associated indeed if X' is an independent copy of X then calculus proves that

$$\mathrm{Cov}\,(f(X), g(X)) = \frac{1}{2}\mathbb{E}(f(X) - f(X')(g(X) - g(X')).$$

Hence for f, g monotonic this expression is nonnegative.
More generally:

Theorem 8.1.1 ([Newman, 1984])**.** *Independent vectors are associated.*

Theorem 8.1.2 ([Newman, 1984])**.** *A limit in distribution of a sequence of associated vectors is associated.*

Proof. A recursion is needed. A careful conditioning is needed. For this one needs to prove that

Lemma 8.1.1. *Let $Z = (X, Y) \in \mathbb{R}^{p+q}$ and $f : g : \mathbb{R}^{p+q} \to \mathbb{R}$ such that $f(Z)$ and $g(Z) \in \mathbb{L}^2$. If X, Y are independent vectors then $F(x) = \mathbb{E}f(x, Y)$ and $G(x) = \mathbb{E}g(x, Y) \in \mathbb{L}^2$ for a.s. each $x \in \mathbb{R}^p$.*

Remark 8.1.2. *In this case by setting $U(x) = Cov(f(x, Y), g(x, Y))$ we derive:*

$$Cov(f(Z), g(Z)) = \mathbb{E}U(x) + Cov(F(X), G(X)).$$

Hint. From Cauchy-Schwarz inequality one derive $F(X), G(X) \in \mathbb{L}^2$.
∎

8.2 Associated processes

Definition 8.2.1. *A process $(X_t)_{t\in\mathbb{T}}$ is associated if for each $S \subset \mathbb{T}$ finite, the vector $(X_t)_{t\in S}$ is associated.*

Remark 8.2.1. *Heredity properties of association is very important to handle applications involving associated processes.*

Example 8.2.1. *The following example inherit association properties*

- *A non-decreasing image of an associated sequence. Such heredity property admits many consequences:*
- *A LARCH(∞)–model with nonnegative coefficients $a_j \geq 0$ and inputs $\xi_j \geq 0$:*

$$X_t = \Big(a_0 + \sum_{j=1}^{\infty} a_j X_{t-j}\Big)\xi_t.$$

  To check this, use a recursion, the point that a linear function

$$(x_1, \ldots, x_p) \mapsto \sum_{j=1}^{p} b_j x_j,$$

  with nonnegative coefficients $b_j (= a_j \xi_t)$ is non-decreasing and the fact that association is stable under limits in distribution.

- *An autoregressive*
- *non-linear AR–model process solution of an equation*

$$X_t = r(X_{t-1}, \ldots, X_{t-p}) + \xi_t,$$

  if the function $r : \mathbb{R}^p \to \mathbb{R}$ is a coordinatewise non-decreasing function,

- *INAR models*

$$X_t = a \circ X_{t-1} + \epsilon_t,$$

  or more general Integer Bilinear models

$$X_t = a \circ X_{t-1} + b \circ (\epsilon_{t-1} X_{t-1}) + \epsilon_t,$$

  are associated if $\epsilon_t \geq 0$ is iid and integer valued, and if $a\circ$ and $b\circ$ are both thinning operators with non-negative random variables.
  Indeed one may write $(X_1, \ldots, X_n)$ as a monotonic function of independent sequences (thus associated).

8.3 A main inequality

A new concept is needed

Definition 8.3.1. *Let $f, f_1 : \mathbb{R}^p \to \mathbb{R}$ the we set $f \ll f_1$ if both function $f \pm f_1$ are coordinatewise nondecreasing.*

Example 8.3.1. *Assume that the function f satisfies*

$$|f(y) - f(x)| \le a_1|y_1 - x_1| + \cdots + a_p|y_p - x_p|$$

for all vectors $x = (x_1, \dots, x_p)$, $y = (y_1, \dots, y_p) \in \mathbb{R}^p$.
Then $f \ll f_1$ if one sets

$$f_1(x) = a_1x_1 + \cdots + a_px_p$$

Proof. In order to prove this only work out inequalities by grouping terms invoking x's or y's only:

$$-a_1(y_1-x_1)-\cdots-a_p(y_p-x_p) \le f(y)-f(x) \le a_1(y_1-x_1)+\cdots+a_p(y_p-x_p).$$

The previous inequalities apply to vectors x, y such that $x_i = y_i$ excepted for only one index $1 \le i \le p$.
The corresponding inequalities exactly write $f \ll f_1 \dots$ ∎

An essential inequality follows:

Lemma 8.3.1 ([Newman, 1984])**.** *Let $X \in \mathbb{R}^p$ be an associated random vector and f, g, f_1, g_1 be measurable functions $\mathbb{R}^p \to \mathbb{R}$ then:*

$$|Cov(f(X), g(X))| \le Cov(f_1(X), g_1(X)),$$

if those function are such that $f(X), g(X), f_1(X), g_1(X) \in \mathbb{L}^2$ and $f \ll f_1$, $g \ll g_1$.

Proof. The 4 covariances

$$\mathrm{Cov}\Big(f(X) + af_1(X), g(X) + bg_1(X)\Big)$$

are non negative if $a, b = -1$ or 1, then add them 2 by 2 yields the result.

For this consider separetely the cases
• $ab = -1$ and
• $ab = 1$
corresponding to the couples $(a, b) = (-1, 1)$, $(1, -1)$ and $(1, 1)$, $(-1, -1)$, respectively. ∎

As a simple byproduct of the previous Lemma implies with Example 8.3.1 we derive the nice following result:

Theorem 8.3.1. *Let $(Y, Z) \in \mathbb{R}^u \times \mathbb{R}^v$ be an associated vector in $\mathbb{L}^2$. If for some constants $a_1, \ldots, a_u, b_1, \ldots, b_v \geq 0$, the functions f and g satisfy respectively:*

$$\begin{aligned} |f(y) - f(y')| &\leq \sum_{i=1}^{u} a_i |y_i - y_i'|, \quad \forall y, y' \in \mathbb{R}^u, \\ |g(z) - g(z')| &\leq \sum_{j=1}^{v} b_j |z_j - z_j'|, \quad \forall z, z' \in \mathbb{R}^v, \end{aligned}$$

then:

$$|Cov(f(Y), g(Z))| \leq \sum_{i=1}^{u} \sum_{j=1}^{v} a_i b_j \, Cov(Y_i, Z_j) \tag{8.1}$$

Remark 8.3.1. *We thus derive that for each associated random vector in $\mathbb{L}^2$:*

- **Independence**
  If the vectors Y, Z admit pairwise orthogonal components then they are stochastically independent as for the Gaussian case.

- **Quasi-independence**

$$\begin{aligned} |Cov(f(Y), g(Z))| &\leq Lip\, f \cdot Lip\, g \sum_{i=1}^{u} \sum_{j=1}^{v} Cov(Y_i, Z_j) \\ &\leq uv Lip\, f \cdot Lip\, g \max_{1 \leq i \leq u} \max_{1 \leq j \leq v} Cov(Y_i, Z_j) \end{aligned}$$

  This inequality means that the asymptotic dependence structure of an associated random vector relies on its second order structure.

This conducted to the definition of weak dependence and proves that κ-weak dependence holds for associated models (see chapter 11).

8.4 Limit theory

[Newman, 1984] proved the following elegant and powerful weak invariance principle.

Theorem 8.4.1 (Newman, 1984)**.** *If the condition*

$$\sigma^2 = \sum_{n=-\infty}^{\infty} Cov(X_0, X_n) < \infty,$$

holds for the stationary and associated process $(X_n)_{n\in\mathbb{Z}}$ then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} X_k \to \sigma W_t \quad \text{in the Skohorod space } D[0,1].$$

Note that the condition precisely extends that obtained for the independent identically distributed case since it reduces to $\mathbb{E}X_0^2 < \infty$ in this case. It cannot be improved which also makes such conditions so attractive.

Part III

Dependences

A first chapter in this part begins with the (central) ergodic theorem which asserts that the strong law of large numbers (SLLN) works for the partial sum process of most of the previously introduced models. Assume that an unknown parameter for $(X_n)_{n\in\mathbb{Z}}$, a stationary sequence, is $\theta = \mathbb{E}X_0$, this writes as:

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \to_{n\to\infty} \theta, \qquad a.s.$$

The question of convergence rates in this results is solved in the forthcoming dependence types for stationary sequences.
Two additional chapter detail as much as possible more precise asymptotic results useful for statistical applications.
According to the fact that they are either LRD or SRD very different asymptotic behaviors will be seen to occur including corresponding rates.

$$n^\alpha(\overline{X}_n - \theta) \to_{n\to\infty} Z, \text{in distribution}$$

with $\alpha = \frac{1}{2}$, or $> \frac{1}{2}$ according to the fact that SRD or LRD holds. Asymptotic confidence bounds may thus be derived. Namely set $\tau > 0$ a confidence level then in case there there exists $t > 0$ with $\mathbb{P}(Z \le t) = 1 - \tau$, then:

$$\mathbb{P}(\theta \in [\overline{X}_n - tn^{-\alpha}, \overline{X}_n + tn^{-\alpha}]) \to_{n\to\infty} \tau.$$

This also yields to goodness of fit tests for this mean parameter θ.

Chapter 9

Dependence

9.1 Ergodic theorem

We thank Jérôme Dedecker for the present presentation.

Definition 9.1.1. *A transformation $T : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is bijective bi-measurable and $\mathbb{P}$-invariant if it is bijective, measurable, admits a measurable inverse and moreover $\mathbb{P}(T(A)) = \mathbb{P}(A)$ for all $A \in \mathcal{A}$.*
Note

$$\mathcal{I} = \{A \in \mathcal{A} \ / \ T(A) = A\}$$

the sub-sigma algebra of $\mathcal{A}$ containing all the $T-$invariant events.
A transformation is ergodic if $A \in \mathcal{I}$ implies $\mathbb{P}(A) = 0$ or 1.

Remark 9.1.1 (Link to stationary processes)**.** *Let $X = (X_n)_{n\in\mathbb{Z}}$ be a real valued stationary process defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the image $\mathbb{P}_X$ is a probability on the space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}})\big)$. The sigma-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$ is generated by elementary events:*

$$A = \prod_{k\in\mathbb{Z}} A_k \qquad \text{with } A_k = \mathbb{R} \quad \text{excepted for finitely many indices } k.$$

The transformation T defined by

$$T(x)_i = (x_{i+1}) \qquad \text{for} \qquad x = (x_i)_{i\in\mathbb{Z}}$$

satisfies

$$T\left(\prod_{k\in\mathbb{Z}} A_k\right) = \prod_{k\in\mathbb{Z}} A_{k+1}.$$

It is bijective bimeasurable and $\mathbb{P}$-invariant ; it is called the **shift operator**.
Note $\mathcal{J} = X^{-1}(\mathcal{I})$ the sigma-algebra image of $\mathcal{I}$ through X.
If T is ergodic (i.e. $\mathbb{P}(A) = 0$ or 1 if $A \in \mathcal{J}$) then the process $X = (X_n)_{n\in\mathbb{Z}}$ is ergodic.

Example 9.1.1 (A non ergodic process). *A very simple example of notn ergodic process is $X_t \equiv \zeta$ for each t and for a non-constant rv. Refining it to $X_t = \xi_t + \zeta$,for each $t \in \mathbb{Z}$ provides a non trivial example if (ξ_t) is independent identically distributed and independent of ζ. In order to make it evident just assume e.g. that $\xi_t \geq 0$ and $\zeta < 0$, a.s. Many other examples may be found in [Kallenberg, 1997].*

Proposition 9.1.1. *Let T be a bijective and bi-measurable $\mathbb{P}$-invariant transformation.*
Let $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be measurable with $\mathbb{E}f^2 < \infty$ then

$$R_n(f) = \frac{1}{n}\sum_{k=1}^{n} f \circ T^k \xrightarrow{\mathbb{L}^2}_{n\to\infty} \mathbb{E}^{\mathcal{I}} f.$$

Proof of Proposition 9.1.1. Let C denote the closure (in $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$) of the convex hull $\mathcal{C}$ of

$$E = \{f \circ T^k \ / \ k \in \mathbb{Z}\}, \qquad (^1).$$

From the orthogonal projection theorem (see e.g. théorème 3.81, page 124 in [Doukhan and Sifre, 2001]) there exists a unique $\overline{f} \in C$ with

$$\|\overline{f}\|_2 = \inf\{\|g\|_2 \ / \ g \in C\}.$$

If one prove

$$\|R_n(f)\|_2 \to_{n\to\infty} \|\overline{f}\|_2$$

[1] $C = \overline{\mathcal{C}}$ with

$$\mathcal{C} = \left\{\sum_{i=1}^{I} a_i x_i;\ a_i \geq 0, x_i \in E, \sum_{i=1}^{I} a_i = 1, I \geq I\right\}$$

then the proof of the projection theorem implies also

$$\|R_n(f) - \overline{f}\|_2 \to_{n\to\infty} 0.$$

Moreover

$$R_n(f) = f + R_{n-1}(f) \circ T.$$

Hence

$$\|\overline{f} \circ T - \overline{f}\|_2 \leq \|\overline{f} \circ T - R_{n-1}(f) \circ T\|_2 + \frac{1}{n}\|f\|_2 + \|R_n(f) - \overline{f}\|_2.$$

$\mathbb{P}$-invariance of T implies that the frist term in the right hand member of this inequality writes $\|\overline{f} - R_{n-1}(f)\|_2 \to 0$. Thus

$$\overline{f} \circ T = \overline{f}.$$

Thus $\overline{f}$ is $\mathcal{I}$-measurable.
Since

$$R_n(f) \to \overline{f}, \quad \text{in} \quad \mathbb{L}^2$$

we also deduce

$$\mathbb{E}^{\mathcal{I}} R_n(f) \to \mathbb{E}^{\mathcal{I}}\, \overline{f} = \overline{f}.$$

The fact that $\mathbb{E}^{\mathcal{I}} R_n(f) = \mathbb{E}^{\mathcal{I}} f$ allows to conclude. ∎

In order to prove

$$\|R_n(f)\|_2 \to_{n\to\infty} \|\overline{f}\|_2$$

consider a convex combination

$$g = \sum_{|j|\leq k} a_j f \circ T^j \in \mathcal{C}, \quad \text{with} \quad \|g\|_2 \leq \|\overline{f}\|_2 + \epsilon.$$

With the invariance of T we derive

$$\|R_n(g)\|_2 \leq \|g\|_2 \leq \|\overline{f}\|_2 + \epsilon.$$

From another hand

$$\begin{aligned}
\|R_n(f-g)\|_2 &= \left\| \sum_{j=-k}^{k} a_j (R_n(f) - R_n(f \circ T^j)) \right\|_2 \\
&\leq \sum_{j=-k}^{k} a_j \|R_n(f) - R_n(f \circ T^j)\|_2
\end{aligned}$$

and using again T's invariance,

$$\begin{aligned} \|R_n(f) - R_n(f \circ T^j)\|_2 &\leq \frac{1}{n} \sum_{i=k+1}^{k+j} (\|f \circ T^j\|_2 + \|f \circ T^{-j}\|_2) \\ &\leq \frac{2j}{n} \|f\|_2 \end{aligned} \tag{9.1}$$

Thus

$$\|R_n(f-g)\|_2 \leq \sum_{|j|\leq k} \frac{2ja_j}{n} \|f\|_2 \leq \frac{2k}{n} \|f\|_2 \to_{n\to\infty} 0.$$

Hence

$$\|\overline{f}\|_2 \leq \limsup_n \|R_n(f)\|_2 \leq \|\overline{f}\|_2 + \epsilon$$

yielding the result. ∎

Corollary 9.1.1. *If we only assume $\mathbb{E}|f| < \infty$ then*

$$R_n(f) \xrightarrow{\mathbb{L}^1}_{n\to\infty} \mathbb{E}^{\mathcal{I}} f.$$

Proof. There exists a sequence $g_m \in \mathbb{L}^2$ such that $\|g_m - f\|_1 \to_{m\to\infty} 0$ (it is even possible to assume that $g_m \in \mathbb{L}^\infty$).
Then

$$\begin{aligned} \|R_n(f) - \mathbb{E}^{\mathcal{I}} f\|_1 &\leq \|R_n(f - g_m)\|_1 + \|R_n(g_m) - \mathbb{E}^{\mathcal{I}}(g_m)\|_1 \\ &\quad + \|\mathbb{E}^{\mathcal{I}}(g_m - f)\|_1 \\ &\leq 2\|f - g_m\|_1 + \|R_n(g_m) - \mathbb{E}^{\mathcal{I}}(g_m)\|_1. \end{aligned}$$

The previous Proposition implies

$$\limsup_n \|R_n(f) - \mathbb{E}^{\mathcal{I}} f\|_1 \leq 2\|f - g_m\|_1.$$

The conclusion follows from a limit argument $m \to \infty$. ∎

The ergodic theorem (aim of this section) is also based upon the next inequality

Lemma 9.1.1 (Hopf maximal inequality)**.** *Let T be a bijective bi-measurable and $\mathbb{P}$-invariant transformation.*
For $f \in \mathbb{L}^1$ set $S_0(f) = 0$ and, for $k \geq 1$ set:

$$S_k(f) = \sum_{j=1}^{k} f \circ T^j, \qquad S_n^+(f) = \max_{0 \leq k \leq n} S_k(f).$$

Then:

$$\mathbb{E}\left(f \circ T \cdot \mathbb{1}_{S_n^+(f)>0}\right) \geq 0.$$

Proof of Lemma 9.1.1. If $1 \leq k \leq n+1$ then

$$S_k(f) \leq f \circ T + S_n^+(f) \circ T.$$

Moreover if $S_n^+(f) > 0$ then

$$S_n^+(f) = \max_{1 \leq k \leq n} S_k(f).$$

Thus

$$S_n^+(f)\mathbb{1}_{S_n^+(f)>0} \leq f \circ T \mathbb{1}_{S_n^+(f)>0} + S_n^+(f) \circ T \, \mathbb{1}_{S_n^+(f)>0}.$$

This entails

$$f \circ T \, \mathbb{1}_{S_n^+(f)>0} \geq (S_n^+(f) - S_n^+(f) \circ T) \, \mathbb{1}_{S_n^+(f)>0}.$$

Now

$$\mathbb{E} f \circ T \, \mathbb{1}_{S_n^+(f)>0} \geq \mathbb{E} S_n^+(f) - \mathbb{E} S_n^+(f) \circ T = 0. \qquad \blacksquare$$

Corollary 9.1.2. *Assume that assumptions in Lemma 9.1.1 hold then*

$$\mathbb{P}\left(\sup_{n \geq 1} |R_n(f)| > c\right) \leq \frac{\mathbb{E}|f|}{c}, \qquad \forall c > 0.$$

Proof. Apply Lemma 9.1.1 to $f - c$:

$$\mathbb{E}\big(f - c\big) \circ T \mathbb{1}_{S_n^+(f-c)>0} \geq 0.$$

Hence

$$\frac{\mathbb{E} f \vee 0}{c} \geq \frac{\mathbb{E} f \circ T \mathbb{1}_{S_n^+(f-c)>0}}{c} \geq \mathbb{P}(S_n^+(f - c) > 0).$$

Thus

$$S_n^+(f-c) = 0 \vee \max_{1\le k\le n} k\left(R_k(f)-c\right) \ge \max_{1\le k\le n}\left(R_k(f)-c\right).$$

Hence

$$\frac{\mathbb{E}f\vee 0}{c} \ge \mathbb{P}\left(\max_{1\le k\le n}\left(R_k(f)-c\right) > 0\right).$$

Replace f by $-f$ one proves analogously:

$$\frac{-(\mathbb{E}f\wedge 0)}{c} \ge \mathbb{P}(S_n^+(f+c)<0) \ge \mathbb{P}\left(\max_{1\le k\le n}\left(R_k(f)+c\right) < 0\right).$$

The result follows from summing up the previous inequalities and for $n\to\infty$.
Indeed $|f| = f\vee 0 - f\wedge 0$ and $\mathbb{P}(R-c>0)+\mathbb{P}(R+c<0) = \mathbb{P}(|R|>c)$ for each random variable R. ∎

Theorem 9.1.1 (Ergodic theorem)**.** *Let T bijective bi-measurable and $\mathbb{P}$-invariant. Let $f\in\mathbb{L}^1$ then*

$$R_n(f) \to_{n\to\infty} \mathbb{E}^{\mathcal{I}}f, \quad a.s.$$

If the process is ergodic the limit is constant almost everywhere for any integrable f.

Proof of Theorem 9.1.1. Assume first that g is bounded. If $n,m\ge 1$ then

$$\left|R_n(g)-\mathbb{E}^{\mathcal{I}}g\right| \le |R_n(g-R_m(g))| + \left|R_n(R_m(g)-\mathbb{E}^{\mathcal{I}}g)\right|.$$

Using the same idea as to derive inequality (9.1) we obtain

$$\|R_n(g)-R_n(g\circ T^j)\|_\infty \le 2j\|g\|_\infty/n.$$

Hence

$$|R_n(g-R_m(g))| \le \frac{\|g\|_\infty}{nm}\sum_{j=1}^m 2j = \frac{(m+1)\|g\|_\infty}{n}.$$

Thus

$$\begin{aligned}\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| &\leq \sup_{n\geq 1} \left|R_n(R_m(g) - \mathbb{E}^{\mathcal{I}} g)\right| \\ &\leq \left|R_n(R_m(g) - \mathbb{E}^{\mathcal{I}} g)\right|, \text{ a.s.}\end{aligned}$$

With Corollary 9.1.2 we derive

$$\mathbb{P}\left(\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| > c\right) \leq \frac{1}{c}\mathbb{E}\left|R_m(g) - \mathbb{E}^{\mathcal{I}} g\right| \to_{m\to\infty} 0.$$

So $\mathbb{P}\left(\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| = 0\right) = 1$.
For the general case ($g \in \mathbb{L}^1$) there exists a sequence of bounded functions g_m which satisfies $\|f - g_m\|_1 \to_{m\to\infty} 0$. Then

$$\left|R_n(f) - \mathbb{E}^{\mathcal{I}} f\right| \leq |R_n(f - g_m)| + \left|R_n(g_m) - \mathbb{E}^{\mathcal{I}} g_m\right| + \left|\mathbb{E}^{\mathcal{I}}(g_m - f)\right|.$$

Hence

$$\limsup_n \left|R_n(f) - \mathbb{E}^{\mathcal{I}} f\right| \leq \sup_{n\geq 1} |R_n(f - g_m)| + \left|\mathbb{E}^{\mathcal{I}}(g_m - f)\right| \quad a.s.$$

1) Markov inequality implies $\mathbb{E}^{\mathcal{I}}(g_m - f) \xrightarrow{\mathbb{P}}_{m\to\infty} 0$. Indeed

$$\mathbb{E}\left|\mathbb{E}^{\mathcal{I}}(g_m - f)\right| \leq \frac{1}{c}\|g_m - f\|_1$$

2) Let $A_m = \sup_{n\geq 1} |R_n(f - g_m)|$ then from Lemma 9.1.1:

$$\mathbb{P}(A_m > c) \leq \frac{1}{c}\|f - g_m\|_1.$$

The previous relations 1) and 2) imply

$$\mathbb{P}\left(\limsup_n \left|R_n(f) - \mathbb{E}^{\mathcal{I}} f\right| > c\right) = 0$$

This holds for each $c > 0$ which implies the result. ∎

In the case of stationary processes this theorem is reformulated with the shift operator T.

Corollary 9.1.3. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary process. If $f:\mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$ is measurable and $\mathbb{E}|f(X)|<\infty$ then*

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\to_{n\to\infty}\mathbb{E}^{\mathcal{J}}f(X)\quad a.s.\qquad \text{and in } \mathbb{L}^1.$$

If $\mathbb{E}f^2(X)<\infty$ the convergence also holds in $\mathbb{L}^2$.

Proof. The only point to notice is that $\mathbb{E}^{\mathcal{J}}f(X)=\mathbb{E}^{\mathcal{I}}_{\mathbb{P}_X}f$. ∎

Example 9.1.2. *Exercise 23 provides us with a non-ergodic sequence satisfying anyway a law of large numbers.*

Remark 9.1.2. *If the process X is ergodic*

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\longrightarrow\mathbb{E}f(X)\quad \text{if } \mathbb{E}|f(X)|<\infty.$$

Ergodicity may also be omitted if $\mathbb{E}f^2(X)<\infty$ and

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\to_{n\to\infty}\mathbb{E}f(X)\ a.s.\Leftrightarrow\frac{1}{n}\sum_{k=1}^{n}\ Cov(f(X),f\circ T^k(X))\to 0$$

Moreover, as a partial converse of Theorem 9.1.1, quote that if the above limit is constant everywhere for any integrable function f then the system is ergodic, see [Kallenberg, 1997].

After those remarks we derive examples of ergodic processes.

