Lectures on non linear time series

Paul Doukhan, University Cergy-Pontoise and IUF,
Lectures in Torun, February 17-22, 2014

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.

Those notes are divided into 8 chapters.

1. A first chapter recalls essential concepts of probability and provides some applications for the case of iid inputs. inequalities or central limit theorems proved in the independent setting. A special emphasis is set on Lindeberg method with easily extends to a dependent setting. The kernel density estimates are described with some details as an application of previous results.

2. Stationarity is an assumption allowing a convenient limit theory. Filtering techniques are provided in Brockwell and Davis [4] to come back to the simpler stationary case. This makes more reasonable to assume the stationarity which is essentially invalid in practice.
   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. In particular we propose a simple definition of long or short range dependence.

3. Due to the CLT the Gaussian case plays a central role in statistics. 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 evocated. The fractional Brownian motion essential hereafter for the long range dependent setting is also introduced.

4. Due to the Lindeberg CLT the linear case first case considered after the Gaussian. Standard linear models of statistics ARMA, ARIMA of FARIMA models are thus defined. We again refer to [4] for further information.

5. Non linear models are then naturally considered as extensions of the previous ones. This chapter proposes a wide botanic for those models as well as the very general class of Bernoulli schemes. 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 view allows a general and simple overview of those various examples.

6. Associated processes are then investigated. 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.

7. Tools for asymptotic theory under long range or short range dependence are then developed in an quick way.
   The ergodic theorem is the first result proposed in this chapter. Precising further asymptotic expansions of the empirical mean is the really impact of short/long range dependence (SRD/LRD).
   Under LRD the more elementary examples are are seen to get such asymptotic explicit expansion in distribution up to nonGaussian limits. In the SRDcase we will give a rapid idea of techniques.

8. A last chapter is devoted to moment and cumulant inequalities developing the more standard spectral ideas of the second chapter.
   Applications of those techniques to spectral estimations are developed in an elegant way in Rosenblatt monographies [22, 23].

Books edited by Doukhan, Oppenheim and Taqqu (2002) [10] provide a wide amount of directions for the study of long range dependent processes. Papers by Doukhan and Louhichi (1999) [12] and Dedecker and Doukhan (2003) [7] as well as the monograph Dedecker, Doukhan, Lang, Leon, Louhichi and Prieur (2007) [8] also consider the weakly dependent setting.
The books by Azencott and Dacunha-Castelle (1984) and Rosenblatt (1985) [22] also lead to a large amount of additional developments.
Functional estimation frames are synthetically described in the great monograph by Rosenblatt [23].

Suggestions are welcome.

Paris, December 31, 2013

Paul Doukhan

Chapter 1

Probability and independence

1.1 Probabilist setting

For any space E, a sigma algebra $\mathcal{E}$ is a subset of $\mathcal{P}(E)$ (set of the subsets of E). such that

- $\emptyset \in \mathcal{E}$
- $\forall A \in \mathcal{E}: \; A^c \in \mathcal{E}$
- $\forall A_n \in \mathcal{E}, n = 1, 2, 3 \ldots : \; \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$

A measured space is such a couple $(E, \mathcal{E})$ ([1]).
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:

- $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
- $\forall A, B \in \mathcal{A}: \; A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
- $\forall A_i \in \mathcal{A}, i = 1, 2, 3 \ldots : \; A_1 \subset A_2 \subset \cdots t \Rightarrow \lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Elements of $\mathcal{A}$ are called events.
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 nulsets).
A random variable (rv) $X : \Omega \to E$ is a measurable function between those two measurable sets: this means that $X^{-1}(\mathcal{E}) \subset \mathcal{A}$. In other terms introducing the probabilist notation $(X \in A) = X^{-1}(A)$ for each $A \in \mathcal{E}$:

$$\forall A \in \mathcal{E}: \quad (X \in A) \in \mathcal{A}$$

[1] If E is a finite set with n elements then a reasonable choice of sigma-algebra is $\mathcal{E} = \mathcal{P}(E)$ which admits 2^n elements as it may be seen from the fact the the application: $A \mapsto 1_A$ defined for $\mathcal{P}(E)$ on the set of functions from E to $\{0, 1\}$ is a bijective function.

Note also that $\sigma(X) = X^{-1}(\mathcal{E})$ is the sigma-algebra generated by the rv $X : \Omega \to E$, that mean it is the smallest sub-sigma-algebra $\mathcal{F}$ of $\mathcal{A}$ which makes the application $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ measurable.

Also the image distribution 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 ([2]), 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. Anyway we assume that students will rectify by themselves the numerous errors of this type in those notes.

Definition 1.1.1 *For $E = \mathbb{R}^d$ in case those integrals converge one defines*

$$\mathbb{E}X = \int_E x P_X(dx) \in \mathbb{R}^d,$$

and

$$Cov(X) = \mathbb{E}XX' - \mathbb{E}X(\mathbb{E}X)', \qquad \text{is a symmetric positive } n \times n-\text{matrix},$$

In case $X = (X_1, X_2)$, we also write:

$$Cov(X) = \begin{pmatrix} Var X_1 & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var X_2 \end{pmatrix}$$

Note that for $X \geq 0, \mathbb{E}X \geq 0$.
A first but essential result is the following

Theorem 1.1.1 (Markov inequality) *Assume that $V \geq 0$ is a real valued non-negative random variable, then its expectation exists (in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$*

$$\mathbb{P}(V \geq u) \leq \frac{\mathbb{E}V}{u}, \qquad \forall u > 0$$

Proof. Set $A = (V \geq u)$,then:

$$\mathbb{E}V = \mathbb{E}V 1_A + \mathbb{E}V 1_{A^c} \geq \mathbb{E}V 1_A \geq u\mathbb{P}(A) \qquad \blacksquare$$

[2]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.

Proposition 1.1.1 (Jensen) *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)

$$\mathbb{E}g(Z) \geq g(\mathbb{E}Z) \tag{1.1}$$

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. More over 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

$$g'(x+) \leq g'(y-) \leq g'(y+) \leq g'(z-), \qquad \text{with} \quad g'(y\pm) = \lim_{h \to 0^0} \frac{g(y \pm h) - g(y)}{\pm h}$$

then for each $x_0 \in C$ choose any $u \in [g'(x_0-), g'(x_0+)]$ 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 $f \leq g$.
Thus 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 (the Hilbert case) of Hahn-Banach theorem the same representation of g holds and the proof is the same. ∎

Remark 1.1.1

- *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 5.3.1.*
- *Let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ algebra of $\mathcal{A}$, a conditional variant of this inequality writes*

$$\mathbb{E}^{\mathcal{B}} g(Z) \geq g\left(\mathbb{E}^{\mathcal{B}} Z\right) \tag{1.2}$$

  (it makes use a conditional version of the dominated convergence theorem).

Definition 1.1.2 *Let $X \in \mathbb{R}^d$ be a vector valued random variable then its characteristic function is defined as*

$$\phi_X(t) = \mathbb{E}e^{it \cdot X}, \qquad \forall t \in \mathbb{R}^d$$

When it exists the Laplace transform of