Example 9.1.3 (ergodic processes)**.**
The following models fit the ergodicity condition.

- *An independent identically distributed sequence is also a stationary and ergodic sequence. For this, use Kolmogorov $0-1$'s law.*

- *Hence Bernoulli schemes are also ergodic. Indeed if $X=(X_i)_{i\in\mathbb{Z}}$ is defined from an independent identically distributed sequence*

$\xi = (\xi_i)_{i\in\mathbb{Z}}$ and a function H through equation (7.11) then $f\circ T^i(X) = f\circ H\circ T^i(\xi)$; hence as soon as $\mathbb{E}|f(X)| < \infty$

$$\frac{1}{n}\sum_{i=1}^{n} f\circ T^i(X) \to \mathbb{E}(f(X)).$$

This is true for bounded measurable functions $\mathbb{R}^{\mathbb{Z}}$ in $\mathbb{R}$ which entails the ergodicity of X.

- *If relation $Cov(f(X_0), f(X_n)) \to 0$ as $n\to\infty$ for $f\in\mathcal{F}$ (this class of functions generates a dense linear vector subspace of $\mathbb{L}^1$). Indeed this relation implies with Cesaro lemma that*

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X) \to_{n\to\infty} \mathbb{E}f(X) \quad \text{in } \mathbb{L}^1.$$

The result still holds for each bounded function from a density argument. Now Corollary 9.1.3 entails $\mathbb{E}^{\mathcal{J}} f(X) = \mathbb{E}f(X)$ and ergodicity follows.
Forthcoming examples follow this scheme:

- *A Gaussian stationary sequence is ergodic if its covariance $r_n \to 0$. Quote that this condition seems necessary since eg. a constant sequence $X_n = \xi_0 \sim \mathcal{N}(0,1)$ is not ergodic.*
Assume $X_0 \sim \mathcal{N}(0,1)$. If f Hermite expansion writes

$$f = \sum_{k=0}^{\infty} c_k H_k$$

then

$$Cov(f(X_0), f(X_n)) = \sum_{k=1}^{\infty} \frac{c_k^2}{k!} r_n^k \ \big(= G(r_n)\big).$$

The function $G(r)$ defined this way is continuous on $[-1,1]$ if one sets $G(1) = \mathbb{E}f^2(X_0)$ and $G(0) = 0$. Ergodicity follows.

- *Strongly mixing sequences, and all the previous examples of weakly dependent sequences (see the definition (11.1)) are ergodic.*

- *A last example is that of stationary associated sequences such that $r_n \to 0$. For this use inequality (8.1).*

Remark 9.1.3. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary and ergodic centered sequence in $\mathbb{L}^2$.*
Then

$$\widehat{r}_{n,p} = \frac{1}{n-|p|}\sum_{k=|p|+1}^{n} X_k X_{k-|p|} \tag{9.2}$$

fits $r_p = \mathbb{E}X_0X_p$ without bias (i.e. $\mathbb{E}\widehat{r}_{n,p} = r_p$) and $\widehat{r}_{n,p} \to r_p$ a.s. and in $\mathbb{L}^1$ (it is consistent).
For this use the previous result with $f(\omega) = \omega_0\omega_p$.
Let $(\xi_n)_{n\in\mathbb{Z}}$ be stationary and ergodic with $\mathbb{E}\xi_0^2 < \infty$.
If $|a| < 1$ then $X_n = \sum_{k=0}^{\infty} a^k\xi_{n-k}$ is stationary and ergodic and $\mathbb{E}X_0^2 < \infty$.
Moreover

$$X_n = aX_{n-1} + \xi_n, \quad \forall n \in \mathbb{Z}.$$

The previous solution is the unique sequence such that this relation holds. It is the first order auto-regressive process.
Previous arguments imply

$$\widehat{a}_n = \frac{\sum_{k=2}^{n} X_nX_{n-1}}{\sum_{k=2}^{n} X_n^2} \longrightarrow_{n\to\infty} a, \qquad a.s.$$

if $\mathbb{E}\xi_0 = 0$ and $\lim_{p\to\infty}\mathbb{E}\xi_0\xi_p = 0$ for the ergodic sequence (ξ_t).

Example 9.1.4. *Chapter 7 includes a wide variety of applications for which Theorem 9.1.1 applies. Essentially the consistence properties of all empirical processes follows from this main result.*

9.2 Range

We aim at providing some ideas yielding definitions for the range of a process. Namely we advocate to define it according a possible limit theorem. As this is claimed at the beginning to the present Part III, a definition through a limit theorem in distribution allows to define an asymptotic confidence interval for testing a mean through the simplest frequentist empirical mean. After Theorem 9.1.1 this is

indeed known that such empirical estimates do converge under mild assumptions.
The classical definition of the long/short range dependence for second order stationary sequences is based on the convergence rates to zero covariances $r_k = \mathrm{Cov}\,(X_0, X_k)$, more precisely the convergence of the following series is of importance:

$$\sum_k r_k.$$

Definition 9.2.1 ($\mathbb{L}^2-$range)**.** *In case the series (r_k) is absolutely convergent the process is short range dependent (SRD) and if the series diverges the process is long range dependent (LRD).*

The proof of Proposition 4.3.2 provides an expression of the square of a convergence rate in $\mathbb{L}^2$ in the ergodic theorem under $\mathbb{L}^2-$stationarity:

$$\mathbb{E}(S_n - n\mathbb{E}X_0)^2 = \mathrm{Var}\left(\sum_{k=1}^n g(X_k)\right) = \sum_{|k|<n} (n-|k|)r_k.$$

Based on the previous definition the partial sums

$$S_n = \sum_{k=1}^n X_k$$

admit variances with order n or $\gg n$ according to either an SRD or an LRD behavior.
A phenomenon of very short range corresponds to $g_X(0) = 0$; in this case $\mathrm{Var}\, S_n \ll n$.
More generally consider L_1, L_2, L_3 are slowly varying functions (typically powers of logarithm) and constants $\alpha, \beta, \gamma > 0$ introduce the relations

$$\sum_{k=-n}^n r_k \quad \sim_{n\to\infty} \quad n^\alpha L_1(n) \tag{9.3}$$

$$r_n \quad \sim_{n\to\infty} \quad k^{-\beta} L_2(n) \tag{9.4}$$

$$g_X(\lambda) \quad \sim_{\lambda\to 0} \quad |\lambda|^{-\gamma} L_3\left(\frac{1}{|\lambda|}\right) \tag{9.5}$$

One may prove (Taqqu in [Doukhan et al., 2002b])

Theorem 9.2.1 (Tauber)**.** *If r_k is monotonous for $k \geq k_o$ then relations (9.3), (9.4) and (9.5) are equivalent with $\alpha = 1 - \beta$, $L_1 = \frac{2}{1-\beta} L_2$, $\gamma = 1 - \beta$ and $L_3 = \frac{\Gamma(\alpha+1)}{2\pi} \sin \frac{\pi(1-\alpha)}{2} L_1$.*

This yields a convergence rate in the precise law of large numbers (Ergodic Theorem 9.1.1), but in case one needs more accurate test for goodness-of-fit, then some more information is needed. This definition is quite unsatisfactory because a user is more involved in the asymptotic behavior of functionals of a process better that its only $\mathbb{L}^2-$behavior. Even for an orthogonal sequence $\mathrm{Var}\, S_n = n \mathrm{Var}\, X_0$ does not imply an asymptotically Gaussian behavior.

Example 9.2.1. *Let (ξ_n) be an independent identically distributed sequence with marginals $\mathcal{N}(0,1)$ and let η be a real valued random variable independent of this sequence then $X_n = \eta \xi_n$ is orthogonal stationary but it is not ergodic since $S_n/\sqrt{n}$ admits the same distribution as $\eta\xi_0$ usually not Gaussian.*

A more attractive definition is thus based on limit theorems relative to the partial sums:

$$S_n = X_1 + \cdots + X_n$$

Definition 9.2.2 (distributional range)**.** *Let $(X_n)_{n\in\mathbb{Z}}$ be a strictly stationary and centered sequence in $\mathbb{L}^2$:*

- *if $\frac{1}{\sqrt{n}} S_n$ is asymptotically Gaussian then we say it is short range dependent. Precisely we may suppose that $Var\, S_n \sim cn$ (as $n \to \infty$) for some constant $c > 0$.*
  Assume that the sequence of processes
  $$t \mapsto Z_n(t) = \frac{1}{\sqrt{Var\, S_n}} S_{[nt]}, \qquad for\ t \in [0,1]$$
  converges toward a Brownian motion in the Skorohod space $D[0,1]$ (see Definition A.3.2).
- *if the sequence of processes*
  $$t \mapsto Z_n(t) = \frac{1}{\sqrt{Var\, S_n}} S_{[nt]}, \qquad for\ t \in [0,1]$$
  does not converge toward a Brownian motion it would be long range dependence

An alternative definition omits the fact that $X_n \in \mathbb{L}^2$. It ask that the previous partial sums process admits a limit with either independent increments or not. This nice proposal is that of Herold Dehling and allows to aggregate cases of heavy tail processes and Lévy processes.

Chapter 10

Long range dependence

Long range pendant phenomena were first exhibited by Hurst for hydrology purposes. This phenomenon occurs from the superposition of independent sources, namely confluent rivers provide this behavior. Such aggregation procedures provide this new phenomenon.
In the present chapter we address the Gaussian and linear cases as well as the case of functions of such processes where such LRD phenomena occur. Due to the technical difficulties we restrict to the initial example of Rosenblatt for functions of Gaussian processes. Finally we also describe some few additional extensions.
The most elementary example is that of Gaussian processes. We follow the presentation in [Rosenblatt, 1985] who discovered long range dependent behaviors. He considered models of instantaneous functions of a Gaussian process.

10.1 Gaussian processes

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary centered Gaussian sequence with $r_0 = \mathbb{E}X_0^2 = 1$ and with covariance

$$r_k \sim ck^{-\beta} \qquad \text{as} \qquad k \to \infty,$$

for $c > 0$, $\beta > 0[$ (Theorem 4.2.1 proves that the sequence $r_k = (1 + k^2)^{-\beta/2}$ is indeed the sequence of covariances of a stationary

Gaussian process: hence there exist such sequences). Tauber theorem 9.2.1 implies $g(\lambda) \sim |\lambda|^{a-1}$. Also $S_n \sim \mathcal{N}(0, \operatorname{Var} S_n)$ with

$$\operatorname{Var} S_n = n \sum_{|k|<n} \left(1 - \frac{|k|}{n}\right) r_k,$$

and

$$Z_n(t) \sim \mathcal{N}\left(0, \frac{\operatorname{Var} S_{[nt]}}{\operatorname{Var} S_n}\right).$$

- Hence if $\beta > 1$, $\operatorname{Var} S_n \sim n\sigma^2$ the sequence is SRD and

$$\frac{1}{n}\operatorname{Var} S_{[nt]} \to t\sigma^2.$$

Now Z_n converges to a Brownian motion with variance

$$\sigma^2 = \sum_{k=-\infty}^{\infty} r_k.$$

First check that

$$\mathbb{E}Z_n(t)Z_n(s) \to (s \wedge t)\sigma^2.$$

Tightness is consequence of

$$\mathbb{E}(Z_n(t) - Z_n(s))^2 \le C|t-s|$$

for $C = \sum_k |r_k|$ and from Chentsov Lemma 10.1.1 .
Indeed for Gaussian processes

$$\mathbb{E}|Z_n(t) - Z_n(s)|^p = \mathbb{E}|N|^p \, [\mathbb{E}(Z_n(t) - Z_n(s))^2]^p$$

for each $p > 2$ if $N \sim \mathcal{N}(0,1)$.

- If now $\beta < 1$ the series of covariances diverges

$$\operatorname{Var} S_n \sim n^{2-\beta} \qquad \text{if} \qquad r_k \sim ck^{-\beta}.$$

Hence

$$Z_n(t) \to \mathcal{N}\left(0, ct^{2-\beta}\right),$$

does not converge to the Brownian motion; indeed contrary to the Brownian motion the previous variance does not increase linearly with t.

Remark 10.1.1 (Chentsov Lemma). *This standard lemma (see e.g. [Billingsley, 1999] and [van der Vaart and Wellner, 1998] for a more complete overview) asserts tightness of the sequence of processes Z_n in $C[0,1]$ if*

$$\mathbb{E}|Z_n(t) - Z_n(s)|^p \leq C|t-s|^a$$

for some $a > 1$.
For SRD sequences this needs $p > 2$ because $a = p/2$. The above mentioned relations imply this holds for a Gaussian process if $p = 2$ and $a > 0$.
The long range dependent case is much nicer since for $p = 2$ one derives $a = 2 - \alpha > 1$.

10.2 Gaussian polynomials

Generally if the process (X_n) is stationary and standard Gaussian (i.e. $\mathbb{E}X_0 = 0$, $\operatorname{Var} X_0 = 1$) with $r_k \sim ck^{-\beta}$ and the function g is such that $\mathbb{E}|g(X_0)|^2 < \infty$ then

- $$\operatorname{Var}\left(\sum_{k=1}^{n} g(X_k)\right) = \mathcal{O}\left(n\right),$$

  if $\beta \cdot m(g) > 1$ and $m(g)$ denote the Hermite rank of g.
  In this first case the diagram formula (§ 5.2.3) allows to prove the convergence in distribution, [Breuer and Major, 1983]:

  $$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} g(X_k) \to_{n\to\infty} \sigma W_t, \qquad \text{in} \quad D[0,1].$$

  The result is also proved in a shorter way in [Nourdin et al., 2011]

- Else say if $\beta \cdot m(g) < 1$, then

  $$\operatorname{Var}\left(\sum_{k=1}^{n} g(X_k)\right) = \mathcal{O}\left(n^{2-m(g)\beta}\right).$$

Here $1-\frac{m(g)\beta}{2}>\frac{1}{2}$ and the convergence still holds

$$\frac{1}{n^{1-\frac{m(g)\beta}{2}}}\sum_{k=1}^{n}g(X_k)\to_{n\to\infty}Z_r,$$

to some non Gaussian distribution in case the rank is >1, see [Dobrushin and Major, 1979]).
The technique is complicated due to the fact that for $k>2$ the Laplace transform for the law of X_0^k is not analytic around 0.
The case $k=1$ is considered in the previous section and the case $k=2$ is the aim of the next one.

10.3 Rosenblatt process

The previous non Gaussian asymptotic may be proved elementary "à la main" in the case enlighted in [Rosenblatt, 1961], see also the nice monograph [Rosenblatt, 1985]. Set $Y_n=X_n^2-1$ then Mehler formula implies that the covariance $\mathrm{Cov}\,(Y_0,Y_k)$ equals $2r_k^2\sim 2c^2k^{-2\beta}$. The series of those covariances is divergent in case $\beta<\frac{1}{2}$.
In this case we aim at proving that

$$U_n=n^{\beta-1}\sum_{k=1}^{n}Y_k$$

converges toward a non-Gaussian limit. More explicitly the normalization should be written $\sqrt{n^{2\beta}}/n$.
Set R_n for the covariance matrix of the vector $(X_1,\ldots,X_n)$, then for t small enough:

$$\begin{aligned}
\mathbb{E}e^{tU_n} &= \mathbb{E}e^{tn^{\beta-1}\sum_{k=1}^{n}(X_k^2-1)}\\
&= e^{-tn^{\beta}}\int_{\mathbb{R}^n}e^{-x^t(R_n^{-1}-2tn^{\beta-1}I_n)x/2}\frac{dx}{(2\pi)^{n/2}\sqrt{\det R_n}}\\
&= e^{-tn^{\beta}}\int_{\mathbb{R}^n}e^{-y^t(I_n-2tn^{\beta-1}R_n)y/2}\frac{dy}{(2\pi)^{n/2}}\\
&= e^{-tn^{\beta}}\det{}^{-\frac{1}{2}}\left(I_n-2tn^{\beta-1}R_n\right)
\end{aligned}$$

Indeed through a linear change in variable for each symmetric definite positive matrix A with order n:

$$\int_{\mathbb{R}^n} e^{-y^t Ay/2} \frac{dy}{(2\pi)^{n/2}} = \frac{1}{\sqrt{\det(A)}}$$

Now denote $(\lambda_{i,n})_{1\le i\le n}$ the eigenvalues (≥ 0) of the symmetric and non-negative matrix R_n (diagonalizable) then

$$\begin{aligned}
\frac{1}{\sqrt{\det\left(I_n - 2tn^{\beta-1}R_n\right)}} &= \prod_{i=1}^{n}(1-2tn^{\beta-1}\lambda_{i,n})^{-1/2} \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\log\left(1-2tn^{\beta-1}\lambda_{i,n}\right)\right)
\end{aligned}$$

Use the following analytic expansion (valid for $|z|<1$)

$$\log(1-z)+z = -\sum_{k=2}^{\infty}\frac{z^k}{k}$$

The simple observation that $\operatorname{trace}(R_n) = n$ follows from the fact that R_n's diagonal elements equal 1; we thus deduce that

$$e^{-tn^\beta} = \exp\left(-2tn^{\beta-1} \text{ trace } R_n\right) = \exp\left(-\sum_{i=1}^{n}(2tn^{\beta-1})\lambda_{i,n}\right)$$

Thus:

$$\begin{aligned}
\mathbb{E}e^{tU_n} &= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left\{\log\left(1-2tn^{\beta-1}\lambda_{i,n}\right)+2tn^{\beta-1}\right\}\right) \\
&= \exp\left(\frac{1}{2}\sum_{k=2}^{\infty}\frac{1}{k}(2tn^{\beta-1})^k \text{ trace } R_n^k\right)
\end{aligned}$$

Quote now that:

$$\begin{aligned}
\frac{n^{k\beta}}{n^k}\text{trace } R_n^k &= n^{k(a-1)}\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n} r_{i_1-i_2}r_{i_2-i_3}\cdots r_{i_{k-1}-i_k}r_{i_k-i_1} \\
&\sim \frac{c^k}{n^k}\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{1}{\left|\frac{i_1}{n}-\frac{i_2}{n}\right|^\beta}\frac{1}{\left|\frac{i_2}{n}-\frac{i_3}{n}\right|^\beta}\cdots\frac{1}{\left|\frac{i_k}{n}-\frac{i_1}{n}\right|^\beta}
\end{aligned}$$

Hence through a discretization of a multiple integral by Riemann sums we derive :

$$\frac{n^{k\beta}}{n^k}\text{trace } R_n^k \to_{n\to\infty}, c_k > 0$$

with

$$c_k = c^k \int_0^1 \cdots \int_0^1 \frac{1}{|x_1 - x_2|^\beta} \cdots \frac{1}{|x_{k-1} - x_k|^\beta} \frac{1}{|x_k - x_1|^\beta} dx_1 \cdots dx_k.$$

For this, simple upper and lower bounds for integrals over cubes with volume n^{-k} allow to derive this convergence; indeed the function to be integrated is locally monotonic with respect to each coordinate. More generally [Polya and Szegö, 1970] prove the validity of such approximations for generalized integrals, in case of functions monotonic around their singularity as this is the case here. Thus for $k = 2$, on obtains that for each $1 \le i \le n$,

$$0 \le \lambda_{i,n}^2 \le \sum_{j=2}^n \lambda_{j,n}^2 \equiv \text{trace } R_n^2 = \mathcal{O}(n^{2(1-\beta)})$$

thus $2|t|\, n^{\beta-1}\lambda_{i,n} < 1$ if $|t| < c$ for some constant $c > 0$.
Thus **all the necessary convergences hold** in order to justify the above mentioned calculation if we assume $|t| \le c/2$.
Hence if t is small enough:

$$\mathbb{E}e^{tU_n} \to_{n\to\infty} \exp\left(\frac{1}{2}\sum_{k=2}^\infty (2t)^k \frac{c_k}{k}\right)$$

Hence this sequence converges in distribution to a non-Gaussian law (this distributeion is therefore named Rosenblatt distribution).
Indeed the logarithm of its Laplace transform is not a polynomial of order 2. ∎

Remark 10.3.1. *This technique does not extend to polynomials with degree* > 2 *since its Laplace transform is not analytic (this is easy to prove that if* $N \sim \mathcal{N}(0,1)$ *then*

$$\mathbb{E}\exp(t|N|^3) = 2\int_0^\infty \exp\left(tx^3 - \frac{1}{2}x^2\right)\frac{dx}{\sqrt{2\pi}} = \infty, \qquad \textit{if } t > 0,$$

and thus the method of moments does not apply to prove convergence in law (see Theorem 12.1.1).
[Dobrushin and Major, 1979] introduced weaker convergences for sequences of multiple Ito integrals in order to derive "non Central Limit Theorems", .

10.4 Linear processes

Other models admit analogue behaviors. Linear processes

$$X_n = \sum_{k=0}^{\infty} c_k \xi_{n-k}$$

for which $c_k \sim_{k\to\infty} ck^{-\beta}$ with $\frac{1}{2} < \beta < 1$ satisfy

$$r_k = \sum_l c_l c_{l+k} \sim_{k\to\infty} ck^{1-2\beta} \int_0^\infty \frac{ds}{(s(s+1))^\beta},$$

hence $\mathrm{Var}\,(X_1 + \cdots + X_n) \sim c' n^{2-2\beta}$ and it is possible to prove

$$n^{\beta-1} \sum_{k=1}^{[nt]} X_k \to_{n\to\infty} B_H(t),$$

with convergence in law in the Skorohod space $D([0,1])$ of right-continuous functions with limit on the left (called càdlàg functions, see Definition A.3.2). It follows from the following simple result in [Davydov, 1970]:

Theorem 10.4.1 (Davydov, 1970)**.** *Let (X_n) be a linear process. Set*

$$S_n = X_1 + \cdots + X_n.$$

If

$$Var\, S_n = n^{2H} L(n) \qquad (n \to \infty)$$

for a slowly varying function L and $0 < H < 1$ then

$$\frac{1}{n^H L(n)} \sum_{k=1}^{[nt]} X_k \xrightarrow{\text{in law}}_{n\to\infty} B_H(t)$$

Hint. This result also relies on the Lindeberg Theorem 2.1.1. ∎

10.5 Functions of linear processes

A martingale based technique was introduced in [Ho and Hsing, 1996] for the extension of such behaviors as previously considered for the Gaussian case. The idea of this section is to give a flavor of results and underlying techniques but the rigorous proofs should be found in the corresponding literature. Using the weak uniform reduction principle, [Giraitis and Surgailis, 1999] establish the same result for a causal linear process. Let

$$X_t = \sum_{s=0}^{\infty} b_s \xi_{t-s},$$

where ξ is independent identically distributed and $b_s = L(s)s^{-(\alpha+1)/2}$.

Theorem 10.5.1 (Causal linear process)**.** *Let $f(x)$ be the density of X_0 and $B_{1-\alpha/2}$ the fractional Brownian motion. If there exists constants $\delta, C > 0$ such that*

$$\left|\mathbb{E}\left(e^{iu\xi_0}\right)\right| \le C(1+|u|)^{-\delta},$$

and if $\mathbb{E}|\xi_0|^9 < \infty$ then there exists c_α an explicit constant with

$$n^{\alpha/2-1}F_n(x,t) \longrightarrow c_\alpha f(x) B_{1-\alpha/2}(t)$$

in $D[-\infty, +\infty] \times D[0,1]$.

A main tool is uniform control of the approximation of the empirical process by the partial sums process:

Proposition 10.5.1 (Uniform reduction principle)**.** *There exist $C, \gamma > 0$ such that for $0 < \varepsilon < 1$:*

$$\mathbb{P}\Big(\frac{n^{\frac{\alpha}{2}-1}}{L(n)} \sup_{\substack{n \,\le\, n \\ x \,\in\, \mathbb{R}}} \left|\sum_{t=1}^{n} \big(\mathbb{I}_{\{X_t \le x\}} - F(x) + f(x)X_t\big)\right| \ge \varepsilon\Big) \le \frac{C}{n^\gamma \varepsilon^3}.$$

Proof. Set

$$S_n(x) = \frac{\sqrt{n^\alpha}}{nL(n)} \sum_{t=1}^{n} \big(\mathbb{I}_{\{X_t \le x\}} - F(x) + f(x)X_t\big).$$

Then

$$\mathrm{Var}\,(S_n(y) - S_n(x)) \le \frac{n}{n^{1+\gamma}}\mu([x,y]).$$

where μ is a finite measure on $\mathbb{R}$. Then a chaining argument is used.

Remark 10.5.1. *[Ho and Hsing, 1996] extend the expansion of the reduction principle:*

$$S_{n,p}(x) = \frac{n^{p\alpha/2}}{nL(n)}\left(\sum_{t=1}^{n} 1\!\!1_{\{X_t \le x\}} - F(x) - \sum_{r=0}^{p}(-1)^r F^{(r)}(x) Y_{n,r}\right),$$

$$Y_{n,r} = \sum_{t=1}^{n}\sum_{1\le j_1<\ldots<j_r}\prod_{s=1}^{r} b_{j_s}\xi_{t-j_s}.$$

Proposition 10.5.2 (Uniform reduction principle)**.** *If the density of ξ_0 is $(p+3)$-times differentiable and if $\mathbb{E}|\xi_0|^4<\infty$, there exist $C,\gamma>0$ such that for $0<\varepsilon<1$:*

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}|S_{n,p}(x)|\ge\varepsilon\right)\le Cn^{-(\alpha\wedge(1-p\alpha))+\gamma}\varepsilon^{-2-\gamma}.$$

Proof methods. Computation of the variance of $S_{n,p}(x)$: for

$$f_t(x) = 1\!\!1_{\{X_t\le x\}} - F(x) - \sum_{r=0}^{p}(-1)^r F^{(r)}(x)Y_{n,r}$$

write the orthogonal decomposition:

$$f_t(x) - \mathbb{E}f_t(x) = \sum_{s=1}^{\infty}\mathbb{E}(f_t(x)|\mathcal{F}_{t-s}) - \mathbb{E}(f_t(x)|\mathcal{F}_{t-s-1}),$$

where $\mathcal{F}_t$ is the σ-field generated by the $\{\xi_s,\ s\le t\}$. Compute the variance of each term using a Taylor expansion. Note that the $\mathcal{F}_t$ are increasing so that many covariances between terms are zero.
[Doukhan et al., 2005] generalize the preceding method to the case of random fields. (X_t) is a *linear* random field:

$$X_t = \sum_{u\in\mathbb{Z}^d} b_u\zeta_{t+u}, \quad t\in\mathbb{Z}^d,$$

where $(\zeta_u)_{u\in\mathbb{Z}^d}$ is an i.i.d. random field with zero mean and variance 1, and $b_u = B_0(u/|u|)|u|^{-(d+\alpha)/2}$, for $u \in \mathbb{Z}^d$, $0 < \alpha < d$ and B_0 is a continuous function on the sphere.
Let $A_n = [1,n]^d \cap \mathbb{Z}^d$ and $F_n(x) = \dfrac{1}{n^d}\sum_{t\in A_n} 1\!\!1_{\{X_t\le x\}}$:

Theorem 10.5.2. *If there exists $\delta, C > 0$ such that*

$$\left|\mathbb{E}\left(e^{iu\xi_0}\right)\right| \le C(1+|u|)^{-\delta},$$

and if $\mathbb{E}|\xi_0|^{2+\delta} < \infty$ then

$$n^{\alpha/2}(F_n(x) - F(x)) \longrightarrow c_\alpha f(x)Z,$$

in $D[-\infty, +\infty]$, where Z is a Gaussian random variable .

Remark 10.5.2. *This is remarkable that the limit distribution is extremely simple in this case, indeed Z does not depend on x. Recall that the weak dependent case yields much more complicated limit behaviors, typically the Brownian bridge which Hölder regularity exponent satisfies $\alpha < \frac{1}{2}$.*

10.6 More non-linear models

The section aims at providing some directions for the extension of LRD. to nonlinear models. Below one addresses more bibliographical comments that rigorous statements.

10.6.1 LARCH–type models

As quoted in § 7.2.2 models solution of the recursion:

$$X_n = \Big(b_0 + \sum_{j=1}^{\infty} b_j X_{n-j}\Big)\xi_n,$$

also admit LRD behaviors if the iid sequence (ξ_t) is centered and

$$\mathbb{E}\xi_0^2 \sum_{j=1}^{\infty} b_j^2 < 1,$$

but

$$\sum_{j=1}^{\infty} |b_j| = \infty.$$

More general volatility models

$$X_t = \sigma_t \xi_t, \quad \sigma_t^2 = G\Big(\sum_{j=1}^{\infty} b_j X_{t-j}\Big),$$

extend on ARCH(∞)–models ($G(x) = b_0 + x^2$), and asymmetric ARCH(∞)–models ($G(x) = b_0 + (c+x)^2$). Again one requires

$$\sum_{j=1}^{\infty} b_j^2 < \infty, \qquad \sum_{j=1}^{\infty} |b_j| = \infty.$$

10.6.2 Randomly fractional differences

[Philippe et al., 2008] introduced time-varying fractional filters $A(\mathbf{d}), B(\mathbf{d})$ defined by

$$A(\mathbf{d})x_t = \sum_{j=0}^{\infty} a_j(t) x_{t-j}, \qquad B(\mathbf{d})x_t = \sum_{j=0}^{\infty} b_j(t) x_{t-j}, \tag{10.1}$$

where $\mathbf{d} = (d_t, t \in \mathbb{Z})$ is a given function of $t \in \mathbb{Z}$, and where we set $a_0(t) = b_0(t) = 1$, and if $j \geq 1$:

$$\begin{aligned} a_j(t) &= \Big(\frac{d_{t-1}}{1}\Big)\Big(\frac{d_{t-2}+1}{2}\Big)\Big(\frac{d_{t-3}+2}{3}\Big)\cdots\Big(\frac{d_{t-j}+j-1}{j}\Big), \\ b_j(t) &= \Big(\frac{d_{t-1}}{1}\Big)\Big(\frac{d_{t-j}+1}{2}\Big)\Big(\frac{d_{t-j+1}+2}{3}\Big)\cdots\Big(\frac{d_{t-2}+j-1}{j}\Big). \end{aligned}$$

If $d_t \equiv d$ is a constant, then $A(\mathbf{d}) = B(\mathbf{d}) = (I-L)^{-d}$ is the usual fractional integration operator ($Lx_t = x_{t-1}$ is the backward shift). [Doukhan et al., 2007] consider for centered independent identically distributed inputs ϵ_t,

$$X_t^A = \sum_{j=0}^{\infty} a_j(t) \epsilon_{t-j}, \qquad X_t^B = \sum_{j=0}^{\infty} b_j(t) \epsilon_{t-j}$$

If d_t is independent identically distributed and $\mathbb{E}d_t = \bar{d} \in (0, \frac{1}{2})$ then the asymptotic behavior of partial sums of this process is the same as for $FARIMA(0,0,\bar{d},0)$ which corresponds to the case of a constant sequence d_t. If ϵ_t is standard normal, then

- X_t is Gaussian with a variance $A(t) = \mathbb{E}(X_t^2|\mathcal{D})$, conditionally wrt $\mathcal{D}$, the past sigma-algebra of d_t,
- $g_k(A) = A^{-2k}\mathbb{E}[h(X)H_k(X;A)]$, where $H_k(x;A) = A^k H_k(x/A)$ (Hermite polynomials with variance A).

Then the Gaussian limit theory extends with Hermite coefficients replaced by $\beta_k = \mathbb{E}[g_k(A(0))Q^k]$ for a random variable Q related with the random coefficients d_t, $\mathbb{E}d_t = \bar{d}$ and d_t admits a finite range, and $\mathbb{E}|h(B\epsilon_t)|^a < \infty$, for some $a > 2$.

10.6.3 Perturbed linear models

[Doukhan et al., 2002a] study the empirical process of perturbed linear models:

$$X_t = Y_t + V_t, \quad t \in \mathbb{Z},$$

where (Y_t) is a long range dependent causal linear process and $V_t = H(\zeta_t, \zeta_{t-1}, \ldots)$ is a short range dependent perturbation. Then the perturbation does not modify the behavior of the empirical process which behaves as for linear LRD processes..

10.6.4 Non-linear Bernoulli shift models

Doukhan, Lang, Surgailis (manuscript) study the partial sums process of

$$X_t = H(Y_t; \zeta_t, \zeta_{t-1}, \ldots), \quad Y_t = \sum_{j=0}^{\infty} b_j \zeta_{t-j},$$

where $b_j \sim c_0 j^{d-1}$, with $d \in (0, 1/2)$, and ζ_t independent identically distributed innovations, and H is a function of infinitely many variables. A main goal of the results was to prove that:

There exists a non Gaussian process X whose partial sums converge to a second order Rosenblatt process while the partial sums of X^2 converge to the fractional Brownian motion.

The technique extends on [Ho and Hsing, 1996]: it is based on a martingale decomposition of the partial sums process.

$$S_n(t) = \sum_{s=1}^{[nt]} (X_s - \mathbb{E}X_s), \quad t \in [0,1].$$

This is possible to give conditions ensuring that, in law:

$$\begin{aligned} S_n(t) &\sim h'_\infty(0) \sum_{s=1}^{[nt]} Y_s, \\ h_\infty(y) &= \mathbb{E}H(y + Y_t, \zeta_t, \zeta_{t-1}, \dots). \end{aligned}$$

An analogue result holds with a second order U-statistic of ζ which asymptotic is related to the Rosenblatt process. There exists a constant $c_d \in \mathbb{R}$ such that if $d \in]\frac{1}{2}, 1]$ and if $h'_\infty(0) \neq 0$ then:

$$n^{-d-1/2} S_n(t) \quad \overset{D[0,1]}{\longrightarrow}_{N\to\infty} \quad c_d h'_\infty(0) B_d(t),$$

if now $d \in]\frac{1}{4}, 1]$ and $h'_\infty(0) = 0$ and $h''_\infty(0) \neq 0$ then:

$$n^{-2d} S_n(t) \quad \overset{D[0,1]}{\longrightarrow}_{N\to\infty} \quad c_d h''_\infty(0) Z_d^{(2)}(t).$$

Chapter 11

Short range dependence

The aim of the Chapter is only to fix some simple ideas. We investigate here Conditions on time series such that the standard limit theorems obtained for independent identically distributed sequences still hold.
After a general introduction to weak dependence conditions an example quotes the fact that the most classical *strong-mixing condition* from [Rosenblatt, 1956] may fail to work, see [Andrews, 1984].

Now when dealing with any weak dependence condition (including strong mixing), additional *decay rates* and *moment conditions* are necessary to ensure CLTs. Thus decay rates need to be known. Coupling arguments as proposed in § 7.4.2 are widely used for this.
Finally to make more clear the need of decay rates, we explain how CLTs may be proved with such assumptions.
The monograph [Dedecker et al., 2007] is used as the reference for weak dependence and formal results should be found there.

11.1 Weak dependence

Looking for asymptotic independence it seems natural to consider conditions such as this in [Doukhan and Louhichi, 1999]:

Definition 11.1.1. *Assume that there exist classes of functions*

$$\mathcal{F}, \mathcal{G} : \bigcup_{u \geq 1} \mathbb{R}^u \to \mathbb{R},$$

one function $\psi : \mathcal{F} \times \mathcal{G} \to \mathbb{R}$ (which depends on f, g and on the number of their arguments u, v) and
a sequence $\epsilon_r \downarrow 0$ as $r \uparrow \infty$.
A random process $(X_t)_{t \in \mathbb{Z}}$ is said to be $(\mathcal{F}, \mathcal{G}, \psi, \epsilon)-$weakly dependent in case

$$|Cov\left(f(X_{i_1}, \ldots, X_{i_u}), g(X_{j_1}, \ldots, X_{j_u})\right)| \leq \epsilon_r \psi(u, v, f, g) \qquad (11.1)$$

for functions f, g belonging respectively to classes $\mathcal{F}, \mathcal{G}$, and

$$i_1 \leq \cdots \leq i_u \leq j_1 - r \leq j_1 \leq \cdots \leq j_v.$$

The scheme of epochs is that reported in 11.1. The forthcoming

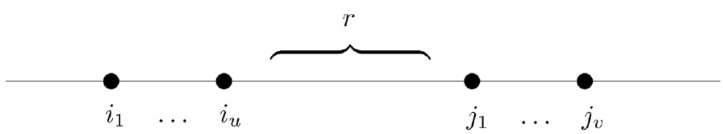


Figure 11.1: Asymptotic independence.

sections are devoted to examples of those generic notions.
First we explicitly consider strong mixing as well as a simple counter-example, and then we develop model based bootstrap as an example of application for which wider weak dependence notions in definition 11.4.1 are a reasonable option.

11.2 Strong mixing

A special case of the previous weak dependence situation is strong mixing for which

$$\mathcal{F} = \mathcal{G} = \mathbb{L}^\infty, \quad \text{and} \quad \psi(u, v, f, g) = 4\|f\|_\infty \|g\|_\infty, \quad \epsilon_r = \alpha_r.$$

Examples of strongly mixing are given in [Doukhan, 1994]. The sup bound of such ϵ_r satisfying this inequality is denoted α_r and also writes t

$$\alpha_r = \sup_{\substack{A \in \sigma(X_i,\, i \le 0) \\ B \in \sigma(X_j,\, j \ge r)}} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right|.$$

Indeed the previous inequality extends to (11.1) for non-negative linear combinations of indicator functions. A density argument make possible to consider arbitrary non-negative functions. A factor 4 appears when one allows functions with values in $[-1, 1]$.

Anyway this condition does not hold for some models. Eg.

$$X_n = \frac{1}{2}(X_{n-1} + \xi_n) \tag{11.2}$$

where the independent identically distributed inputs (ξ_n) admit a Bernoulli distribution with parameter $\frac{1}{2}$.
Quote on the simulation (11.2) that this model admits quite chaotic samples while its covariances decay quite fast ($\mathrm{Cov}\,(X_0, X_t) = 2^{-t}$).

Proposition 11.2.1. *The stationary solution of the equation (11.2) exists and is uniform on the unit interval moreover it is not strong mixing, more precisely* $\alpha_r \ge \frac{1}{4}$.

Note. In this case of equation (11.2) the process is however weakly dependent under **alternative dependence conditions**, see Example 11.4.1. More precisely $\epsilon_r (= \theta_r) \le 2^{1-r}$ for $r \in \mathbb{N}$, holds under a dependence assumption for which the considered classes of functions are Lipschitz, see below in § 11.4 some more precise statements.

Proof. The function $x \mapsto \frac{1}{2}(x + u)$ maps $[0, 1]$ in a subset of $[0, 1]$ which implies that applying recessively the equation (11.2) yields

$$X_n = \sum_{k=0}^{p} 2^{-1-k} \xi_{n-k} + 2^{-1-p} X_{n-p}.$$

Hence if we assume that initial values of the model are in the unit interval, the remainder term is $\le 2^{-1-p} \to_{p \to \infty}$.

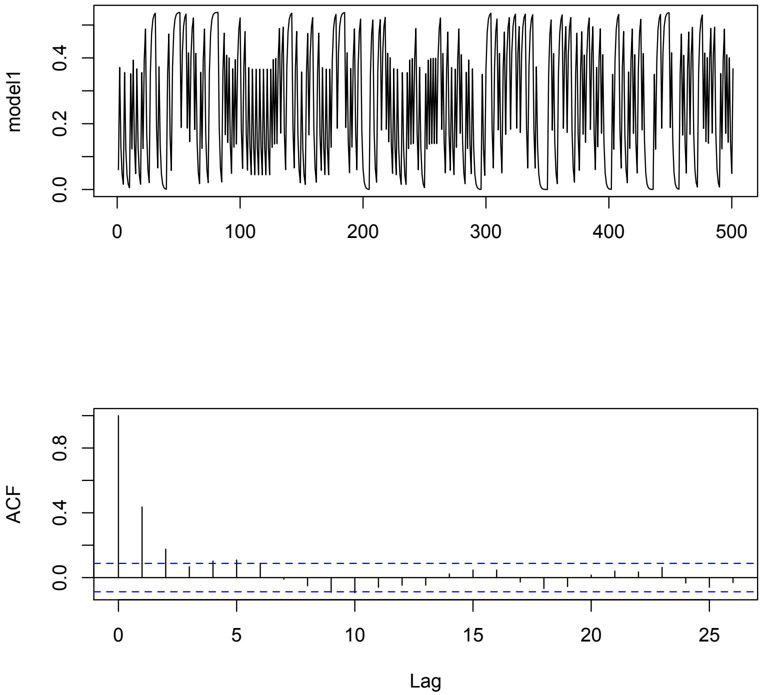


Figure 11.2: A non-mixing AR(1)–process.

The stationary solution of the previous equation writes

$$X_n = \sum_{k=0}^{\infty} 2^{-1-k} \xi_{n-k} = 0, \xi_n \xi_{n-1} \cdots \text{ in the numeration basis 2.}$$

The expansion of a real number in $x = 0.x_1 x_2 x_3 \ldots \in [0, 1)$ in the base 2 is in fact unique if one adopts the convention that there does not exist an integer p with $x_k = 1$ for each $k \geq p$.

This does not matter much since such event admits zero probability. The marginals of this process are easily proved to be uniformly distributed on $[0,1]$ :: for example choose an interval with dyadic extremities makes it evident and such intervals generate the Borel sigma field of $[0,1]$.
Now the previous condition writes in terms of sigma-algebra generated by the processes and X_{t-1} is the fractional part of $2X_t$ which implies the inclusion of sigma algebras generated by marginals of such processes.
More precisely

$$X_0 = 0, \xi_0\xi_{-1}\xi_{-2}\dots,$$

and

$$X_r = 0, \xi_r\xi_{r-1}\xi_{r-2}\cdots\xi_0\xi_{-1}\xi_{-2}\dots$$

thus the event $A = (X_0 \le \frac{1}{2})$ writes as $\xi_0 = 0$ and thus $\mathbb{P}(A) = \frac{1}{2}$; now there exists a measurable function (namely the $r-$th iterate of $x \mapsto$frac$(2x)$) such that $X_0 = f_r(X_r)$ and thus $A = f_r^{-1}([0, \frac{1}{2}]) \in \sigma(X_r)$.
If $r = 1$ then

$$A = \left(X_1 \in \left[0, \frac{1}{4}\right] \cup \left[\frac{1}{2}, \frac{3}{4}\right]\right)$$

and if $r = 2$ we also easily check that

$$A = \left(X_2 \in \left[0, \frac{1}{8}\right] \cup \left[\frac{1}{4}, \frac{3}{8}\right] \cup \left[\frac{1}{2}, \frac{5}{8}\right] \cup \left[\frac{3}{4}, \frac{7}{8}\right]\right).$$

More generally $A = B$ with $A_r = (X_r \in I_r)$ where I_r is the union of 2^r intervals with dyadic extremities and with the same length 2^{-r-1}. Thus

$$\alpha_r \ge \sup\{A \in \sigma(X_0), B \in \sigma(X_r)\} \ge \mathbb{P}(A \cap B_r) - \mathbb{P}(A)\mathbb{P}(B_r) = \frac{1}{4}. \qquad \blacksquare$$

Such examples prove that strong mixing type notions are not enough to consider a wide class of statistical models.

11.3 Bootstraping autoregressive models

A main problem for time series is that the exact distribution of many useful functionals are unknown. Such functionals are important since they usually appear as limits (in distribution) of some convergent sequences of functionals

$$G_k = g_k(X_1, \dots, X_k) \to_{k\to\infty} \Gamma.$$

We already considered subsampling in § 4.6 as an easy process. A common way to estimate the quantiles of Γ is due to Efron (1982) and it is known as Bootstrap.
From the previous convergence in distribution this will be used to determine distribution of simply high order quantiles useful to determine the property of tests of goodness of fit.
To this aim this is important to be able to simulate many samples

$$\{X_1(\omega_i), \dots, X_k(\omega_i)\}, \quad \text{for } 1 \le i \le I.$$

Indeed the simple law of large numbers (for independent identically distributed samples) allows to prove the consistency of the empirical quartiles derived from such samples.
We do not intend to provide an abstract theory for the bootstrap but rather to explain how to implement it over a very simple example.
Consider first the estimation the model

$$X_n = aX_{n-1} + \xi_n, \tag{11.3}$$

for an independent identically distributed and centered sequence with first order finite moment.
Then analogously for $|a| < 1$ this equation admits a solution

$$X_n = \sum_{k=0}^{\infty} a^k \xi_{n-k}.$$

In order to bootstrap we proceed to the following steps.

- The estimate $\widehat{a}_n$ in Remark 9.1.3 is proved to be *a.s.* convergent by a simple use of the ergodic theorem. Assume that one observes a sample $\{X_1(\omega), \dots, X_N(\omega)\}$ (which means that $\omega \in \Omega$ is fixed outside of some negligible event) of the stationary solution of the AR(1) process (11.3).

- Then let us use the first n data $\{X_1(\omega), \dots, X_n(\omega)\}$ to estimate $\widehat{a}_n$ and thus from *a.s.* convergence we may suppose that N is

large enough in order that for $n \geq N/3$, $|\widehat{a}_n| < 1$. We thus only used the first third of the sample to estimate $\widehat{a}_n$ and this allows to estimate residuals $\widehat{\xi}_j = X_j - \widehat{a}_{N/3}X_{j-1}$ for $j > 2N/3$.
We just forget one third of the data to assume that the random variables $\{\widehat{\xi}_j;\, j > 2N/3\}$ are almost independent of $\widehat{a}_{N/3}$ (in some sense to be precised).

- Now assume that $N = 3n$ we even may consider conditionally centered residuals by setting

$$\tilde{\xi}_j = \widehat{\xi}_j - \frac{1}{n}\sum_{k=2n+1}^{N} \widehat{\xi}_k, \quad 2n < j \leq N.$$

- To the end of resampling statistics we simulate independent identically distributed sequences $(\xi^*_{i,j})_{(i,j)\in\mathbb{N}\times\mathbb{Z})}$ with uniform distribution on the set $\{\tilde{\xi}_j; 2n+1 \leq j \leq 3n\}$.

- This means that we may simply simulate trajectories of the stationary solution of eqn. (11.3):

$$X^*_{i,n} = \widehat{a}_n X^*_{i,n-1} + \xi^*_{i,n}, \qquad i \geq 0,\ n \in \mathbb{Z},$$

  which exists since we have chosen n large enough so that $|\widehat{a}_n| < 1$.

As a final remark quote that the previous stationary solutions of eqn. (11.3) are shown to be strongly mixing only in the case when ξ_0's distribution admits an absolutely continuous part, see [Doukhan, 1994]. This is not the case for the resampled process which led authors like Jens Peter Kreiss and Michael Neumann to simply smooth the discrete distribution ν_* of ξ^*_0 in order to be able to use the necessary asymptotic properties shown under strong mixing in order to prove the consistency of those techniques. They simply convolute ν_* with any probability density to get an absolutely continuous distribution. They thus just replace ξ^*_0 distribution by $\xi^*_0 + \zeta_0$ for a small random variable ζ_0 independent of ξ^*_0 admitting a density wrt Lebesgue measure (think of $\zeta_0 \sim \mathcal{N}(0, \epsilon^2)$), then the Markov chain obtained is ergodic and strong mixing applies (see [Doukhan, 1994] for details) but it is not clear whether this distribution admits a real sense wrt bootstrap.

11.4 Weak dependence conditions

This sections aims at proving on the simple example of linear processes that an alternative to mixing defined in Definition 11.1.1 is indeed more adapted to such resampling questions.
The main point on the understanding of the concept of weak dependence is the following: for some processes we are able to get fast speed of decay just for some classes of functions. In this way it is natural to restrict the class of observables that we want to analyze to functions f and g which are on some special classes $\mathcal{F}$ and $\mathcal{G}$.
Such a simple model will help us to introduce suitable weak dependence conditions for model-based bootstrap procedures.
Those lectures essentially aim at proving that such weak dependence conditions also allow to develop a simple asymptotic theory (see section 11.5).
Return to inequality (11.1), which right hand side will be written for simplicity $\mathrm{Cov}\,(\mathbf{f},\mathbf{g})$ with $\mathbf{f}\equiv f(X_{i_1},\dots,X_{i_u})$ and $\mathbf{g}\equiv g(X_{j_1},\dots,X_{j_v})$.
Then we consider a simple linear (infinite moving average) model defined through an independent identically distributed sequence with finite first order moments $(\xi_t)_{t\in\mathbb{Z}}$:

$$X_t=\sum_{k\in\mathbb{Z}} a_k\xi_{t-k}.$$

The previous series converge in $\mathbb{L}^1$ in case

$$\|\xi_0\|_1<\infty,\qquad\text{and}\qquad\sum_t |a_t|<\infty,$$

and the considered process is then stationary; it corresponds to

$$H((u_t)_{t\in\mathbb{Z}})=\sum_{t\in\mathbb{Z}} a_t u_{-t}.$$

Then the model is said to be causal in case $a_k=0$ for $k<0$ since X_t is measurable with respect to $\mathcal{F}_t=\sigma(\xi_s;\, s\le t)$. Set

$$X_t^{(p)}=\sum_{|k|\le p} a_k\xi_{t-k},\quad \tilde X_t^{(p)}=\sum_{0\le k\le p} a_k\xi_{t-k},$$

then this is simple to check that $X_s^{(p)}$ and $X_t^{(p)}$ are independent if $|t-s|>2p$ and in case $r>2p$ this also implies that also $\mathbf{f}'$ and $\mathbf{g}'$ are independent when setting $\mathbf{f}'=f(X_{i_1}^{(p)},\ldots,X_{i_u}^{(p)})$ and $\mathbf{g}'=g(X_{j_1}^{(p)},\ldots,X_{j_v}^{(p)})$ and thus

$$\mathrm{Cov}\,(\mathbf{f},\mathbf{g})=\mathrm{Cov}\,(\mathbf{f},\mathbf{g}-\mathbf{g}')+\mathrm{Cov}\,(\mathbf{f}-\mathbf{f}',\mathbf{g}').$$

Now if the involved functions are both bounded above by 1 then

$$|\mathrm{Cov}\,(\mathbf{f},\mathbf{g}-g')|\le 2\mathbb{E}|\mathbf{g}-\mathbf{g}'|,\quad |\mathrm{Cov}\,(\mathbf{f}-\mathbf{f}',\mathbf{g}')|\le 2\mathbb{E}|\mathbf{f}-\mathbf{f}'|.$$

If now those functions are Lipschitz then *eg.*

$$|\mathbf{f}-\mathbf{f}'|\le \mathrm{Lip}\,f\sum_{s=1}^{u}|X_{i_s}-X_{i_s}^{(p)}|.$$

We should also notice that for each t,

$$\mathbb{E}|X_t-X_t^{(p)}|\le \mathbb{E}|\xi_0|\sum_{k>p}|a_k|.$$

Hence doing the same with g we derive:

$$|\mathrm{Cov}\,(\mathbf{f},\mathbf{g}-\mathbf{g}')|\le 2v\mathbb{E}|\xi_0|\sum_{k>p}|a_k| \text{ and } |\mathrm{Cov}\,(\mathbf{f}-\mathbf{f}',\mathbf{g}')|\le 2u\mathbb{E}|\xi_0|\sum_{k>p}|a|.$$

(we use the bound $|\mathrm{Cov}\,(U,V)|\le 2\|U\|_\infty\,\mathbb{E}|V|$).
From another hand, in the causal case we derive that $\mathbf{f}$ and $\tilde{\mathbf{g}}$ are independent for $\tilde{\mathbf{g}}=g(\tilde X_{j_1}^{(p)},\ldots,\tilde X_{j_v}^{(p)})$ which implies $\mathrm{Cov}\,(\mathbf{f},\mathbf{g})=\mathrm{Cov}\,(\mathbf{f},\mathbf{g}-\tilde{\mathbf{g}})$ and thus analogously we obtain:

- $|\mathrm{Cov}\,(\mathbf{f},\mathbf{g})|\le(u\mathrm{Lip}\,f+v\mathrm{Lip}\,g)\epsilon_r$ if we set $\epsilon_r=2\mathbb{E}|\xi_0|\sum_{|i|>2r}|a_i|$ for non causal linear processes

$$X_n=\sum_{i=-\infty}^{\infty}a_i\xi_{n-i}$$

- $|\mathrm{Cov}\,(\mathbf{f},\mathbf{g})|\le v\mathrm{Lip}\,g\cdot\epsilon_r$ if respectively $\epsilon_r=2\mathbb{E}|\xi_0|\sum_{i>r}|a_i|$ for the causal case, $a_i=0$ if $i<0$.

Most of the previous models satisfy such conditions as for Bernoulli schemes. A similar proof follows the same lines for general Bernoulli shifts. Set here

$$\begin{aligned}
\epsilon_r &= 2\sup_{q>2r}\mathbb{E}\left|H\left((\xi_i)_{i\in\mathbb{Z}}\right)-H\left((\xi_i)_{|i|\le q}\right)\right|, \\
&\quad\ \text{for non-causal Bernouilli shifts, } X_t = H(\ldots,\xi_{t+1},\xi_t,\xi_{t-1},\ldots). \\
&= 2\sup_{q>r}\mathbb{E}\left|H\left((\xi_i)_{i\in\mathbb{N}}\right)-H\left((\xi_i)_{0\le i\le q}\right)\right|, \\
&\quad \text{for causal Bernouilli shifts } X_t = H(\xi_t,\xi_{t-1},\ldots),
\end{aligned}$$

the sequence $(\xi_i)_{|i|\le r}$ is obtained by setting 0 for indices with $|i|>r$. Then $\epsilon_r \downarrow 0$ as $r\uparrow\infty$ ([1]). and the forthcoming conditions ψ_θ or ψ_η apply according the fact the the Bernoulli is causal or not.

[Doukhan and Louhichi, 1999] synthesize such easy conditions in terms of Lipschitz classes.
The present lectures are not exhaustive so that we will restrain the really general notions in Definition 11.1.1 to some few examples of such weak dependence.

Definition 11.4.1. *Set $\mathcal{L}$ the class of functions $g:\mathbb{R}^v\to\mathbb{R}$ for some integer $v\ge 1$, with $\|g\|_\infty\le 1$ and $Lip(g)<\infty$ where:*

$$Lip(g) = \sup_{(x_1,\ldots,x_v)\neq(y_1,\ldots,y_v)}\frac{|g(x_1,\ldots,x_v)-g(y_1,\ldots,y_v)|}{|x_1-y_1|+\cdots+|x_v-y_v|}$$

Some weak dependence conditions correspond to $\mathcal{G}=\mathcal{L}$ and respectively $\mathcal{F}=\mathcal{L}$ (non causal case) or

$$\mathcal{F}=\mathbb{B}_\infty=\{f:\mathbb{R}\to\mathbb{R},\ measurable, \|f\|_\infty\le 1\} \qquad (causal\ case).$$

Here respectively

$$\begin{aligned}
\psi(u,v,f,g) &= \psi_\eta(u,v,f,g)\equiv uLip(f)+vLip(g), \\
&= \psi_\theta(u,v,f,g)\equiv vLip(g \\
&= \psi_\kappa(u,v,f,g)\equiv uvLip(g)Lip(g), \\
&= \psi_\lambda(u,v,f,g)\equiv uLip(f)+vLip(g)+uvLip(g)
\end{aligned}$$

[1]Quote that for the special case of the previous linear processes, the present bound ϵ_r is in fact sharper that the previous series remainders.

Then the process $(X_t)_{t\in\mathbb{Z}}$ is η (resp. θ,κ or λ)$-$weakly dependentttt in case the least corresponding sequence ϵ_r given from relation (11.1) converges to 0 as $r\uparrow\infty$, the respective coefficients will be denoted η_r, θ_r,κ_r or λ_r.

Example 11.4.1 (Dependence decay rates)**.** *To derive limit theorems this will be essential to know decay rates of decorrelation as well as the existence of moments. The following examples aim at feeling this important gap.*

- *Conditions η, and θ are checked before to hold for linear causal or non-causal sequences.*
  They also hold analogously for Bernoulli shifts under assumptions (7.14) if they are $\mathbb{L}^1-$weakly dependent (see Definition 7.4.3. Here respectively

$$\begin{array}{rcll}\theta_r &=& 2\delta_r^{(1)}, & \text{under condition } \psi_\theta,\\ \eta_r &=& 2\delta_{[r/2]}^{(1)}, & \text{under condition } \psi_\eta.\end{array}$$

  - *Examples of such causal models are Markov stable processes (see Definition 7.3.1) satisfy those relations as proved in Theorem 7.3.1. Such Markov models 7.2 indeed satisfy the inequality 7.5.*
    item Linear and Volterra type processes are also weakly dependent and tails of coefficients allow to bound ϵ_r in both the causal and the non-causal case.
  - *In order to consider an explicit example of a chaotic expansion. LARCH(∞)–models in § 7.2.2 are solutions of the recursion*

$$X_n = \left(b_0 + \sum_{j=1}^{\infty} b_j X_{n-j}\right)\xi_n.$$

    The $\mathbb{L}^p-$valued strictly stationary solution of this recursion writes $X_n=\sum_{k=1}^{\infty} S_n^{(k)}$ with:

$$S_n^{(k)} = b_0 \sum_{l_1,..,l_k=1}^{\infty} b_{l_1}\cdots b_{l_k}\xi_{n-l_1}\xi_{n-l_1-l_2}\cdots\xi_{n-(l_1+\cdot\cdot+l_k)}.$$

Under the condition $B = \|\xi_0\|_p \sum_{l=1}^{\infty} |b_l| < 1$, *this is simple to derive with independence of all those products that* $\|S_n^{(k)}\|_p \le |b_0| a^k$. *Now set* $S_n^{(k,L)}$ *for the finite sum where each of the indices satisfy* $1 \le l_1, \ldots, l_k \le L$ *then analogously*

$$\|S_n^{(k)} - S_n^{(k,L)}\|p \le k|b_0|a^{k-1}B_L, \ \text{with } B_L = \|\xi_0\|_p \sum_{l>L} |b_l|,$$

where the factor k *comes from the fact that in order that only the tail of a series appears, it may appear at any position in those multiple sums.*
Turning now to $p = 1$, *we approximate* X_n *by the following* $(LK)-$*dependent sequence*

$$X_n^{(K,L)} = S_n^{(0)} + S_n^{(1,L)} + \cdots + S_n^{(K,L)}$$

then previous calculations prove that for a constant $C > 0$, $\|X_n - X_n^{(K,L)}\|_1 \le C(B_L + a^K)$. *Let* $L, K \ge 1$ *be such that* $LK \le r$ *then this implies that wrt* ψ_θ,

$$\theta_r \le C \inf_{1 \le L \le r} \left(B_L + a^{\frac{r}{L}} \right).$$

Eg. if $b_l = 0$ *for* $l > L$ *large enough then* $\theta_r \le C a^{\frac{r}{L}}$, *if* B_L *h decays geometrically to 0, then* $\theta_r \le C e^{-c\sqrt{r}}$, *and if* $B_L \le cL^{-\beta}$ *then* $\theta_r \le c' r^{-\beta}$.

- *Either Gaussian processes or associated random processes (in* $\mathbb{L}^2$*) are* $\kappa-$*weakly dependent because of Lemma 8.1. Here*

$$\kappa_r = \max_{|j-i| \ge r} |Cov(X_i, X_j)|,$$

is convenient from inequality (8.1) (in this case absolute values are useless); this inequality holds in the Gaussian case as this is proved eg. in [Dedecker et al., 2007].

- *Now the function* ψ_λ *allows to combine both difficulties. For example the sum of a Bernoulli shift and of an independent associated process may satisfy such conditions.*

Remark 11.4.1 (A comparison). *A strict comparison of the previous strong mixing conditions and all such weak dependence is not not always possible. Anyway α_r and θ_r are obtained in inequality (11.1)tt as the supremum of covariances $|Cov(f(P), g(F)|$ for functions of past and future where $\|f\|_\infty \le 1$ and where $\|g\|_\infty \le 1$ under mixing or where moreover $Lip\, g \le 1$ under weak dependence, thus*

$$\theta_r \le \alpha_r.$$

Various applications of those notions are considered in our monograph [Dedecker et al., 2007]. It is however simple to quote that such properties are stable through Lipschitz images as an extension of Lemma 7.4.1.
The function $g(x_1, \ldots, x_u) = x_1 \cdots x_u$ is more specifically used in the next Chapter, it is usually unbounded and non-Lipschitz so that truncations will be needed to derive moment inequalities for partial sums. The following Exercise is a first step to consider the empirical cdf. Various generalizations of which may be found in [Dedecker et al., 2007].

Exercise 22 (Heredity). *Let $(X_t)_{t\in\mathbb{Z}}$ be a real valued and $\theta-$weakly dependent process. Assume that there exists a constant $C > 0$ such that $\mathbb{P}(X_i \in [a, b]) \le C(b - a)$ for each $-\infty < a < b < \infty$.*
Then: $|Cov(g(X_0), g(X_r))| \le (1 + C)\sqrt{\theta_r}$.

Proof. Set g_ϵ the continuous function such that $g_\epsilon(x) = g(x)$ if $x < u$ and $x > u + \epsilon$, and g_ϵ is affine on $[u, u + \epsilon]$ then $\mathrm{Lip}\, g_\epsilon = 1/\epsilon$:

$$\begin{aligned}
&|\mathrm{Cov}\,(g(X_0), g(X_r))| \\
&\qquad \le |\mathrm{Cov}\,(g(X_0), g(X_r) - g_\epsilon(X_r))| + |\mathrm{Cov}\,(g(X_0), g_\epsilon(X_r))| \\
&\qquad\qquad\qquad C\epsilon + \frac{1}{\epsilon}\theta_r = (1 + C)\sqrt{\theta_r}, \quad \text{with } \epsilon^2 = \theta_r.
\end{aligned}$$

(use the bound $|\mathrm{Cov}\,(U, V) \le 2\|U\|_\infty \mathbb{E}|V||$ to conclude.) ∎

For non-causal weak dependences, use

$$\begin{aligned}
|\mathrm{Cov}\,(g(X_0), g(X_r))| &\le |\mathrm{Cov}\,(g(X_0), g(X_r) - g_\epsilon(X_r))| \\
&\quad + |\mathrm{Cov}\,(g(X_0) - g_\epsilon(X_0)), g_\epsilon(X_r))| + |\mathrm{Cov}\,(g_\epsilon(X_0)), g_\epsilon(X_r))|
\end{aligned}$$

11.5 Proving limit theorems

Here follows a simple way to derive CLTs. The situation chosen is that of stationary and centered processes.
In fact ergodicity really allows to recenter such processes.
Moment inequalities proved later in Chapter 12 yield good controls for $\mathbb{E}|S_n|^p$ and a central limit theorem may be derived by using the following simple dependent Lindeberg CLT:

Lemma 11.5.1 (Dependent Lindeberg [Bardet et al., 2006])**.** *We set $f(x) = e^{i<t,x>}$ for each $t \in \mathbb{R}^d$ (with $< a, b >$ scalar product in $\mathbb{R}^d$) and we consider an integer $k \in \mathbb{N}^*$.*
Let $(X_i)_{1\le i\le k}$ be $\mathbb{R}^d-$valued centered random variables such that:

$$A_k = \sum_{i=1}^{k} \mathbb{E}\|X_i\|^{2+\delta} < \infty.$$

Set

$$T(k) = \sum_{j=1}^{k} \big| Cov(e^{i<t,X_1+\cdots+X_{j-1}>}, e^{i<t,X_j>}) \big|.$$

Then

$$\Delta_k \le T(k) + 6\|t\|^{2+\delta} A_k.$$

Proof. Going ahead with the proof of Lemma 2.1.1 we will only reconsider the bound of $\mathbb{E}\delta_j$, for this let a random variable U_j^* be independent of all the other random variables already considered and with the same distribution as U_j.
Then we decompose:

$$\delta_j = (g(Z_j + U_j) - g(Z_j + U_j^*)) + (g(Z_j + U_j^*) - g(Z_j + V_j))$$

The second term admits the bound provided in Lemma 2.1.1, which writes as stated above since for $f(x) = e^{i<t,x>}$ one easily derive that $\|f^{(p)}\|_\infty = |t|^p$. Now the first term is the "dependent" one and from the independence of V's and multiplicative properties of the exponential:

$$\begin{aligned}|\mathbb{E}(g(Z_j+U_j)-g(Z_j+U_j^*))| &\le |\mathbb{E}g(U_1+\cdots+U_{j-1})(g(U_j)-g(U_j^*))| \\ &= |\mathrm{Cov}\,(g(U_1+\cdots+U_{j-1}), g(U_j))|. \qquad \blacksquare\end{aligned}$$

In case the series of covariances is summable we already quoted that

$$\mathbb{E}S_m^2 \sim \sigma^2 m, \qquad \text{for large values of} \quad m.$$

The idea is to calculate

$$\Delta_n = \mathbb{E}\left(f\left(\frac{S_n}{\sqrt{n}}\right) - f(\sigma N)\right)$$

for enough functions in the class $\mathcal{C}^3$ class.
We need to derive $\Delta_n \to 0$ as $n \to \infty$. For this we sketch Bernstein blocks technique To this end consider sequences

$$q = q(n) \ll p = p(n) \ll n \qquad \text{as} \qquad n \uparrow \infty.$$

Then we may write

$$\frac{S_n}{\sqrt{n}} = U_1 + \cdots + U_k + V,$$

with $k = k(n) = \left[\frac{n}{p(n) + q(n)}\right]$ and

$$U_j = \frac{1}{\sqrt{n}} \sum_{i=(j-1)(p+q)+1}^{(j-1)(p+q)+p} X_i$$

In this case the remainder $\|V\|_2 \to 0$ because $V = \frac{1}{\sqrt{n}} \sum_{u \in E} X_u$ is a sum over some set E with cardinality $m \le q + p = o(n)$.
Indeed

$$\begin{aligned} n\mathrm{Var}\, V \le \sum_{u,v \in E} |\mathrm{Cov}\,(X_u, X_v)| \\ = \sum_{u,v \in E} |\mathrm{Cov}\,(X_0, X_{v-u})| \\ \le m \sum_{j=-\infty}^{\infty} |\mathrm{Cov}\,(X_0, X_j)|. \end{aligned}$$

The variables U are almost independent since they are all distant at least q so that Lemma 11.5.1 may be applied. ∎

To conclude we cite a result which is adapted to causal cases (following from a very different proof to see eg. in [Rio, 2000]):

Theorem 11.5.1 (Dedecker & Rio, 1998)**.** *Let $(X_n)_{n\in\mathbb{Z}}$ be an ergodic stationary sequence with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$. Set $S_n = X_1 + \cdots + X_n$. Assume that the random series $\sum_{n=0}^{\infty} X_0 \mathbb{E}\Big(X_n \Big| \sigma(X_k/k \le 0)\Big)$ converges in $\mathbb{L}^1$.*
Then the sequence $\mathbb{E}(X_0^2 + 2X_0 S_n)$ converges to some σ^2 and

$$\frac{1}{\sqrt{n}} S_{[nt]} \to_{n\to\infty} \sigma W_t, \qquad \text{in distribution in } D[0,1] \quad (^2).$$

In [Dedecker and Doukhan, 2003] for the case of $\theta-$weak dependence and in [Dedecker et al., 2007] we derive analogue CLTs; quote that assumptions needed to replace such abstract conditions always write in terms of decay rates and moment conditions. Previously introduced Conditions take into account most of the standard models in a statistics. We do not recall those results because the aim is more pedagogical here that formal and we defer a reader to the monograph [Dedecker et al., 2007].

Exercise 23. *Consider a sequence of iid random variables R_i (with finite mean) and an independent standard normal random variable N then we may set $X_k = R_k N$,*

1. *Set $\overline{X}_n = (X_1 + \cdots + X_n)/n$ then $\lim_{n\to\infty} \overline{X}_n = \mathbb{E}R_0 \cdot N$ a.s.*
2. *Deduce that this sequence is not ergodic in case $\mathbb{E}R_0 \neq 0$.*
3. *If R_k follows Rademacher distribution ($\mathbb{P}(R_k = \pm 1) = \frac{1}{2}$) prove that $Cov(X_k, X_\ell) = 0$ for all $k \neq \ell$ and the sequence is not independent.*
4. *In this Rademacher case prove that $\sqrt{n}\overline{X}_n$ converges in distribution to the product of 2 standard normal random variables.*
5. *Prove that the sequence $(X_k)_k$ is never ergodic.*

[2]For Skorohod space see Definition A.3.2) and the Remark following it.

Hints for Exercise 23.

1. The first point follows from the Strong Law of Large Numbers.
2. Ergodic theorem (Corollary 9.1.3) does not hold since the limit is non-deterministic, which proves non-ergodicity.
3. This point is proved in Exercise 2.
4. It follows from the CLT.
5. This sequence is never ergodic since conditionally to N it is ergodic and the tail sigma-field is always the sigma-field generated by N. ∎

Example 11.5.1. *With the previous Exercise 23 we obtain an orthogonal stationary sequence of Gaussian random variables such that the law of large numbers holds, but which is not ergodic and does not satisfy the CLT. This sequence is thus not a Gaussian process.*

Remark 11.5.1. *The empirical process*

$$Z_n(x) = \sqrt{n}(F_n(x) - F(x)), \qquad F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1\!\!I_{\{X_k \le x\}}$$

is also of a main interest and one may consider it as above but in this case heredity of weak dependence conditions is no more ensured directly since the function $u \mapsto 1\!\!I_{\{u \le x\}}$ is not Lipschitz but concentration conditions as in Lemma 7.4.2 allow to work out the asymptotic properties for such processes.
In the Remark following Definition A.3.2 we recall a criterion for the convergence of this cumulative repartition distribution.
Anyway we refer again to [Dedecker et al., 2007] for more details.

Exercise 24 (subsampling)**.** *Prove eqn. (11.4) for a bounded function g, and deduce asymptotic sketched in § 4.6:*

- *Under strong mixing this is enough to assume* $\lim_r \alpha_r = 0$*, and does not rely on the properties of functions t_b.*

- *Under $\theta-$weak dependence, use Lipschitz properties of the functions t_m (set $L_m = \mathrm{Lip}\, t_m$) to derive a consistency result both under overlapping and non-overlapping schemes.*

Hint. The bound eqn. (4.2) yields with $N = \text{Card}\, E_{m,n}$ for each case and with immediate notations:

$$\mathrm{Var}\, \widehat{K}(g) \le \frac{1}{N} \sum_{i=1}^{N} |\mathrm{Cov}\,(g_{0,m}, g_{i,m})|.$$

In the first strong mixing case

$$|\mathrm{Cov}\,(g_{0,m}, g_{i,m})| \le \alpha_{r-m+1}, \tag{11.4}$$

(resp. $\le \alpha_{r/m+1}$) which does not depend on m for this special mixing case. Hence Cesaro lemma yields the result for this case.
The cases of weak dependence are more complex, here $\mathrm{Lip}\, g_m \le L_m \mathrm{Lip}\, g$ and thus:

$$|\mathrm{Cov}\,(g_{0,m}, g_{i,m})| \le m L_m \mathrm{Lip}\, g\, \theta_{r-m+1}, \tag{11.5}$$

(resp. $\le m L_m \mathrm{Lip}\, g\, \theta_{\frac{r}{m}+1}$)here in the overlapping scheme e.g. we otain for some constant:

$$\mathrm{Var}\, \widehat{K}(g) \le \frac{Cm}{n-m}\Big(1 + L_m \mathrm{Lip}\, g \sum_{i=1}^{n-m} |\mathrm{Cov}\,(g_{0,m}, g_{i,m})|\Big).$$

The cases of sub empirical means $T_m = \sqrt{m}\overline{X}_m$, and $L_m = \frac{C}{\sqrt{m}}$ and for kernel density estimates $\sqrt{mh}(f_{m,h} - \mathbb{E} f_{m,h})$, $L_m = \frac{C}{h\sqrt{mh}}$.
Eg. in the first case of subsampling means with overlapping scheme, the assumptions

$$\frac{m_n}{n} \to 0, \quad \frac{1}{\sqrt{m_n}} \sum_{i=1}^{n} \theta_i \text{ is bounded,}$$

together imply consistency of subsampling. This hold for instance if $\sum_{i=1}^{\infty} \theta_i < \infty$ and $\lim_n m_n/n = 0$.
To derive the above inequality for discontinuous functions $g = 1\!\!1_{\{\cdot \le u\}}$ one additional step is necessary, use Exercise 22.
Finally uniform convergence use again by using Exercise 24.
Higher order moments are considered in Exercise 31. ∎

Chapter 12

Moments and cumulants

The Chapter is devoted to moment methods: this tool is essential for time series with finite moments.
Use of moments relies on their importance to derive asymptotic of several estimators, based on moments and limit distributions.
Cumulants are linked with spectral or multispectral estimation which are main tools of time series analysis.

$$g(\lambda) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\,(X_0, X_k)e^{-ik\lambda}.$$

Such functions do not characterize the dependence of non linear processes: indeed we already had examples of orthogonal but not independent sequences. This motivated the introduction of higher order characteristics.
E.g. a multispectral density is defined over $\mathbb{C}^{p-1}$ by

$$g(\lambda_2, \ldots, \lambda_p) = \sum_{k_2=-\infty}^{\infty} \cdots \sum_{k_p=-\infty}^{\infty} \kappa(X_0, X_{k_2}, \ldots, X_{k_p})e^{-i(k_2\lambda_2+\cdots+k_p\lambda_p)}.$$

Remark 12.0.2. *The periodogram in Definition 4.4.1 may be not enough to deal with non-Gaussian stationary time series because:*

- *As quoted in Exercise 2 covariances are not enough to prove independence. Thus in order to test for stochastic independence the use of higher order spectral estimates appear reasonable.*

- *Gaussian laws are characterized by the fact that cumulants with order > 2 vanish. This provides hints to test Gaussianness.*

12.1 Method of moments

Recall that the method of moments yields limit theorems:

Theorem 12.1.1 (Feller). *If the sequence of real valued random variables U_n is such that*

$$\mathbb{E}U_n^p \to_{n\to\infty} \mathbb{E}U^p, \qquad \text{for each integer } p \geq 0,$$

then

$$U_n \to_{n\to\infty} U, \qquad \text{in law.}$$

If moreover U admits an analytic Laplace transform around 0. This holds if there exists $\alpha > 0$ with $\mathbb{E}e^{\alpha|U|} < \infty$.

Hint. Indeed from the *analytic continuation theorem* implies that U's distribution is characterized through its moments. ∎

12.1.1 Notations

Let $Y = (Y_1, \ldots, Y_k) \in \mathbb{R}^k$ be a random vector we set

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}e^{it\cdot Y} = \mathbb{E}\exp\Big(i\sum_{j=1}^k t_j Y_j\Big), \\ m_p(Y) &= \mathbb{E}Y_1^{p_1}\cdots Y_k^{p_k}, \end{aligned}$$

where

$$\begin{array}{rclcl} p &=& (p_1, \ldots, p_k), & t = (t_1, \ldots, t_k) & \in \ \mathbb{R}^k, \\ |p| &=& p_1 + \cdots + p_k = r, & \mathbb{E}(|Y_1|^r + \cdots + |Y_k|^r) & < \ \infty. \end{array}$$

Denote

$$p! = p_1!\cdots p_k!, \qquad t^p = t_1^{p_1}\cdots t_k^{p_k},$$

for $t = (t_1, \dots, t_k) \in \mathbb{R}^k$, and $p = (p_1, \dots, p_k)$.
In case the previous condition holds for some integer $r \in \mathbb{N}^*$, the function $t \mapsto \log \phi_Y(t)$ admits a Taylor expansion

$$\log \phi_Y(t) = \sum_{|p| \le r} \frac{i^{|p|}}{p!}\, \kappa_p(Y)\, t^p + o(|t|^r), \qquad \text{as } t \to 0.$$

Coefficients $\kappa_p(Y)$ are named cumulants of Y with order $p \in \mathbb{N}^k$ and they exist if $|p| \le r$.
Replace Y by a vector with higher dimension $s = |p|$ with p_1 repetitions for Y_1, ..., p_k repetitions for Y_k allows to consider $p = (1, \dots, 1)$ and we set $\kappa_{(1,\dots,1)}(Y) = \kappa(Y)$.
If $\mu = \{i_1, \dots, i_u\} \subset \{1, \dots, k\}$ set:

$$\kappa_\mu(Y) = \kappa(Y_{i_1}, \dots, Y_{i_u}), \qquad m_\mu(Y) = m(Y_{i_1}, \dots, Y_{i_u}).$$

Identifying Taylor expansions Leonov and Shyraev (1959) (for a proof see [Rosenblatt, 1985], pages 33–34), derived the relations

$$\kappa(Y) = \sum_{u=1}^{k} (-1)^{u-1}(u-1)! \sum_{\mu_1,\dots,\mu_u} \prod_{j=1}^{u} m_{\mu_j}(Y) \tag{12.1}$$

$$m(Y) = \sum_{u=1}^{k} \sum_{\mu_1,\dots,\mu_u} \prod_{j=1}^{u} \kappa_{\mu_j}(Y) \tag{12.2}$$

Previous sums are taken over all the partitions $\mu_1, \dots, \mu_u$ of the set $\{1, \dots, k\}$. The Taylor expansion of $s \mapsto \log(1+s)$ as $t \to 0$ successively yields

$$\begin{aligned}
\phi_Y(t) &= 1 + \sum_{0<|p|\le r} \frac{i^{|p|}}{p!} m_p(Y) t^p + o(|t|^r), \\
\log \phi_Y(t) &= \sum_{u=1}^{r} \frac{(-1)^{u-1}}{u} \left(\sum_{0<|p|\le r} \frac{i^{|p|}}{p!} m_p(Y) t^p \right)^u + o(|t|^r) \\
&= \sum_{u=1}^{r} \frac{(-1)^{u-1}}{u} \sum_{\substack{0<|p|\le r \\ p_1+\cdots+p_u=p}} \frac{(it)^{|p|}}{p!} \prod_{j=1}^{u} m_{p_j}(Y) + o(|t|^r)
\end{aligned}$$

hence identifying the coefficient corresponding to $p = (1, \ldots, 1)$ for $u-$tuples such that $p_1 + \cdots + p_u = p$; choose $r = k$ to derive relation (12.1). Indeed then $|p| = k, p! = 1$ and $(it)^p = i^k t^k$.
A combinatoric coefficient $u!$ appears, it corresponds to the number of the permutations in a partition. ∎

12.1.2 Combinatorics of moments

Recall now some notions from [Saulis and Statulevicius, 1991]'s book.

Definition 12.1.1. *Centered moments of the random vector $Y = (Y_1, \ldots, Y_k)$ are defined with $\widehat{\mathbb{E}}\,(Y_1, \ldots, Y_l) = \mathbb{E}Y_1 c(Y_2, \ldots, Y_l)$ where centered random variable $c(Y_2, \ldots, Y_l)$ are recursively identified by setting $c(\xi_1) = \overbrace{\ \xi_1\ } = \xi_1 - \mathbb{E}\xi_1$ and*

$$\begin{aligned} c(\xi_j, \xi_{j-1}, \ldots, \xi_1) &= \xi_j\, \overbrace{c(\xi_{j-1}, \ldots, \xi_1)} \\ &= \xi_j \left(c(\xi_{j-1}, \ldots, \xi_1) - \mathbb{E}c(\xi_{j-1}, \ldots, \xi_1) \right). \end{aligned}$$

Consider $Y_\mu = (Y_j / j \in \mu\}$ as a $p-$tuple l for $\mu \subset \{1, \ldots, k\}$.

For example $\widehat{\mathbb{E}}\,(\xi) = 0, \widehat{\mathbb{E}}\,(\eta, \xi) = \mathrm{Cov}\,(\eta, \xi)$,

$$\widehat{\mathbb{E}}\,(\zeta, \eta, \xi) = \mathbb{E}(\zeta\eta\xi) - \mathbb{E}(\zeta)\mathbb{E}(\eta\xi) - \mathbb{E}(\eta)\mathbb{E}(\zeta\xi) - \mathbb{E}(\xi)\mathbb{E}(\zeta\eta).$$

Centered moments are a way to generalize covariances. They also say about independence of the coordinates for a random vector.

The nice following result explains the nature of cumulants . This provides a representation in terms of centered moments.

Theorem 12.1.2 ([Saulis and Statulevicius, 1991])**.**

$$\kappa(Y_1, \ldots, Y_k) = \sum_{u=1}^{k} (-1)^{u-1} \sum_{\mu_1, \ldots, \mu_u} N_u(\mu_1, \ldots, \mu_u) \prod_{j=1}^{u} \widehat{\mathbb{E}}\, Y_{\mu_j}$$

sums are over all the partitions $\mu_1, \ldots, \mu_u$ of the set $\{1, \ldots, k\}$ and the integers $N_u(\mu_1, \ldots, \mu_u) \in \left[0, (u-1)! \wedge \left[\frac{k}{2}\right]!\right]$ defined for each par-

tition satisfy

$$\begin{aligned} N(k,u) &= \sum_{\mu_1,\ldots,\mu_u} N_u(\mu_1,\ldots,\mu_u) \\ &= \sum_{j=1}^{u-1} C_k^j (u-j)^{k-1}, \end{aligned}$$

and $\sum_{u=1}^{k} N(k,u) = (k-1)!$.

The following bound is a simple consequence of this result

Lemma 12.1.1. *Let $Y_1,\ldots,Y_k \in \mathbb{R}$ be centered random variables. For each $k \geq 1$ set $M_k = (k-1)! 2^{k-1} \max_{1\leq i\leq k} \mathbb{E}|Y_i|^k$ then*

$$\begin{aligned} |\kappa(Y_1,\ldots,Y_k)| &\leq M_k, & (12.3)\\ M_k M_l &\leq M_{k+l}, \text{ for } k,l \geq 2. & (12.4) \end{aligned}$$

Hence:

$$\prod_{i=1}^{u} \left|\kappa(Y_1,\ldots,Y_{p_u})\right| \leq M_{p_1+\cdots+p_u}. \tag{12.5}$$

Proof of Lemma 12.1.1. The second point follow from inequality $a!\,b! \leq (a+b)!$ also written $C_{a+b}^a \geq 1$ and the second comes from Lemma 12.1.2 and of the forthcoming Lemma. ∎

Lemma 12.1.2. *For each $j,p \geq 1$ and for all the real valued random variables*

$$\|c(\xi_j,\xi_{j-1},\ldots,\xi_1)\|_p \leq 2^j \max_{1\leq i\leq j} \|\xi_i\|_{pj}^j \qquad (\textit{with } \|\xi\|_q = \mathbb{E}^{1/q}|\xi|^q).$$

Proof of Lemma 12.1.2. Jensen inequality (Proposition A.1.1) leads to

$$\|c(\xi_1)\|_p \leq \|\xi_1\|_p + |\mathbb{E}\xi_1| \leq 2\|\xi_1\|_p.$$

Set $Z_j = c(\xi_j,\xi_{j-1},\ldots,\xi_1)$ then $Z_j = \xi_j(Z_{j-1} - \mathbb{E}Z_{j-1})$ and from Hölder inequality

$$\|\xi_j Z_{j-1}\|_p^p \leq \|\xi_j\|_{pj}^p \|Z_{j-1}\|_{\frac{pj}{j-1}}^p.$$

Thus using the recursion assumption for the pair $(q, j-1)$ where $q = pj/(j-1)$ Minkowski and Hölder inequalities give

$$\begin{aligned}
\|Z_j\|_p &\le \|\xi_j Z_{j-1}\|_p + \|\xi_j\|_p |\mathbb{E} Z_{j-1}|, \\
&\le 2\|\xi_j\|_{pj} \|Z_{j-1}\|_q, && (q = p \cdot \frac{j}{j-1}) \\
&\le 2^j \|\xi_j\|_{pj} \max_{0\le i<j} \|\xi_i\|_{q(j-1)}^{j-1}, \\
&\le 2^j \max_{0\le i\le j} \|\xi_i\|_{pj}^j,
\end{aligned}$$

because $q(j-1) = pj$ which allows to conclude. ∎

Proof of Lemma 12.1.1. Omit suprema by replacing $\max_{j\le J} \|Y_j\|_p$ by $\|Y_0\|_p$ for the sake of simplicity.
Lemma 12.1.2 yields $|\widehat{\mathbb{E}}\ Y_\mu| \le 2^{l-1}\|Y_0\|_l^l$ with $l = \#\mu$. Indeed write $Z = c(Y_2, \ldots, Y_l)$ and define p through the identity $\frac{1}{p} + \frac{1}{l} = 1$.
Then

$$\left|\widehat{\mathbb{E}}\ (Y_1, \ldots, Y_l)\right| = |\mathbb{E} Y_1 Z| \le \|Y_0\|_l \|Z\|_p \le 2^{l-1}\, \|Y_0\|_l^l$$

since $p(l-1) = l$. Theorem 12.1.2 implies

$$\begin{aligned}
|\kappa(Y)| &\le \sum_{u=1}^k \sum_{\mu_1,\ldots,\mu_u} N_u(\mu_1, \ldots, \mu_u) \prod_{i=1}^u 2^{\#\mu_i - 1} \|Y_0\|_{\#\mu_i}^{\#\mu_i} \\
&\le \sum_{u=1}^k 2^{k-u}\, N(k,u) \|Y_0\|_k^k \\
&\le 2^{k-1}\, \|Y_0\|_k^k \sum_{u=1}^k N(k,u) \\
&= 2^{k-1}(k-1)!\|Y_0\|_k^k
\end{aligned}$$

ending the proof. ∎

12.2 Dependence and cumulants

The following Lemmas are essentially proved for sequences of real valued random variables $(X_n)_{n\in\mathbb{Z}}$ in [Doukhan and León, 1989].

12.2.1 More dependence coefficients

Consider a stationary real valued sequence $(X_n)_{n\in\mathbb{Z}}$. Then as in [Doukhan and Louhichi, 1999] considert

$$c_{X,q}(r) = \max_{1\le l<q} \sup_{\substack{t_1 \le \cdots \le t_q \\ t_{l+1}-t_l \ge r}} \left|\mathrm{Cov}\left(X_{t_1}\cdots X_{t_l}, X_{t_{l+1}}\cdots X_{t_q}\right)\right| \quad (12.6)$$

Example 12.2.1. *Assume that the $\eta-$weak dependenceindexdependence coefficient, η_r condition (11.1) associated with the functional ψ_η and with the classes of function $\mathcal{F}=\mathcal{G}=\mathcal{L}$ holds. If $Y_i = h(X_i)$ for some Lipschitz function h bounded by M, we get*

$$c_{Y,q}(r) \le M^{q-1} Lip(h)\theta_r.$$

The following coefficients are also useful

$$c^{\star}_{X,q}(r) = \max_{1\le l\le q} c_{X,l}(r)\cdot\mu_{q-l}, \qquad \text{with } \mu_t = \mathbb{E}|X_0|^t. \quad (12.7)$$

Define

$$\kappa_q(t_2,\ldots,t_q) = \kappa_{(1,\ldots,1)}(X_0, X_{t_2},\ldots,X_{t_q}).$$

The forthcoming decomposition explain the way cumulants behave as covariances.
Precisely it proves that cumulants $\kappa_Q(X_{k_1},\ldots,X_{k_Q})$ are small if for some index l the lag $k_{l+1}-k_l$ is large. Here $k_1\le\cdots\le k_Q$ and a weak dependence condition will be is assumed.
This is also a natural extension of an important property of cumulants. A cumulant vanishes is it is derived from a couple of independent vectors.

Definition 12.2.1. *Let $t=(t_1,\ldots,t_p)$ be any $p-$tuple in $\mathbb{Z}^p$ with $t_1\le\cdots\le t_p$. Set $r(t)=\max_{1\le l<p}(t_{l+1}-t_l)$, the maximal lag in the*

sequence $(t_1, \ldots, t_p)$.
Define the other alternative dependence coefficient:

$$\kappa_p(r) = \max_{\substack{t_1 \le \cdots \le t_p \\ r(t_1, \ldots, t_p) \ge r}} \left|\kappa_p\left(X_{t_1}, \ldots, X_{t_p}\right)\right|. \tag{12.8}$$

Lemma 12.2.1. *If* $(X_n)_{n\in\mathbb{Z}}$ *is a centered and stationary process with finite moments up to order* Q.
If $Q \ge 2$ *by using notation in Lemma 12.1.1 we derive*

$$\kappa_{X,Q}(r) \le c_{X,Q}(r) + \sum_{s=2}^{Q-2} M_{Q-s}\left[\frac{Q}{2}\right]^{Q-s+1} \kappa_{X,s}(r).$$

Proof of Lemma 12.2.1. Set $X_\eta = \prod_{i\in\eta} X_i$ if $\eta \in \mathbb{Z}^p$ (η may include repetitions). If $k_1 \le \cdots \le k_Q$ is such that $k_{l+1} - k_l = r = \max_{1\le s<p}(k_{s+1} - k_s) \ge 0$.
Assume that $\mu = \{\mu_1, \ldots, \mu_u\}$ runs over all partitions of $\{1, \ldots, Q\}$ then one of those μ_i (denoted by ν_μ) satisfies

$$\nu_\mu^- = [1, l] \cap \nu_\mu \ne \emptyset \quad \text{and} \quad \nu_\mu^+ = [l+1, Q] \cap \nu_\mu \ne \emptyset.$$

From formula (12.2) we obtain with $\eta = \{1, \ldots, l\}$,

$$\kappa(X_{k_1}, \ldots, X_{k_Q}) = \mathrm{Cov}\left(X_{\eta(k)}, X_{\overline{\eta}(k)}\right) - \sum_u \sum_{\{\mu\}} \kappa_{\nu_\mu(k)} K_{\mu,k}, \tag{12.9}$$

with $K_{\mu,k} = \prod_{\mu_i \ne \nu_u} \kappa_{\mu_i(k)}$ where he previous sum extends to all partitions $\mu = \{\mu_1, \ldots, \mu_u\}$ of $\{1, \ldots, Q\}$ such that $\mu_i \cap \nu \ne \emptyset$ for some $i \in [1, u]$ and $\mu_i \cap \overline{\nu} \ne \emptyset$.
From $r\big(\nu_\mu(k)\big) \ge r(k)$ it is easy to derive $|\kappa_{\nu_\mu(k)}| \le \kappa_{X,\#\nu_\mu}(r)$.
This allow to let the size of lags increase.
With Lemma 12.1.1 we arrive to $|M_\mu| \le M_{Q-\#\mu_\nu}$ as in (12.5) and

the following bound is proved:

$$\begin{aligned}
\left|\kappa\left(X_{k_1},\ldots,X_{k_Q}\right)\right| &\le C_{X,Q}(r)+\sum_{u=2}^{[Q/2]}(u-1)!\sum_{\mu_1,\ldots,\mu_u} M_{Q-\#\nu_\mu}|\kappa_{\nu_\mu(k)}(X)| \\
&\le C_{X,Q}(r)+\sum_{u=2}^{[Q/2]}(u-1)!\sum_{s=2}^{Q-2} M_{Q-s}\kappa_{X,s}(r)\sum_{\substack{\mu_1,\,\ldots,\,\mu_u\\ \#\nu_\mu\,=\,s}} 1 \\
&\le C_{X,Q}(r)+\sum_{u=2}^{[Q/2]}(u-1)!\sum_{s=2}^{Q-2}(u-1)^{Q-s}M_{Q-s}\kappa_{X,s}(r) \\
&\le C_{X,Q}(r)+\sum_{s=2}^{Q-2}\frac{1}{Q-s+1}\left[\frac{Q}{2}\right]^{Q-s+1}M_{Q-s}\kappa_{X,s}(r)
\end{aligned}$$

Inequality

$$\sum_{u=1}^{U}(u-1)^p \le \frac{1}{p+1}U^{p+1}$$

follows from comparison of a series with an integral. ∎

Write Lemma 12.2.1 as

$$\kappa_{X,Q}(r) \le c_{X,Q}(r)+\sum_{s=2}^{Q-2} B_{Q,s}\kappa_{X,s}(r),$$

then

$$\begin{aligned}
\kappa_{X,2}(r) &\le c_{X,2}(r), \\
\kappa_{X,3}(r) &\le c_{X,3}(r), \\
\kappa_{X,4}(r) &\le c_{X,4}(r)+B_{4,2}\kappa_{X,2}(r) \\
&\le c_{X,4}(r)+B_{4,2}c_{X,2}(r), \\
\kappa_{X,5}(r) &\le c_{X,5}(r)+B_{5,3}\kappa_{X,3}(r)+B_{5,2}\kappa_{X,2}(r) \\
&\le c_{X,5}(r)+B_{5,3}c_{X,3}(r)+B_{5,2}c_{X,2}(r),
\end{aligned}$$

$$\begin{aligned}\kappa_{X,6}(r) &\leq c_{X,6}(r)+B_{6,4}\kappa_{X,4}(r)+B_{6,3}\kappa_{X,3}(r)+B_{6,2}\kappa_{X,2}(r)\\ &\leq c_{X,6}(r)+B_{6,4}\left(c_{X,4}(r)+B_{4,2}c_{X,2}(r)\right)\\ &\quad +B_{6,3}c_{X,3}(r)+B_{6,2}c_{X,2}(r)\\ &\leq c_{X,6}(r)+B_{6,4}c_{X,4}(r)+B_{6,3}c_{X,3}(r)\\ &\quad +(B_{6,2}+B_{6,4}B_{4,2})c_{X,2}(r).\end{aligned}$$

Lemma 12.2.1 implies the following important Corollary derived from a recursion with the previous inequalities.

Corollary 12.2.1. *For each $Q\geq 2$ there exists a constant $A_Q\geq 0$ only depending on Q and such that*

$$\kappa_{X,Q}(r)\leq A_Q\cdot c^*_{X,Q}(r).$$

Remark 12.2.1.

- *This Lemma proves the equivalence between coefficients $c_{X,Q}(r)$ and $\kappa_Q(r)$.*
  Precise upper bounds follow from Theorem 12.1.2.
  Decompose for it the sums corresponding to centered moments in 2 terms among which one explicitly depends on the maximal lag.
  Formula (12.9) implies with $B_{Q,Q}=1$,

$$c_{X,Q}(r)\leq\sum_{s=2}^{Q}B_{Q,s}\,\kappa_{X,s}(r).$$

  Thus there exists a constant $\widetilde{A}_Q$ with

$$c_{X,Q}(r)\leq\widetilde{A}_Q\kappa^*_{X,Q}(r),\quad \kappa^*_{X,Q}(r)=\max_{2\leq l\leq Q}\kappa^*_{X,l}(r)\mu_{Q-l}.$$

  Hence constants $a_Q,A_Q>0$ satisfy

$$a_Qc^*_{X,Q}(r)\leq\kappa^*_{X,Q}(r)\leq A_Qc^*_{X,Q}(r).$$

  Those coefficients are equivalent up to constants only depending on Q.

- *The previous formula (12.9) implies that a cumulant*

$$\kappa(X_{k_1},\ldots,X_{k_Q}) = \sum_{\alpha,\beta} K_{\alpha,\beta,k}\, Cov\,(X_{\alpha(k)}, X_{\beta(k)})$$

  is a linear combination of such covariances with $\alpha \subset \{1,\ldots,l\}$ and $\beta \subset \{l+1,\ldots,Q\}$ for which coefficients $K_{\alpha,\beta,k}$ are polynomials dof cumulants.
  For this replace the $Q-$tuple $(X_{k_1},\ldots,X_{k_Q})$ by $(X_i)_{i\in\nu_\mu(k)}$ for each partition μ in formula (12.9) and use recursion.

  This representation is useful if one knows the covariances.

  The advantage of cumulants over covariances of products is that given a vector $(X_{k_1},\ldots,X_{k_q})$ the behavior of the cumulant is that of $c_{X,q}(r(k))$. It does not need to know where occurs the maximal lag in indices.

Example 12.2.2. *Constants A_Q are not explicit but more tight bounds are derived from the previous proof for small values of Q*

$$\begin{aligned}
\kappa_{X,2}(r) &= c_{X,2}(r)\\
\kappa_{X,3}(r) &= c_{X,3}(r)\\
\kappa_{X,4}(r) &\le c_{X,4}(r) + 3\mu_2 c_{X,2}(r)\\
\kappa_{X,5}(r) &\le c_{X,5}(r) + 10\mu_2 c_{X,3}(r) + 10\mu_3 c_{X,2}(r)\\
\kappa_{X,6}(r) &\le c_{X,6}(r) + 15\mu_2 c_{X,4}(r) + 20\mu_3 c_{X,3}(r)) + 150\mu_4 c_{X,2}(r)
\end{aligned}$$

However the heavy combinatorics gives an advantage to the rough bound in Lemma 12.2.1 to bound high order cumulants.

12.2.2 Sums of cumulants

The previous bounds yield

Lemma 12.2.2. *Let*

$$\kappa_Q = \sum_{k_2=0}^{\infty}\cdots\sum_{k_Q=0}^{\infty}\left|\kappa\left(X_0, X_{k_2},\ldots,X_{k_Q}\right)\right|, \tag{12.10}$$

with notation (12.7) for each $Q \geq 2$ there exists a constant B_Q such that

$$\kappa_Q \leq B_Q \sum_{r=0}^{\infty} (r+1)^{Q-2} C^*_{X,Q}(r).$$

Proof of Lemma 12.2.2. Decompose sums:

$$\kappa_Q \leq (Q-1)! \sum_{k_2 \leq \cdots \leq k_Q} \left| \kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right) \right| \equiv (Q-1)!\, \widetilde{\kappa}_Q$$

considering the following partition of the index set

$$E = \{k = (k_2, \ldots, k_Q) \in \mathbb{N}^{Q-1} / k_2 \leq \cdots \leq k_Q\}$$

as $E_r = \{k \in E / r(k) = r\}$ for $r \geq 0$ (according to the size of the maximal lag) then

$$\widetilde{\kappa}_Q = \sum_{r=0}^{\infty} \sum_{k \in E_r} \left| \kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right) \right|.$$

The previous Lemma implies

$$\sum_{k \in E_r} \left| \kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right) \right| \leq A_Q \# E_r C^*_{X,Q}(r),$$

for a constant $A_Q > 0$ and the elementary bound

$$\# E_r \leq (Q-1)(r+1)^{Q-2},$$

yields the result. ∎

12.2.3 Moments of sums

Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary and centered sequence one expects an asymptotic behavior analogue to the CLT for partial sums

$$\frac{1}{\sqrt{n}} (X_1 + \cdots + X_n) \longrightarrow_{n \to \infty} \mathcal{N}(0, \sigma^2), \qquad \text{in law.}$$

The behavior of moments in $\mathbb{L}^p$−norm is of importance.

Cumulants allow an elementary approach of such expressions.

Lemma 12.2.3. *If the series (12.10) are summable for each $Q \le p$ set $q = [p/2]$ ([1]) then*

$$\left|\mathbb{E}\left(\sum_{j=1}^{n} X_j\right)^p\right| \le \sum_{u=1}^{q} n^u \gamma_u, \qquad \text{with} \tag{12.11}$$

$$\gamma_u = \sum_{v=1}^{2q} \sum_{p_1+\cdots+p_u=p} \frac{p!}{p_1!\cdots p_u!}\kappa_{p_1}\cdots\kappa_{p_u}.$$

Proof. As in [Doukhan and Louhichi, 1999] bound

$$\begin{aligned} |\mathbb{E}(X_1+\cdots+X_n)^p| &= \left|\sum_{1\le k_1,\ldots,k_p\le n} \mathbb{E}X_{k_1}\cdots X_{k_p}\right| \\ &\le p! A_{p,n} \equiv p! \sum_{1\le k_1,\ldots,k_p\le n} \left|\mathbb{E}X_{k_1}\cdots X_{k_p}\right| \end{aligned}$$

Let also $\mu = \{i_1,\ldots,i_v\} \subset \{1,\ldots,p\}$ and $k = (k_1,\ldots,k_p)$ set

$$\mu(k) = (k_{i_1},\ldots,k_{i_v}) \in \mathbb{N}^v \tag{12.12}$$

To enumerate the terms with their multiplicity this is simpler to consider multi-indices than partitions.
Cumulants and moments are defined analogously.
As in [Doukhan and León, 1989] with formula (12.2) and partitions

[1] This expression equals $q = p/2$ for p even and $q = (p-1)/2$ else.

$\mu_1, \ldots, \mu_u$ of $\{1, \ldots, p\}$ with exactly $1 \le u \le p$ elements,

$$
\begin{aligned}
A_{p,n} &= \sum_{1\le k_1,\ldots,k_p\le n} \sum_{u=1}^{p} \sum_{\mu_1,\ldots,\mu_u} \prod_{j=1}^{u} \kappa_{\mu_j(k)}(X) \\
&= \sum_{u=1}^{p} \sum_{\mu_1,\ldots,\mu_u} \sum_{1\le k_1,\ldots,k_p\le n} \prod_{j=1}^{u} \kappa_{\mu_j(k)}(X) \\
&= \sum_{r=1}^{p} \sum_{p_1+\cdots+p_r=p} \frac{p!}{p_1!\cdots p_r!} \times \qquad (12.13) \\
&\quad \times \prod_{u=1}^{r} \sum_{1\le k_1,\ldots,k_{p_u}\le n} \kappa_{p_u}(X_{k_1}, \ldots, X_{k_{p_u}}) \\
|A_{p,n}| &\le \sum_{u=1}^{q} n^u \sum_{p_1+\cdots+p_u=p} \frac{p!}{p_1!\cdots p_u!} \prod_{j=1}^{u} \kappa_{p_j} \qquad (12.14)
\end{aligned}
$$

Identity (12.13) follows from a change of variable and takes into account the fact that the number of partitions for $\{1, \ldots, p\}$ into u sets with respective cardinalities $p_1, \ldots, p_u$ is the multinomial coefficient. For $\lambda \in \mathbb{N}$ one may deduce from the stationarity of X that

$$\sum_{1\le k_1,\ldots,k_\lambda\le n} |\kappa_{p_u}(X_{k_1}, \ldots, X_{k_\lambda})| \le n\kappa_\lambda.$$

Cumulants with order 1 always vanish and non zero terms are such that if there exist u indices $p_j \ge 2$ ($p_1, \ldots, p_u \ge 2$ thus $2u \le p$) then $u \le q$.
We thus get (12.14). ∎

Remark 12.2.2. *If their exists $C > 0$ with $\kappa_s \le C^s$ for each $s \le p$ then due to the multinomial identity the bound (12.14) simply writes*

$$C^p \sum_{u=1}^{q} u^p n^u.$$

12.2.4 Rosenthal inequality

Again as in [Doukhan and Louhichi, 1999] we derive a Rosenthal type inequality using coefficients $c_{X,l}(r)$.
As in the previous proof

$$|\mathbb{E}(X_1 + \cdots + X_n)^p| \le p! A_{p,n} \equiv p! \sum_{1\le k_1,\ldots,k_p\le n} \left|\mathbb{E}X_{k_1}\cdots X_{k_p}\right|.$$

Each term T_k $(k = (k_1, \ldots, k_p))$ in the sum $A_{p,n}$ admits a maximal lag $r = r(k) = \max_j(k_{j+1} - k_j) < n$,

$$T_k \le c_{X,p}(r) + |\mathbb{E}X_{k_1}\cdots X_{k_l}| \cdot \left|\mathbb{E}X_{k_{l+1}}\cdots X_{k_p}\right|.$$

Partition those multi-indices k according to the value of $r(k)$ and the smallest index $l = l(k)$ such that $r(k) = k_{l+1} - k_l$ (for r and l fixed there exists less that $n(r+1)^{p-2}$ such multi-indices).
We obtain

$$A_{p,n} \le (p-1)n \sum_{r=0}^{n-1} (r+1)^{p-2} c_{X,p}(r) + \sum_{l=2}^{p-2} A_{l,n} A_{p-l,n}.$$

A Rosenthal is thus proved in [Doukhan and Louhichi, 1999].
Iterating the previous relation yields

$$\begin{array}{rcl}
A_{2,n} & \le & nC_{2,n} \\
A_{3,n} & \le & 2nC_{3,n} \\
A_{4,n} & \le & 3nC_{4,n} + A_{2,n}^2 \le 3nC_{4,n} + n^2 C_{2,n}^2 \\
A_{5,n} & \le & 4nC_{5,n} + 2A_{2,n}A_{3,n} \le 5nC_{5,n} + 4n^2 C_{2,n}C_{3,n} \\
A_{6,n} & \le & 5nC_{6,n} + 2A_{2,n}A_{4,n} + A_{3,n}^2 \le 5nC_{6,n} \\
 & & +2n^2\left(2C_{3,n}^2 + 3C_{2,n}C_{4,n}\right) + 8n^3 C_{2,n}^3
\end{array}$$

We denote (for a fixed q)

$$C_{m,n}^{(q)} = \sum_{k=0}^{n-1} (r+1)^{m-2} c_{X,q}(r).$$

Generally if $p = 2q$ or $p = 2q+1$ we obtain

$$A_{p,n} \le \sum_{j=1}^{q} c_{j,n} n^j,$$

with $c_{j,n}$ a polynomial wrt expressions $C_{i,n}^{(q)}$ for $i \le j$. Precisely this is a linear combination of expressions $\prod_{s=1}^{t} C_{i_s,n}^{(q_s)}$ with $i_1+\cdots+i_t = j$ and $q_1 + \cdots + q_t = p$.
Hence for $c_{X,p}(r) = \mathcal{O}(r^{-q})$ one deduces Marcinkiewicz-Zygmund inequality

$$\left|\mathbb{E}(X_1 + \cdots + X_n)^p\right| = \mathcal{O}(n^q).$$

Rosenthal type inequalities yield sharp bounds for centered moments of kernel density estimators or for the empirical process.

12.3 Dependent kernel density estimation

Better that atomizing the different items related to kernel density estimation under dependence extending on section 3.3, in various parts of those notes we set them all as a specific section. Assume that the marginals of the stationary process (X_n) admit a density f. Let the kernel K is symmetric compactly supported and Lipschitz and a window sequence $h_n \downarrow 0$ with $nh_n \to \infty$ and $x \in \mathbb{R}$.
Set $U = (U_j)_{j\in\mathbb{Z}}$ with

$$U_j = K\left(\frac{X_j - x}{h_n}\right) - \mathbb{E}K\left(\frac{X_j - x}{h_n}\right).$$

Use eg. $\theta-$weakly dependence.
This is easy to prove that:

$$\begin{aligned} \operatorname{Lip} h \quad &\le \quad l2^{p-1}\cdot\frac{\operatorname{Lip} K}{h_n}, \qquad \text{if we denote:} \\ h(t_1,\ldots,t_l) \quad &= \quad \prod_{j=1}^{l}\left\{K\left(\frac{t_j - x}{h_n}\right) - \mathbb{E}K\left(\frac{X_j - x}{h_n}\right)\right\} \end{aligned}$$

If for each $n > 0$ the joint density $f_n(x,y)$ of the couple (X_0, X_n) exists and

$$\|f_n(\cdot,\cdot)\|_\infty \le M. \tag{12.15}$$

Exercise 25. *Assume that (X_n) is a stationary real valued Markov chain with a transition kernel $P(x,A) = \mathbb{P}(X_1 \in A|X_0 = x)$, with a* *density; this means that one may write*

$$\mathbb{P}(X_1 \in A | X_0 = x) = \int_A p(x,y)\,dy,$$

for a measurable function p.
Condition (12.15) holds if $\|f\|_\infty < \infty$ and the transition probabilities admit a transition density such that $\|p\|_\infty < \infty$.
Assume e.g. that $X_n = r(X_{n-1}) + \xi_n$ with Lip $r < 1$, $\mathbb{E}|\xi_0| < \infty$, and ξ_0 admits a bounded density g wrt Lebesgue measure then use the previous results to derive those bounds ([2]).

Integrating of the relation (12.15) yields $\|f(\cdot)\|_\infty \le M$ and

$$c_{U,p}(0) \le 2^p f(x) \int K^2(s)ds.$$

A direct calculation coupled with a weak dependence inequality yield 2 distinct controls of $c_{U,p}(r)$ for $r > 0$

$$c_{U,p}(r) \le 2^{p-1}\left(p \cdot \mathrm{Lip}\, K \frac{\theta_r}{h_n}\right) \wedge (2Mh_n^2),$$

thus there exists a constant $C > 0$ with

$$C_{p,n} = Ch_n\left(1 + \sum_{k=1}^{n-1}(r+1)^{p-2}\left(h_n \wedge \frac{\theta_r}{h_n^2}\right)\right).$$

The following elementary inequality is often useful when two different bounds of a quantity are available. In our case, two inequality may appear either from dependence properties or from simple analytic tricks.

Exercise 26.

$$u \wedge v \le u^\alpha v^{1-\alpha}, \qquad \textit{if } u, v \ge 0, \quad 0 \le \alpha \le 1. \tag{12.16}$$

[2] Check that Proposition 7.3.1 implies the existence of a stationary distribution and the relation $f(x) = \int_{\mathbb{R}} p(x,y)f(y)\,dy$ implies with $p(x,y) = g(y - r(x))$ that $M = \|g\|_\infty$.

Hint. From the symmetry of u, v's roles assume that $u \leq v$ then $u \leq 1$ implies $u \wedge v = u = u^\alpha \cdot u^{1-\alpha} \leq u^\alpha \cdot v^{1-\alpha}$. ∎

As a simple application of the previous inequalities, for $p = 2$ we derive the following result

Proposition 12.3.1. *Assume that $\theta_r \leq Cr^{-a}$ for some $a > 3$, then*

$$\lim_{n\to\infty} nh_n \, Var \, \widehat{f}(x) = f(x) \int K^2(t)\, dt.$$

Proof. First quote that $c_{U,2}(0) \sim h_n f(x) \int K^2(s) ds$ and thus one simply need to derive that

$$\lim_{n\to\infty} \frac{1}{h_n} \sum_{r=1}^{\infty} c_{U,2}(r) = 0$$

but for some constant and from relation (12.16) one derives

$$\frac{1}{h_n} c_{U,2}(r) \leq C \frac{\theta_r}{h_n^2} \wedge h_n \leq h_n^{1-3\alpha} \theta_r^\alpha$$

The assumption $a > 3$ implies that there exists some $\alpha < \frac{1}{3}$ such that

$$\sum_{r=1}^{\infty} \theta_r^\alpha < \infty.$$

Hence the dependent part of those variances is indeed negligible and the behavior of kernel density estimates is the same as under independence. ∎

Exercise 27. *Using inequality $\theta/h^2 \wedge h \leq \theta^{1/3}$, prove that $C_{p,n} = \mathcal{O}(h_n)$ if*

$$\sum_{r=0}^{\infty} (r+1)^{p-2} \theta_r^{1/3} < \infty \tag{12.17}$$

More generally if $p \geq 2$, from recursion and by using assumption (12.17) and Exercise 27 we get

$$\left| \mathbb{E}(\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x))^p \right| \leq C(nh_n)^{p-q}.$$

This bound has order $(nh_n)^{-p/2}$ for even p and $(nh_n)^{-(p-1)/2}$ if p is odd.
Consider now some even integer $p > 2$. Almost sure convergence of such estimates also follows from Markov inequality and Borel-Cantelli lemma in case:

$$\sum_{n=1}^{\infty} \frac{1}{(nh_n)^{p/2}} < \infty.$$

Exercise 28. *Derive a.s. uniformly for $x \in [a, b]$ over a compact interval.*

Hint. Use Remark 3.3.5. ∎

Those bounds fit with the underlying CLT:

Theorem 12.3.1 ([Bardet et al., 2006])**.** *Assume the assumption in Proposition 12.3.1 hold then:*

$$\sqrt{nh_n}(\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)) \to_{n\to\infty} \mathcal{N}\left(0, f(x)\int K^2(t)dt\right).$$

Proof. Use Lemma 11.5.1 then arguing as in Proposition 12.3.1 allows a tight control of the dependent terms again.
We leave the result as an exercise.
Let

$$x_\ell = z_\ell - \mathbb{E}z_\ell, \qquad \text{with} \qquad z_\ell = \frac{1}{\sqrt{nh_n}} K\left(\frac{X_\ell - x}{h_n}\right),$$

$s_1 = 0$ and $s_\ell = x_1 + \cdots + x_\ell$, then simply quote that the needed controls of the dependence terms are obtained through the relation

$$\mathrm{Cov}\,(e^{its_k}, e^{itx_k}) = \sum_{0\le \ell\le k} \mathrm{Cov}\,(e^{its_\ell} - e^{its_{\ell-1}}, e^{itx_k}),$$

which concludes. ∎

Exercise 29. *Extend this whole section to the case of associated processes. In particular precise how eqn. (12.17) should be modified in this case.*

Exercise 30. *Extend those results to the case of regression estimates (3.5), both under weakly dependent or under associated frameworks.*

Remark 12.3.1. *Standard extensions are possible for the other weak dependence conditions as well as under strong mixing.*

We leave all such extensions as exercises. Moreover the case of subsampling is analogue:

Exercise 31. *Moreover subsampling from § 4.6 may also be considered as in Exercise 24 and higher order moments may be accurately bounded (as in [Doukhan et al., 2011]) under weak dependence conditions in order to derive almost sure convergence from Borel-Cantelli lemma in case some even number $p \in 2\mathbb{N}$ satisfies*

$$\sum_{n=1}^{\infty} \mathbb{E}\big(\widehat{K}_n(g) - \mathbb{E}\widehat{K}_n(g)\big)^p < \infty.$$

Hints. The proof is as for kernel density estimation based on bounds for coefficients $c_{Z,r}(r)$ in eqn. (12.7) with $Z_i = g(t_m(X_{i+1}, \ldots, X_{i+m}))$ in the overlapping scheme.
For $h(x) = \mathbb{1}_{\{x \le z\}}$ analogously to Exercise 22, bounds of

$$c_{Y,r}(r) = \sup_{\mathbf{i},j} |\mathrm{Cov}\,(h(X_{i_1}) \cdots h(X_{i_u}), h(X_{j_1}) \cdots h(X_{j_v}))|$$

$u + v = p$, $i_1 \le \cdots \le i_u$, $j_1 \le \cdots \le j_v$ with $j_1 - i_u = r$ allow to bound higher order moments.
Set $I_1 = h(X_{i_u})$, $I_1^\epsilon = h_\epsilon(X_{i_u})$, $J_1 = h(X_{j_u}), \ldots$ since those functions are bounded and $\mathrm{Lip}\, h_\epsilon = 1/\epsilon$, using the following inequalities yields eg. under $\eta-$weak dependence:

$$\begin{aligned}
|\mathrm{Cov}\,(I_1 \cdots I_u, J_1 \cdots J_v))| &\le |\mathrm{Cov}\,(I_1^\epsilon \cdots I_u^\epsilon, J_1^\epsilon \cdots J_v^\epsilon))| \\
&\quad + 2\sum_{s=1}^{u} \mathbb{E}|I_s - I_s^\epsilon| + 2\sum_{t=1}^{u} \mathbb{E}|J_t - J_t^\epsilon| \\
&\le \frac{u+v}{\epsilon}\eta_r + (u+v)\epsilon = 2(u+v)\sqrt{\eta_r}, \ (\text{with } \epsilon^2 = \eta_r).
\end{aligned}$$

The same job does not need this last step under strong mixing. ∎

Appendix A

A.1 Probability

A.1.1 Notations

For a space E, recall that a sigma-algebra $\mathcal{E}$ is a subset of $\mathcal{P}(E)$, set of the subsets of E), such that

- $\emptyset \in \mathcal{E}$,
- $\forall A \in \mathcal{E}: \ A^c \in \mathcal{E}$, where we denote by $A^c = E \backslash A$ the complementary set of A.
- $\forall A_n \in \mathcal{E}, n = 1, 2, 3 \ldots :$

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}.$$

A measured space is such a couple $(E, \mathcal{E})$.
A probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a measured space $(\Omega, \mathcal{A})$ equipped with a probability, that is a function $\mathbb{P} : \mathcal{A} \to [0, 1]$ such that:

- $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
- $\forall A, B \in \mathcal{A} :$

$$A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

- $\forall A_i \in \mathcal{A}, i = 1, 2, 3 \ldots :$

$$A_1 \subset A_2 \subset \cdots \Rightarrow \lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A), \quad A = \bigcup_{n=1}^{\infty} A_n.$$

Elements of $\mathcal{A}$ are called events.

Example A.1.1. *Examples of measurable spaces $(\Omega, \mathcal{A})$ follow. One should remember that one usually needs two measurable spaces: a space of values or realizations $(E, \mathcal{A})$ and an abstract probability space $(\Omega, \mathcal{A})$ which needs a probability function $\mathbb{P}$.*
We list some such simple spaces.

- *If Ω is a finite set with n elements then a reasonable choice of sigma-algebra is $\mathcal{A} = \mathcal{P}(\Omega)$ which admits 2^n elements as it may be seen from the fact the the application: $A \mapsto 1\!\!I_A$ defined for $\mathcal{P}(E)$ on the set of functions from E to $\{0,1\}$ is a bijective function.*
- *For denumerable finite sets Ω again $\mathcal{A} = \mathcal{P}(\Omega)$ is a suitable frame.*
- *$\mathbb{R}$ may be equipped with its Borel $\sigma-$field, the smallest containing intervals.*
- *More generally a topological space Ω is measurable with $\mathcal{A}$ the smallest $\sigma-$field containing all the open sets. This $\sigma-$field is called the Borel $\sigma-$field.*
- *Products of 2 measurable spaces are still measurable, and here the $\sigma-$field is again the smallest containing products $A \times B$ with clear notations.*
- *Infinite products are again possible; for a family of measurable spaces $(\Omega_i, \mathcal{A}_i)_{i\in I}$ the product $\Omega = \prod_{i\in I} \Omega_i$ and $\mathcal{A}$ is equipped with the smallest $\sigma-$field containing all the events $\prod_{i\in I} A_i$ with $A_i \in \mathcal{A}_i$ for each $i \in I$ and $A_i = \Omega_i$ for each $i \notin J$ with $J \subset I$, a finite subset of I.*
- *Examples of probability spaces are various and some of them are linked with the generation of random variables. They will be considered in Example A.1.3.*

The sigma-algebra $\mathcal{A}$ is complete in case $A \in \mathcal{A}$, $\mathbb{P}(A) = 0$ and $B \subset A$ imply $B \in \mathcal{A}$ (roughly speaking, it contains the nullsets).
The $\sigma-$field considered are usually those obtained from a measurable space equipped with some measure (often, probability measures): the completed $\sigma-$field is the smallest containing both all the events $A \in \mathcal{A}$ and each set $B \subset A$ for each $A \in \mathcal{A}$ with $\mathbb{P}(A) = 0$.

A.1.2 Random variables

Let $X : \Omega \to E$ be an arbitrary function defined on the measurable space $(\Omega, \mathcal{A})$, taking values in another measurable space $(\mathbb{E}, \mathcal{E})$.

We introduce the probabilist notation:

$$(X \in A) = X^{-1}(A) = \{\omega \in \Omega;\ X(\omega) \in A\}, \qquad \text{for each A } \subset E.$$

A random variable $X : \Omega \to E$ is a measurable function between those two measurable sets: this means that $X^{-1}(\mathcal{E}) \subset \mathcal{A}$. In other terms:

$$\forall A \in \mathcal{E} : \qquad (X \in A) \in \mathcal{A}.$$

Note also that $\sigma(X) = X^{-1}(\mathcal{E})$ is the $\sigma-$algebra generated by the random variable $X : \Omega \to E$, that mean it is the smallest sub-σ-algebra $\mathcal{F}$ of $\mathcal{A}$ which makes the application $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ measurable.

Also the image distribution or the law of X is the probability distribution defined as

$$P_X(A) = \mathbb{P}(X \in A), \qquad \forall A \in \mathcal{E}.$$

In most of the cases $E = \mathbb{R}$ will be endowed with its Borel sigma-algebra ([1]), completed when necessary. X's distribution probability is also defined through its cumulative distribution function:

$$F(x) = \mathbb{P}(X \leq x) \equiv P_X((-\infty, x]), \qquad \forall x \in \mathbb{R}.$$

In some cases $E = \mathbb{R}^d$ is a finite dimensional vector space but we shall avoid to make the situations more complicated. For a column vector $v \in \mathbb{R}^d$, set v' the corresponding row vector ([2]).

Definition A.1.1. *For $E = \mathbb{R}^d$ one defines the mean of a random variable $X \in \mathbb{R}^d$:*

$$\mathbb{E}X = \int_E x P_X(dx) \in \mathbb{R}^d,$$

in case those integrals converge ([3]).
And moreover the covariance:

$$Cov(X) = \mathbb{E}XX' - \mathbb{E}X(\mathbb{E}X)' \quad \textit{is a symmetric positive } n \times n\textit{-matrix,}$$

In case $X = (X_1, X_2)$, we also write:

$$Cov(X) = \begin{pmatrix} Var\,X_1 & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var\,X_2 \end{pmatrix}$$

[1] Let E be any topological space, its Borel sigma algebra $\mathcal{E}$ is the smallest sigma-algebra containing all the open sets; it contains thus also intersections of open sets but also much more complicated sets.

[2] Anyway we assume that students will rectify by themselves the numerous errors of this type in those notes.

[3] or if they can be defined, as this is is the case for $d = 1$ and $E = \mathbb{R}^+ \equiv [0, +\infty)$. In this case integrals take values in the space $[0, +\infty]$.

Note that if $d = 1$, then

$$X \geq 0 \implies \mathbb{E}X \geq 0. \tag{A.1}$$

A first but essential result is the following

Theorem A.1.1 (Markov inequality)**.** *Assume that $V \geq 0$ is a real valued non-negative random variable, then its expectation exists in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ and:*

$$\mathbb{P}(V \geq u) \leq \frac{\mathbb{E}V}{u}, \qquad \forall u > 0.$$

Proof. Set $A = (V \geq u)$ then using (A.1) we derive:

$$\mathbb{E}V = \mathbb{E}V1_A + \mathbb{E}V1_{A^c} \geq \mathbb{E}V1_A \geq u\mathbb{P}(A) \qquad \blacksquare$$

Proposition A.1.1 (Jensen inequality)**.** *Jensen inequality holds for each function $g : C \to \mathbb{R}$ convex and continuous on the convex set $C \subset \mathbb{R}^d$.*
If $Z \in C$ a.s. (and if the following expectations are well defined)

$$\mathbb{E}g(Z) \geq g(\mathbb{E}Z) \tag{A.2}$$

Proof. We begin with the case $d = 1$. In this case we assume that $C = (a, b)$ is an interval then $g : (a, b) \to \mathbb{R}$ is derivable excepted possibly on some denumerable set.
Moreover on each point of C the left and right derivatives exist (at the extremities, only one of them may be defined; moreover for any $x, y, z \in C$ if $x < y < z$ then

$$g'(x+) \leq g'(y-) \leq g'(y+) \leq g'(z-),$$

with

$$g'(y\pm) = \lim_{h \to 0^+} \frac{g(y \pm h) - g(y)}{\pm h},$$

then for each $x_0 \in C$ choose any $u \in [g'(x_0-), g'(x_0+)]$, then the affine function

$$f(x) = u(x - x_0) + g(x_0)$$

satisfies $f \leq g$ and $f(x_0) = g(x_0)$ by convexity.
Thus each convex function g is the upper bound of affine functions $f \leq g$.
From linearity of integrals $f(\mathbb{E}Z) = \mathbb{E}f(Z)$ and thus $f(\mathbb{E}Z) \leq \mathbb{E}g(Z)$.
Now the relation $\sup_f f(\mathbb{E}Z) = g(\mathbb{E}Z)$ allows to conclude.
If now $d \geq 1$ then from the most elementary variant of Hahn-Banach

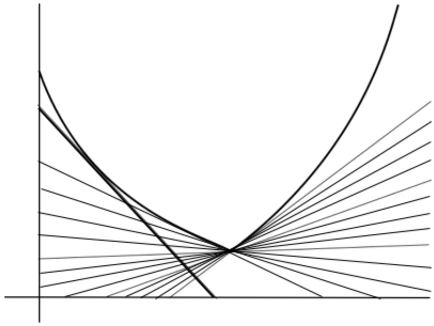


Figure A.1: Convex functions are suprema of affine functions.

theorem ([4]) the same representation of g holds and the proof is the same see Figure A.1. ∎

Remark A.1.1.

- *This inequality is an equality for each affine function.*
- *The inequality is strict if g is strictly convex and Z is not a.s. constant. The case of power functions is investigated in lemma 7.3.1.*
- *Let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ algebra of $\mathcal{A}$, a conditional variant of this inequality writes([5]) :*

$$\mathbb{E}^{\mathcal{B}} g(Z) \geq g\left(\mathbb{E}^{\mathcal{B}} Z\right). \tag{A.3}$$

The following standard inequality is important :

Proposition A.1.2 (Hölder inequality)**.** *Let $X_1 \in \mathbb{L}^{p_1}, \ldots, X_u \in \mathbb{L}^{p_u}$ be real valued random variables, then:*

$$\mathbb{E}|X_1 \cdots X_u| \leq \|X_1\|_{p_1} \cdots \|X_1\|_{p_u}, \quad \textit{if} \quad \frac{1}{p_1} + \cdots + \frac{1}{p_u} = 1.$$

[4] In the Hilbert case the orthogonal projection provides a elementary way to separate a point from a disjoint closed convex set: take its orthogonal projection y of x then the hyperplane with direction $x^{\perp}$ and containing the middle of the interval $[x, y]$ is a valuable solution of Hahn-Banach separation problem.

[5] use of a conditional version of the dominated convergence theorem.

Hint. For $z_1, \ldots, z_u > 0$ the convexity of the exponential function implies $z_1 \cdots z_u \le \frac{1}{p_1} z_1^{p_1} + \cdots + \frac{1}{p_u} z_u^{p_u}$. Now set $z_j = |X_j| / \|X_j\|_{p_j}$ to conclude. ∎

Definition A.1.2. *Let $X \in \mathbb{R}^d$ be a vector valued random variable then its characteristic function is defined as*

$$\phi_X(t) = \mathbb{E} e^{it \cdot X}, \qquad \forall t \in \mathbb{R}^d.$$

The Laplace transform of the law of X is :

$$L_X(z) = \mathbb{E} e^{z.X}, \qquad \text{for all } z \in Dom(L_X) \subset \mathbb{C}^d,$$

($Dom(L_X)$ if the set of such z such that this expression is well defined.)

Remark A.1.2. *First quote that the characteristic function always exists and $\phi_X(t) = L_X(it)$.*
If 0 is interior to the domain of definition of L_X then this function is analytic around 0 as well as ϕ_X.
Thus interchanging derivation and integrals is legitimate:

$$\frac{\partial}{\partial t_j} \phi(0) = i \cdot \mathbb{E} X_j.$$

Moreover Fourier integral theory implies that inversion if possible and thus in this case ϕ_X determines X's distribution.

Simple examples of probability distributions are

- *Discrete random variables*: there exists a finite or denumerable set S such that $\mathbb{P}(X \notin S) = 0$ and in case the following series is absolutely convergent we denote

$$\mathbb{E} X = \sum_{x \in S} x \mathbb{P}(X = x).$$

  In the case when $S \subset \mathbb{Z}$ the generating function $g_X(z) = \mathbb{E} z^X$ will be preferred to the Laplace transform and this function is also defined for $|z| \le 1$.
  Examples are

  – Bernoulli law $b(p)$ with parameter $p \in [0,1]$ is the law of a random variable with values in $\{0,1\}$ with
  $\mathbb{P}(X = 1) = p$ and thus $\mathbb{P}(X = 0) = 1 - p$.
  Here $g_X(z) = pz + q$.

– Binomial law $B(n,p)$ with parameters $n \in \mathbb{N}^*, p \in [0,1]$ is the law of a random variable with values in $\{0,1,\ldots,n\}$ with

$$\mathbb{P}(X=k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

Here, with $g_X(z) = (pz+q)^n$ and, if $X_1, \ldots, X_n \sim b(p)$ are independent identically distributed random variables then

$$X_1 + \cdots + X_n \sim B(n,p).$$

– A Poisson distributed random variable X with parameter λ takes values in $\mathbb{N}$ and $\mathbb{P}(X=k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.

- *Absolutely continuous distributions.*

Definition A.1.3. *We assume here that there exists a measurable function $f : E \to \mathbb{R}^+$ such that for each $A \in \mathcal{E}$:*

$$P_X(A) = \int_A f(x)\, dx$$

this function is called the density of X distribution.

Remark A.1.3.

We also derive that for each measurable function $g : E \to \mathbb{R}$,

$$\mathbb{E}g(X) = \int_E g(x) f(x)\, dx.$$

This relation is also the definition of a density.

Example A.1.2. *Some examples follow (see also Appendix § A.2):*

– *Uniform distribution on the unit interval, it admits a density wrt the Lebesgue measure $f(x) = \mathbb{1}_{[0,1]}(x)$*

– *Exponential law $\mathcal{E}(\lambda)$ with parameter λ admits the density $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$.*

– *The Normal law $\mathcal{N}(0,1)$ is the simplest Gaussian law which admits the density $f(x) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$.*

– *Cauchy distribution is defined with $f(x) = \dfrac{1}{\pi} \cdot \dfrac{1}{1+x^2}$. The mean of such Cauchy distributed random variables is thus not defined.*

Remark A.1.4 (Simulation)**.** *If a cdf F is one-to-one on its image then for each uniform random variable $U \sim U[0,1]$ the random variable $X = F^{-1}(U)$ admits the cumulative distribution function F.*
This is an easy way to simulate real random variables with marginal distribution.
Quote that the same holds for more general cases when defining:

$$F^{-1}(t) = \inf\{x \in \mathbb{R} |\ F(x) \geq t\}$$

Simple examples prove that other possibilities are available:

- *Assume that $X \sim b(p)$ then $F(t) = 1$ for $t \geq p$ then one simulates a $b(p)-$distributed random variable by setting $X' = 1\!\!1_{\{U \leq p\}}$. Anyway other possibilities are $1\!\!1_{\{U<p\}}$, $1\!\!1_{\{U \geq 1-p\}}$ and $1\!\!1_{\{U>1-p\}}$, since $1-U$ also admits a $U[0,1]-$distribution.*
- *For $\mathcal{E}(\lambda)-$distributions $F(t) = 1 - e^{-\lambda t}$ so that*

  $$F^{-1}(t) = -\ln(1-t)/\lambda;$$

  again simulations of such exponential random variables write $X = -\ln(1-U)/\lambda$, or more accurately $-\ln(U)/\lambda$.

Example A.1.3 (Probability spaces)**.** *An example of probability space is $\Omega = [0,1]^{\mathbb{Z}}$ endowed with its product σ-algebra.*
This is the smallest sigma-algebra containing cylinder events $\prod_{n\in\mathbb{Z}} A_n$ where A_n is a Borel set of $[0,1]$ and $A_n \neq [0,1]$ for only finitely many indices n.
Then a sequence of random variables X_n is defined as the $n-$th coordinate function $X_n(\omega) = \omega_n$ for all $\omega = (\omega_n)_{n\in\mathbb{Z}}$.
In this case each of the coordinates X_n admits the uniform distribution μ, the Lebesgue measure on $[0,1]$.
Let now F be the cumulative repartition of the law ν of a real valued random variable then setting instead $X_n(\omega) = F^{-1}(\omega_n)$ make that

$$\mathbb{P}(X_n \in A) = F(A) = \nu(A) = \mathbb{P}(X \in A).$$

Thus one may assign any distribution to those coordinates.

Lemma A.1.1 (Hoeffding)**.**

1. *Let $Z \geq 0$ be a (a.s.-)non-negative random variable then*

   $$\mathbb{E}Z = \int_0^\infty \mathbb{P}(Z \geq t)\, dt.$$

2. *Let $X, Y \in \mathbb{L}^2$ be two real valued random variables*

$$Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)\mathbb{P}(Y \geq t)\right) ds\, dt.$$

Proof.

1. Let $z \geq 0$ then $z = \int_0^\infty \mathbb{I}_{(z \geq t)}\, dt$. Set λ the Lebesgue measure on the line. Without any convergence assumption of those integral, Fubini-Tonnelli theorem applies to the non-negative function $(t, \omega) \mapsto \mathbb{I}_{(Z(\omega) \geq t)}$. This allows to conclude.
2. First for $X, Y \geq 0$ the same trick as above works and

$$\mathbb{E}XY = \int_0^\infty \int_0^\infty \mathbb{P}(X \geq s, Y \geq t)\, ds\, dt.$$

Write $X = X^+ - X^-, Y = Y^+ - Y^-$ for nonnegative random variables $X^\pm, Y^\pm$.

The formula holds for each of them and

$$\mathbb{P}(X \geq s) = \begin{cases} \mathbb{P}(X^+ \geq s), & \text{if} \quad s \geq 0 \\ 1 - \mathbb{P}(X^- > -t), & \text{if} \quad s < 0 \end{cases}$$

Now for an arbitrary couple of real valued random variables Cov (X, Y) writes as a linear combination of four such integrals with respective coefficients ± 1. ∎

A.2 Distributions

We provide a short introduction to Gaussians with first the standard normal and then it vector extension essential to define Gaussian processes and finally $\gamma-$laws which permit many explicit calculations.

A.2.1 Normal distribution

A standard normal random variable is a real valued random variable such that $N \sim \mathcal{N}(0,1)$ admits the density

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

wrt Lebesgue measure on $\mathbb{R}$. The norming factor $\sqrt{2\pi}$ may be checked through the computation of a square as follows:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy\\ &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\\ &= 2\pi.\end{aligned}$$

To this aim, use a change in variables with polar coordinates

$$(r,\theta)\mapsto(x,y)=(r\cos\theta, r\sin\theta), \qquad \mathbb{R}^+\times[0,2\pi[\to\mathbb{R}^2.$$

This is easy to check that the Jacobian is simply r in this case.

Lemma A.2.1. *The characteristic function of this normal distribution writes*

$$\phi_N(s)=\mathbb{E}e^{isN}=e^{-\frac{s^2}{2}} \tag{A.4}$$

Proof. Indeed the Laplace transform $L_N(z)=\mathbb{E}e^{zN}$ is easy to compute in case $z\in\mathbb{R}$:

$$L_N(z)=\mathbb{E}e^{zN}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{zx-\frac{x^2}{2}}\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{z^2}{2}-\frac{(x-z)^2}{2}}\,dx=e^{\frac{z^2}{2}}$$

with the binomial formula $(x-z)^2=x^2-2zx+z^2$ and after a change in variable $x\mapsto x-z$.
From dominated convergence theorem the application $z\mapsto L_N(z)$ is an entire function over $\mathbb{C}$.
The principle of analytic continuation implies that this formula remains valid for each $z\in\mathbb{C}$, and in particular we obtain

$$\phi_N(s)=L_N(is)=e^{-\frac{s^2}{2}}. \qquad \blacksquare$$

Equation (A.4) may also be rewritten:

$$\mathbb{E}e^{zN-\frac{z^2}{2}}=1, \qquad \forall z\in\mathbb{C}. \tag{A.5}$$

From the analicity of ϕ_N over the whole complex plane $\mathbb{C}$, the distribution of a Normal random variable is given from its characteristic function.

Definition A.2.1. *A random variable Y admits the Gaussian law $Y\sim\mathcal{N}(m,\sigma^2)$ if it can be written $Y=m+\sigma N$ for $m,\sigma\in\mathbb{R}$ and for a Normal random variable N.*

The density and the characteristic function of such distributions are derived from linear changes in variable:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(y-m)^2}{2\sigma^2}}, \qquad \phi_Y(t) = e^{itm} e^{-\frac{1}{2}\sigma^2 t^2}.$$

Gaussian samples, Gaussian densities and a Normal repartition functions are reproduced in Figures A.2, A.3, and A.4.

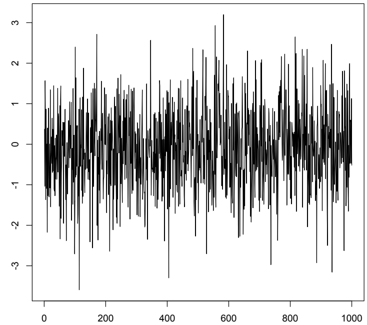


Figure A.2: Gaussian white noise (iid sample).

Exercise 32 (Similarity properties of the Normal law)**.**

1. *An important property is that if random variables $Y_j \sim \mathcal{N}(m_j, \sigma_j^2)$ are independent for $j = 1, 2$, then*

$$Y_1 + Y_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

2. *A converse of this result is that if Y_1, Y_2 are independent and with a same distribution μ. If $(Y_1 + Y_2)/\sqrt{2} \sim \mu$ admits the same distribution then μ is a centered Gaussian distribution.*
Hint. This property follows from a property of characteristic functions.
The characteristic function $\gamma(t) = \int e^{tx}\mu(dx)$, satisfies

$$\gamma(t) = \gamma^2\Big(\frac{t}{\sqrt{2}}\Big),$$

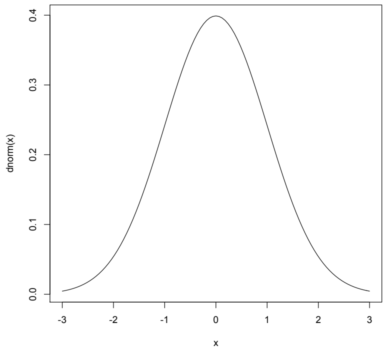


Figure A.3: Normal density.

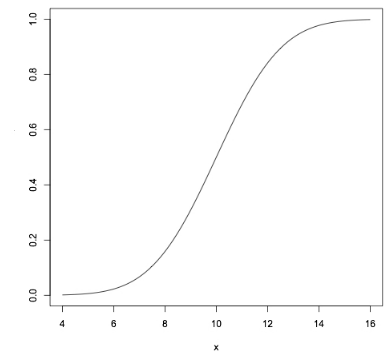


Figure A.4: The $\mathcal{N}(10, 4)-$Normal cumulative distribution function.

from independence.
To prove that this characterizes Gaussians, it may be shown that the log-characteristic function is a second degree polynomial.
With this formula, a simple recursion entails that there exists a con-

stant $a \in R$ such that $\log \gamma(t) = at^2$ for $t = k2^n$ with $k, n \in \mathbb{Z}$. A continuity argument allows to conclude. ∎

A.2.2 Multivariate Gaussians

Definition A.2.2. *A random vector $Y \in \mathbb{R}^k$ is Gaussian if the scalar product $Y \cdot u = Y^t u$ admits a Gaussian distribution for each $u \in \mathbb{R}^k$.*

Some essential features of Gaussian laws follow.

- **Gaussian laws only depends on second order properties**

  The law of a Gaussian Y random variables only depends on the mean and on the variance matrix. For

  $$u \in \mathbb{R}^k, \qquad \Sigma = \mathbb{E}(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)',$$

  we easily check that $Y \cdot u \sim \mathcal{N}(\mathbb{E}Y \cdot u, u^t \Sigma u)$ only depends on u, $\mathbb{E}Y$, and on Σ.

  An important consequence is that, for Gaussian vectors, orthogonality and independence coincide (alternatively this property may also be derived from the expression of characteristic functions).

- **Reduction of Gaussian vectors**

  Let Y be such a Gaussian vector, $\Sigma = \mathbb{E}(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^t$, admits a symetric nonnegative square root R, $\Sigma = R^2$. Indeed Σ is nonnegative symmetric ([6]), thus it is diagonalisable in an orthonormal basis thus there exists an orthogonal matrix Ω and a diagonal matrix D with $\Sigma = \Omega' D \Omega$ and $\Omega'\Omega = I_k$.

  Since Σ is nonnegative, the matrix D admits nonnegative diagonal coefficients (positive if Σ is a definite matrix). The nonnegative diagonal matrix Δ with elements the square roots of those of D satisfies $D = \Delta^2$.

  Thus $R = \Omega^t \Delta \Omega$, is a convenient square root (nonnegative symmetric) of Σ. This solution may be proved to be unique in case Σ is definite, because eigenspaces of R and Σ coincide from the fact that those matrices commute.

  In this case $Z = R^{-1}(Y - \mathbb{E}Y)$ is a Gaussian vector with orthogonal and normal $\mathcal{N}(0, 1)$ coordinates. The previous remark proves that those components are independent identically distributed thus $Z \sim \mathcal{N}_k(0, I_k)$.

---

[6]Indeed $u'\Sigma u = \mathrm{Var}\,(Y \cdot u) \geq 0$ for each $u \in \mathbb{R}^k$,

- **Density**

  Through a change in variables: if Σ is invertible then Y admits a density on $\mathbb{R}^k$:

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} e^{-\frac{1}{2}(y-\mathbb{E}Y)^t \Sigma^{-1}(y-\mathbb{E}Y)} \tag{A.6}$$

- **Characteristic function**

  Even for Σ non invertible we may write $Y = \mathbb{E}Y + RZ$. Thus for each $s \in \mathbb{R}^k$:

$$\begin{aligned}
\phi_Y(s) &= \mathbb{E}e^{is\cdot Y} \\
&= e^{is\cdot\mathbb{E}Y}\mathbb{E}e^{is\cdot RZ} \\
&= e^{it\cdot\mathbb{E}Y}\mathbb{E}e^{iZ\cdot Rs} \\
&= e^{is\cdot\mathbb{E}Y - \frac{1}{2}(Rs)\cdot(Rs)} \\
\phi_Y(s) &= e^{is\cdot\mathbb{E}Y - \frac{1}{2}(s^t\Sigma s)} \qquad (\text{A.7})
\end{aligned}$$

- **Conditioning**

  Let $(X,Y) \sim \mathcal{N}_{a+b}(0,\Sigma)$ be a Gaussian vector with covariance matrix written in blocs

$$\begin{pmatrix} I_a & C \\ C' & B \end{pmatrix}$$

  for some symmetric positive definite matrix B ($b \times b$) and a rectangular matrix C with order $a \times b$.

  Then $Z = Y - C'X$ est is orthogonal to X; hence from Gaussianness of this vector they are independent. Thus $\mathbb{E}(Y|X) = C'X$.

A.2.3 $\gamma-$distributions

As an example of the previous sections we introduce another important class of distributions.

Definition A.2.3. *The Euler function* Γ *of the first kind is defined over* $]0,+\infty[$ *by the relation*

$$\Gamma(t) = \int_0^\infty e^{-x}x^{t-1}\,dt.$$

Hints. Let $t \in \mathbb{R}$. The integral $\Gamma(t)$ is that of a positive and continuous function over $]0,+\infty[$. This is always a convergent integral at infinity but $t > 0$ is necessary to ensure the convergence at origin.

Integration by parts together with the relation $\dfrac{d}{dx}x^t = tx^{t-1}$ entails

$$\Gamma(t+1) = \int_0^\infty \frac{d}{dx}\{-e^{-x}\}x^t\,dt = \left[(-e^{-x})x^t\right]_0^\infty - t\int(-e^{-x})x^{t-1}\,dx.$$

Moreover a simple calculation proves tat $\Gamma(1) = 1$ thus a recursion using the previous identity entails $\Gamma(k) = (k-1)!$ for $k \in \mathbb{N}$, thus:

Lemma A.2.2. *Let $t > 0$ then $\Gamma(t+1) = t\Gamma(t)$ and $\Gamma(k) = (k-1)!$ for each $k \in \mathbb{N}^*$ (with the convention $0! = 1$).*

Definition A.2.4. *Set for $b > 0$, $c_{a,b} = b^a/\Gamma(a)$.*
In case $b > 0$, the law $\gamma(a,b)$ is the law with density

$$f_{a,b}(x) = c_{a,b}\, e^{-bx}x^{a-1}\,1\!\!1_{\{x>0\}}.$$

Proof. The function $f_{a,b}$ is integrable around infinity in case $b > 0$ and this integral converges at 0 if $a > 0$.
As a density admits the integral 1, we compare both integrals to get:

$$c_{a,b}^{-1} = \int_0^\infty e^{-bx}x^{a-1}dx = b^{-a}\int_0^\infty e^{-y}y^{a-1}dy = b^{-a}\Gamma(a),$$

by using a change of variable $y = bx$. Thus $c_{a,b} = b^a/\Gamma(a)$. ∎

Some simple facts are easily derived:

Lemma A.2.3. *Let $Z \sim \gamma(a,b)$ then for $m > 0$ and $\Re(u) < b$:*

$$\begin{aligned} \mathbb{E}Z^m &= \frac{\Gamma(a+m)}{b^m\Gamma(a)}, \\ L_{a,b}(u) &= \mathbb{E}e^{uZ} = \left(\frac{b}{b-u}\right)^a. \end{aligned}$$

Proofs.

$$\mathbb{E}Z^m = c_{a,b}\int_0^\infty x^m e^{-bx}x^{a-1}dx = \frac{c_{a,b}}{c_{a+m,b}} = \frac{\Gamma(a+m)}{b^m\Gamma(a)}.$$

To compute Laplace transform $L_{a,b}(u) = \mathbb{E}e^{uZ}$ of Z, we first assume that $u \in \mathbb{R}$:

$$L_{a,b}(u) = c_{a,b}\int_0^\infty e^{(u-b)x}x^{a-1}\,dx = \frac{c_{a,b}}{c_{a,b-u}} = \left(\frac{b}{b-u}\right)^a.$$

This is an analytic function in case $\Re(u) < b$ since integrals defining $L_{a,b}(u)$ is absolutely convergent because of

$$\left| e^{(u-b)x} x^{a-1} \right| = e^{(\Re u - b)x} x^{a-1}$$

The same holds for the complex derivative $ue^{(u-b)x}x^{a-1}$.
Analytic continuation allows to conclude. ∎

Easy consequences of this lemma follow:

Corollary A.2.1. *Let Z, Z' be two independent random variables with respective distributions $\gamma(a, b)$ and $\gamma(a', b)$, then*

$$Z + Z' \sim \gamma(a + a', b).$$

Proof. The previous lemma implies

$$\mathbb{E}e^{u(Z+Z')} = L_{a,b}(u)L_{a',b}(u) = L_{a+a',b}(u)$$

if $\Re u < a \wedge a'$, then the result follows from uniqueness of Laplace transforms in case they and analytic on a domain with non-empty interior. ∎

Exercise 33. *An alternative proof of Corollary A.2.1.*

- *Define Euler function of the second kind for $a, a' > 0$:*

$$B(a, a') = \int_0^1 u^{a-1}(1-u)^{a'-1}\, du.$$

  (Prove that the above expression is well defined.)

- *Prove that for $a, a' > 0$:*

$$B(a, a') = \frac{\Gamma(a)\Gamma(a')}{\Gamma(a + a')}.$$

- *Prove again Corollary A.2.1 without using the notion of Laplace transform and the principle of analytical continuation.*

Proofs. If $a, a' > 0$, the function

$$B(a, a') = \int_0^1 u^{a-1}(1-u)^{a'-1}\, du,$$

is well defined, indeed such integrals converge at origin since $a > 0$ and at point 1, it is due to the fact that $a' > 0$.
Let g be a continuous and bounded function then for such independent $Z \sim \gamma(a, b)$ and $Z' = \gamma(a', b)$ one derives:

$$\begin{aligned}
\mathbb{E}g(Z+Z') &= \int_0^\infty \int_0^\infty g(z+z') f_{a,b}(z) f_{a',b}(z')\, dz dz' \\
&= \int_0^\infty g(u)\, du \int_0^u f_{a,b}(z) f_{a',b}(u-z)\, dz \\
&= c_{a,b} c_{a',b} \int_0^\infty e^{-bu} g(u)\, du \int_0^u z^{a-1}(u-z)^{a'-1}\, dz, \\
&= c_{a,b} c_{a',b}\ B(a,a') \int_0^\infty u^{a+a'-1} e^{-bu} g(u)\, du \text{ with } \ z = ut \\
&= \frac{b^{a+a'} B(a,a')}{\Gamma(a)\Gamma(a')} \int_0^\infty u^{a+a'-1} e^{-bu} g(u)\, du
\end{aligned}$$

thus $Z + Z'$ admits a $\gamma(a + a', b)-$distribution. Now the normalization constant writes in two different ways which entails:

$$\frac{b^{a+a'} B(a,a')}{\Gamma(a)\Gamma(a')} = \frac{b^{a+a'}}{\Gamma(a+a')},$$

so that $B(a, a') = \dfrac{\Gamma(a)\Gamma(a')}{\Gamma(a+a')}$. ∎

Example A.2.1. *From the above results*

- *The density of the sum S_k of k independent random variables with exponential distribution $\mathcal{E}(\lambda)$ for $\lambda > 0$ is $\gamma(k, \lambda)$.*
- *Define χ_k^2 distribution as the distribution of*

  $$T_k = N_1^2 + \cdots + N_k^2$$

  for independent and normally distributed $\mathcal{N}(0, 1)$ random variables $N_1, \ldots, N_k$. The law of χ_k^2 is $\gamma(\frac{k}{2}, \frac{1}{2})$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
- *For $m \in \mathbb{N}$ and $N \sim \mathcal{N}(0, 1)$, $\mathbb{E}N^m = 0$ for m odd and $\mathbb{E}N^m = (2p)!/2^p p!$ for $m = 2p$ an even number.*

Hints. For $k = 1$, $S_1 \sim \mathcal{E}(\lambda)$ admits a $\gamma(1, \lambda)-$distribution, and $T_1 = N^2$ is the square of a standard Normal; we compute its density from the expression of $\mathbb{E}g(T_1)$ for each bounded and continuous function $g : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned}
\mathbb{E}g(T_1) &= \mathbb{E}g(N^2) \\
&= \int_{-\infty}^{\infty} g(x^2) e^{-x^2/2} \frac{dx}{\sqrt{2}\pi} \\
&= 2\int_{0}^{\infty} g(x^2) e^{-x^2/2} \frac{dx}{\sqrt{2}\pi} \\
&= 2\int_{0}^{\infty} g(z) \frac{1}{2\sqrt{z}} e^{-z/2} \frac{dz}{\sqrt{2}\pi}, \text{ (with } \ z = x^2) \\
&= \int_{0}^{\infty} g(z) z^{\frac{1}{2}-1} e^{-z/2} \frac{dz}{\sqrt{2}\pi}
\end{aligned}$$

Thus the density function of T_1's distribution is $z^{\frac{1}{2}-1}e^{-z/2}/\sqrt{2}\pi$ for $z \geq 0$. Up to a constant this is the density $f_{\frac{1}{2},\frac{1}{2}}$ of a law $\gamma(\frac{1}{2},\frac{1}{2})$.
Since they are both densities it implies $c_{\frac{1}{2},\frac{1}{2}} = 1/\sqrt{2}\pi$. and thus $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
Now addition formulas allow to conclude for $k > 1$ that $S_k \sim \gamma(k, \lambda)$ and $T_k \sim \chi_k^2 \sim \gamma(\frac{k}{2}, \frac{1}{2})$. The last result follows either from the previous result Lemma A.2.3 since

$$\mathbb{E}T_1^p = \frac{\Gamma(\frac{1}{2}+p)}{2^{-p}\Gamma(\frac{1}{2})},$$

but it needs some additional effort. A simpler way to proceed is to use relation A.4 and from comparing both sides of the expansion of $\mathbb{E}e^{itN} = e^{-t^2/2}$,

$$\mathbb{E}e^{itN} = \sum_m \frac{1}{m!}(it)^m \mathbb{E}N^m, \quad e^{-t^2/2} = \sum_p \frac{1}{p!}\left(\frac{-t^2}{2}\right)^p.$$

Clearly the parity of the characteristic function implies that all odd moments vanish. Now for $m = 2p$, we obtain:
$\dfrac{((-t^2)/2)^p}{p!} = \mathbb{E}N^{2p}(-1)^p \dfrac{t^{2p}}{(2p)!}$. ∎

A.3 Convergence

A.3.1 Convergence in distribution

We consider a sequence of random variables X_n and a random variable X with values in an arbitrary complete separable metric space (E, d) .

Definition A.3.1. *The sequence X_n converges in distribution to X, which we denote*

$$X_n \to^{\mathcal{L}}_{n\to\infty} X, \qquad \text{if} \qquad \mathbb{E}g(X_n) \to_{n\to\infty} \mathbb{E}g(X).$$

for any continuous and bounded function $g : E \to \mathbb{R}$.

This definition does not depend on the random variables but really only on their distribution $P_{X_n} \to P_X$ and thus we really define the convergence of probability measures on a metric space.
An example of important metric space follows.

Definition A.3.2. *The Skorohod space $D[0,1]$ is the space of functions $[0,1] \to \mathbb{R}$ continuous from the right and admitting a limit on the left at each point $t \in [0,1]$. For short they are called cadlag functions.*

Example A.3.1. *Such cadlag functions are:*

- *Continuous functions are cadlag, $C[0,1] \subset D[0,1]$,*
- *Indicators are also cadlag, $x \mapsto g_t(x) = \mathbb{1}_{\{x\le t\}}$ for each $t \in [0,1]$.*

The metric $d(f,g) = \sup_t |f(t) - g(t)|$ is natural on the space $C[0,1]$ of continuous real valued functions on the interval.
Note that the indicator function $g_{\frac12}$ may be *approximated* by a sequence of piecewise affine functions f_n with Lip $f_n = n$ and $f_n(x) = \mathbb{1}_{\{x\le\frac12\}}$ for $|x-\frac12| \ge \frac1n$ but this sequence is not $d-$Cauchy. If $\lim_n d(f_n, g_{\frac12}) = 0$ then f_n should also have a jump at $\frac12$ for large values of n.
$D[0,1]$ is not separable with the metric d since $d(g_s, g_t) = 1$ if and only if $s \ne t$.

Remark A.3.1 (Prohorov metric)**.** *Let $\mathcal{H}$ be the set of monotonic homeomorphisms ([7]) $[0,1] \to [0,1]$ then a reasonable metric on $D[0,1]$ is*

$$\delta(f,g) = \inf_{\lambda\in\mathcal{H}} \left\{ d(f\circ\lambda, g) + \sup_{t\in[0,1]} |\lambda(t) - t| \right\}.$$

This metric makes $D[0,1]$ complete and separable. We shall not prove this but this is simple to prove that $\lim_n \delta(g_t, g_{t+\frac1n}) = 0$.
Quote that $\delta \le d$ thus for example the function $f \mapsto \sup_{0\le t\le 1} f(t)$ is a continuous function on this space $(D[0,1],\delta)$. Convergence in this space is addressed in [Billingsley, 1999]; it is not in the scope of those notes.
A criterion for the convergence of empirical repartition functions $Z_n(t) = n^{-\frac12}(F_n(t) - t)$ of a stationary sequence with uniform marginal distribution (see in [Dedecker et al., 2007]) is

[7] i.e. bijective continuous functions with a continuous inverse.

- *For each $d-$tuple $t_1, \dots, t_d \in [0,1]$ the sequence of random vectors $(Z_n(t_1), \dots, Z_n(t_d))$ converges in distribution to some Gaussian random variable in $\mathbb{R}^d$.*
- *There exist constants $a, b, p > 1$ and $C > 0$ such that for each $s, t \in [0,1]$*
$$\mathbb{E}|Z_n(t) - Z_n(s)|^p \leq C\left(|t-s|^a + n^{-b}\right).$$

From now on, we shall restrict to the case $E = \mathbb{R}^d$. In this case,

Lemma A.3.1 (Tightness)**.** *Let X be a rv on $\mathbb{R}^d$. For each $\epsilon > 0$ there exists a compact subset of E such that $\mathbb{P}(X \notin K) < \epsilon$.*

Proof. Note that $\Omega = \bigcup_{n=1}^{\infty} A_n$ with $A_n = (|X| \leq n)$. Hence from the sequential continuity of the probability $\mathbb{P}$ there exists n such that $\mathbb{P}(A_n^c < \epsilon)$. The closed ball with radius n is now a convenient choice $K = B(0,n)$. ∎

Remark A.3.2. *This result allows to* **restrict to a compact set**. *This is easy to prove that the previous convergence holds in case the class of continuous and bounded test functions is replaced by a smaller class of functions:*

- *The class of uniformly continuous and bounded functions ([8]).*
- *The class of functions C_b^3 with third order continuous and bounded partial derivatives ([9]).*
- *If $\phi_{X_n}(t) \to \phi_X(t)$ for each $t \in \mathbb{R}^d$.*
  Indeed, from Stone-Weierstrass theorem which asserts the density on trigonometric polynomials on the space $C(K)$ of continuous real valued functions on a compact $K \subset \mathbb{R}^d$, equipped with the uniform norm $\|g\|_K = \sup_{x \in K} |g(x)|$.

---

[8] The restriction of a continuous function over a compact set is uniformly continuous. Indeed, from Heine theorem, a continuous over a compact set is uniformly continuous.

[9] From a convolution approximation with a bounded and indefinitely derivable function with integral 1 ϕ, $f_\epsilon = f \star \phi_\epsilon$ converges uniformly over compact subsets to f as $\epsilon \downarrow 0$, if one sets $\phi_\epsilon(u) = \frac{1}{\epsilon}\phi(u/\epsilon)$.
Now convolution inherits of ϕ's regularity. Indeed the Lebesgue dominated convergence applies to prove that eg.

$$f_\epsilon'(u) = \lim_{h \to 0} \frac{1}{h}(f_\epsilon(u+h) - f_\epsilon(u)) = f \star \phi_\epsilon'(u).$$

- *If a sequence of characteristic functions converges uniformly on a neighborhood of 0 then its limit is also the characteristic function of a law μ (Paul Lévy).*

A.3.2 Convergence in probability

From now on we shall consider **pathwise convergence** only.

Definition A.3.3. *The sequence X_n converges in probability to X, which we denote*

$$X_n \to^{\mathbb{P}}_{n\to\infty} X$$

if, for each $\epsilon > 0$:

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

Lemma A.3.2. *If a real valued sequence of random variables X_n converges in probability to X, then it converges in distribution.*

Proof. Assume that convergence in probability holds then lemma A.3.1 then we may assume that g is uniformly continuous in the definition of convergence in distribution. Let $\epsilon > 0$ we set $A = (|X_n - X| \geq \epsilon)$

$$\begin{aligned} |\mathbb{E}(g(X_n) - g(X))| &= \left|\mathbb{E}(g(X_n) - g(X))1_A + \mathbb{E}(g(X_n) - g(X))1_{A^c}\right| \\ &\leq 2\|g\|_\infty \mathbb{P}(A_n) + \sup_{|x-y|<\epsilon} |g(x) - g(y)|. \end{aligned}$$

Uniform continuity of g yields convergence in law.
An alternative proof makes use of Lévy theorem, see Remark A.3.2 for details. ∎

Definition A.3.4. *If $\mathbb{E}|X_n - X|^p \to_{n\to\infty} 0$ we say that the sequence X_n converges to X in $\mathbb{L}^p$.*

Remark A.3.3 (relations between convergences)**.**

- *Convergence in probability implies convergence in probability, see Lemma A.3.2.*
- *Convergence in distribution does not imply convergence in probability.*
  A dyadic scheme allows to write $(0,1]$ as the union of the 2^n disjoint intervals

  $$I_{j,n} =]j2^{-n}, (j+1)2^{-n}], \quad (0 \leq j < 2^n),$$

with the same measure 2^{-n}.
This is thus possible to write $[0,1] = A_n \bigcup B_n$ *where both sets admit the measure* $\lambda(A_n) = \lambda(B_n) = \frac{1}{2}$, *by setting eg.*

$$A_n = \bigcup_{j=1}^{2^{n-1}} I_{2j,n}, \qquad B_n = \bigcup_{j=1}^{2^{n-1}} I_{2j-1,n}.$$

On the probability space $((0,1], \mathcal{B}((0,1], \lambda)$, *the sequence* $X_n = 1\!\!1_{A_n}$ *follows the same Bernoulli distribution* $b(\frac{1}{2})$ *thus it converges in distribution to* X_0.
Now the sequence X_n *does not converge in probability since*

$$\lambda\left(A_n \cap \left[0, \frac{1}{2}\right]\right) = \frac{1}{4} < \frac{1}{2} = \lambda(A_0).$$

Indeed $\mathbb{P}(X_n < \frac{1}{2}) = \frac{1}{4}$ *cannot converge to* $\frac{1}{2}$ *thus no subsequence of* X_n *may be convergent in probability to* X_0. ∎

- *From Markov inequality applied to* $V = |X_n - X|$ *this is immediate to derive that* $\mathbb{L}^p$ *convergence implies convergence in probability.*
- *However if the random variable* Z *satisfies* $\mathbb{E}|Z|^p = \infty$ *and* $\mathbb{E}|Z|^q < \infty$ *for each* $q < p$ *then the sequence* $X_n = Z/n$ *converges to* $X = 0$ *in probability but not in* $\mathbb{L}^p$.
  Indeed Markov inequality implies

  $$\mathbb{P}(|X_n| > \epsilon) \leq \frac{\mathbb{E}|Z|^q}{n^q \epsilon^q} \to_{n\to\infty} 0,$$

  for each $\epsilon > 0$ *in case* $q \in (0, p[$. ∎

As an example think of Z *with a Cauchy distribution and* $p = 1$.

A.3.3 Almost sure convergence

Definition A.3.5. *The sequence* X_n *converges almost surely to* X, *which we denote*

$$X_n \to^{a.s}_{n\to\infty} X$$

if there exists an event A *with* $\mathbb{P}(A) = 0$ *such that for each* $\omega \notin A$

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Again a.s. convergence implies convergence in probability.

Definition A.3.6 (Limit superior). *For a sequence of events B_n, set*

$$\overline{\lim_n} B_n = \bigcap_n \bigcup_{k \geq n} B_k.$$

Remark A.3.4. *Note that $A_n = \bigcup_{k \geq n} B_k$ is a decreasing sequence of events.*

Lemma A.3.3 (Borel-Cantelli). *If $(B_n)_{n \in \mathbb{N}}$ is a sequence of events such that $\sum_n \mathbb{P}(B_n) < \infty$ then*

$$\mathbb{P}(\overline{\lim_n} B_n) = 0.$$

Remark A.3.5.

- *If $X_n \to X$ in probability then some subsequence of X_n also converges a.s.*
  Indeed for each $k > 0$ $\lim_n \mathbb{P}(kZ_n > 1) = 0$ with $Z_n = |X_n - X|$. This is possible to extract a subsequence $\phi_k(m)$ such that

  $$\sum_m \mathbb{P}(kZ_{\phi_k(m)} > 1) < \infty.$$

  Then from Borel-Cantelli lemma it is left to the reader to prove that the diagonal scheme $T_m = Z_{\phi_m(m)}$ is almost surely convergent. ∎

- *In Remark A.3.3 we use a dyadic scheme $(I_{j,n})_{0 \leq j < 2^n}$ for $n = 1, 2, 3, \ldots$, now the sequence $X_n = \mathbb{1}_{A_n}$ does not have any a.s. convergent subsequence.*

Bibliography

[Andrews, 1984] Andrews, D. (1984). Non strong mixing autoregressive processes. *J. Appl. Prob*, 21:930–934.

[Azencott and Dacunha-Castelle, 1987] Azencott, R. and Dacunha-Castelle, D. (1987). *Séries d'observations irrégulières; modélisation et prévisions.* Masson, Paris.

[Bardet et al., 2006] Bardet, J. M., Doukhan, P., Lang, G., and Ragache, N. (2006). Dependent Lindeberg central limit theorem and some applications. *ESAIM P&S*, 12:154–171.

[Billingsley, 1999] Billingsley, P. (1999). *Convergence of probability measures.* Second edition, Wiley and sons.

[Bradley and Pruss, 2009] Bradley, R. and Pruss, A. (2009). A strictly stationary, $n-$tuplewise independent counterexample in the central limit theorem. *Stoch. Proc. Appl.*, 119(10):3300–3318.

[Breuer and Major, 1983] Breuer, P. and Major, P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13:425–441.

[Brockwell and Davis, 1991] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer Series in Statistics. Second edition, Springer.

[Choquet, 1973] Choquet, G. (1973). *Topologie, Volume II.* Masson, Paris.

[Colomb, 1977] Colomb, G. (1977). Estimation non paramétrique de la régression par la méthode du noyau: propriété de convergence asymptotiquememt normale indépendante. *Ann. Fac. sc. Univ. Clermont. Série Mathématique*, 65-5:24–46.

[Davydov, 1970] Davydov, Y. A. (1970). Invariance principle for stationary processes. *Theor. Probab. Appl.*, 15-3:498–509.

[Dedecker and Doukhan, 2003] Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stoch. Proc. Appl.*, 106:63–80.

[Dedecker et al., 2007] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., and Prieur, C. (2007). *Weak dependence: With Examples and Applications.* Lecture Notes in Statistics **190**, Springer-Verlag, New York.

[Derriennic and Klopotowski, 2000] Derriennic, Y. and Klopotowski, A. (2000). On bernstein's example of three pairwise independent random variables. *Sankhya, Series A*, 62(3):318–330.

[Diaconis and Freedman, 1995] Diaconis, P. and Freedman, D. (1995). Iterated random functions. *SIAM Rev.*, 41-1:45–78.

[Dobrushin and Major, 1979] Dobrushin, R. L. and Major, P. (1979). Non-central limit theorems for non-linear functions of Gaussian fields. *Z. Wahrsch. und verw. Gebiete*, 50:27–52.

[Doukhan, 1988] Doukhan, P. (1988). Formes de Toëplitz associées à une analyse multiéchelle. *CRAS Paris Série 1*, 306:663–666.

[Doukhan, 1994] Doukhan, P. (1994). *Mixing: properties and examples*, volume 85 of *Lecture Notes in Statistics.* Springer-Verlag.

[Doukhan and Lang, 2009] Doukhan, P. and Lang, G. (2009). Evaluation for moments of a ratio with application to regression estimation. *Bernoulli*, 15-4:1259–1286.

[Doukhan et al., 2005] Doukhan, P., Lang, G., and Surgailis, D. (2005). Functional CLTs for short or long memory linear sequences. *Ann. Inst. H. Poincaré B*, 38-6:879–896.

[Doukhan et al., 2007] Doukhan, P., Lang, G., and Surgailis, D. (2007). Randomly fractionally integrated processes. *Lithuanian Mathematical Journal*, 47:3–28.

[Doukhan et al., 2002a] Doukhan, P., Lang, G., Surgailis, D., and Viano, M. C. (2002a). Functional limit theorem for the empirical process of a class of Bernoulli shifts with long memory. *Journal of Theoretical Probability*, 18-1:161–186.

[Doukhan and León, 1989] Doukhan, P. and León, J. R. (1989). Cumulants for mixing sequences and applications to empirical spectral density. *Probab. Math. Statist.*, 10-1:11–26.

[Doukhan and Louhichi, 1999] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stoch. Proc. Appl.*, 84:313–342.

[Doukhan et al., 2009] Doukhan, P., Mayo, N., and Truquet, L. (2009). Weak dependence, models and some applications. *Metrika*, 69(2-3):199–225.

[Doukhan et al., 2002b] Doukhan, P., Oppenheim, G., and Taqqu, M. (2002b). *Theory and Applications of Long-Range Dependence*. Birkhaüser.

[Doukhan et al., 2011] Doukhan, P., Prohl, S., and Y., R. C. (2011). Extremes of weakly dependent times series, discussion paper. *TEST*, 20(3):447–502.

[Doukhan and Sifre, 2001] Doukhan, P. and Sifre, J.-C. (2001). *Cours d'Analyse, Volumes 1*. Masson, Paris.

[Duflo, 1996] Duflo, M. (1996). *Random Iterative Models*. Springer-Verlag.

[Esary et al., 1967] Esary, J., Proschan, F., and Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.*, 38:1466–147.

[Fortuin et al., 1971] Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22-2:89–103.

[Giraitis et al., 2012] Giraitis, L., Koul, H., and Surgailis, D. (2012). *Large sample inference for long memory processes.* Imperial College Press.

[Giraitis and Surgailis, 1999] Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *J. Stat. Plan. Inf.*, 100:81–93.

[Ho and Hsing, 1996] Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long memory moving averages. *Ann. Stat.*, 24:992–1024.

[Kallenberg, 1997] Kallenberg, O. (1997). *Foundations of modern probability.* Springer-Verlag.

[Kazmin, 1969] Kazmin, J. A. (1969). Appell polynomial series expansions. *Mat. Zametki*, 5:509–520.

[Kedem and Fokianos, 2002] Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis.* Wiley, Hoboken, NJ.

[Lorentz, 1966] Lorentz, G. G. (1966). *Approximation of functions.* AMS Providence, Chelsea Publishing company.

[Major, 1981] Major, P. (1981). *Multiple Ito-Wiener Integrals.* Number 849 in Lecture Notes in Mathematics. Springer-Verlag, Berlin.

[Massart, 2007] Massart, P. (2007). *Concentration Inequalities and Model Selection; Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003, editor Jean Picard*, volume 1896 of *Lecture Notes in Mathematics.* Springer-Verlag.

[Newman, 1984] Newman, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In Tong, Y. L., editor, *Inequalities in Statistics and Probability*, volume 5 of *IMS Lecture Notes-Monograph Series*, pages 127–140. Elsevier.

[Nourdin et al., 2011] Nourdin, I., Peccati, G., and Podolskij, M. (2011). Quantitative breuer-major theorems. *Stochastic Processes and their Applications*, 121-4:793–812.

[Petrov, 1975] Petrov, V. V. (1975). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables.* Oxford Studies in Probability.

[Philippe et al., 2008] Philippe, A., Surgailis, D., and Viano, M. C. (2008). Time-varying fractionally integrated processes with nonstationary long memory. *Theory Probab. Appl.*, 52-4:651–673.

[Politis, 2003] Politis, D. N. (2003). Adaptive bandwidth choice. *J. Nonparam. Statist.*, 15:517–533.

[Polya and Szegö, 1970] Polya, G. and Szegö, G. (1970). *Problems and theorems in analysis, 4th edition*, volume 1. Classics in mathematics, Springer, New-York.

[Rio, 2000] Rio, E. (2000). *Théorie asymptotique pour des processus aléatoires faiblement dépendants.* Number 31 in Mathématiques et Applications. Springer-Verlag.

[Rosenblatt, 1956] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Set. USA*, 42:43–47.

[Rosenblatt, 1961] Rosenblatt, M. (1961). Independence and dependence. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, Univ. California Press:431–443.

[Rosenblatt, 1985] Rosenblatt, M. (1985). *Stationary processes and random fields.* Birkhäuser, Boston.

[Rosenblatt, 1991] Rosenblatt, M. (1991). *Stochastic curve estimation*, volume 3. NSF-CBMS Regional Conference Series in Probability and Statistics.

[Sansone, 1959] Sansone, G. (1959). *Orthogonal functions.* Intersciences, New-York.

[Saulis and Statulevicius, 1991] Saulis, L. and Statulevicius, V. A. (1991). *Limit theorems for large deviations.* Birkhäuser.

[Slepian, 1972] Slepian, D. (1972). On the symmetrized kronecker power of a matrix and extensions of Mehler's formula for Hermite polynomials. *SIAM Journal on Mathematical Analysis*, 3:606–616.

[Soulier, 2001] Soulier, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statistics & Probability Letters*, 54(2):193–203.

[Straumann, 2005] Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models*, volume 181 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

[Szegö, 1959] Szegö, G. (1959). *Orthogonal Polynomials*, volume 23. Colloquium Publication, American mathematical society, New-York.

[van der Vaart, 1998] van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics.

[van der Vaart and Wellner, 1998] van der Vaart, A. and Wellner, J. A. (1998). *Weak convergence and empirical processes*. Springer Verlag series in statistics.

Index

Back Cover: Modeling nonlinear time series *by Paul Doukhan.*

The main focus of this book is to present examples and tools for modeling non-linear time series. The volume is divided 3 parts and an Appendix.
The main standard tools of probability and statistics which directly apply to the time series case are rapidly described in order to get a wide panel of modeling possibilities. For example functional estimation and bootstrap are rapidly recalled. Stationarity is first introduced and this volume will keep this assumption for clarity.
We then describe some tools from Gaussian chaos. Then we give a fast tour on linear time series models. Nonlinearity then appears from polynomial or chaotic models for which explicit expansions are available. Then we turn to Markov and non-Markov linear models; we provide standard examples such as ARCH-type, integer valued models... and their estimation. We also develop an useful tool: Bernoulli shifts time series models.
A third part addresses more theoretical tools with first the ergodic theorem which is seen as the first step for statistics of time series. Then distributional range is defined to get generic tools for limit theory under Long or Short Range dependences (LRD/SRD). Examples of LRD behaviors are made explicit. More general techniques to prove (central-)limit theorems are described under SRD; mixing and weak dependence are also recalled. Finally moment techniques are described together with there relations to cumulant sums as well as an application to kernel type estimation.
A short Appendix recalls basic facts of probability theory and useful laws issued from the Gaussian laws are discussed.
The text is illustrated with simulations and an index aims at helping a reader. This is an advanced master course. Expected reader are thus either mathematician intending to enter the field of time series as well as statisticians wanting to get a more mathematical background.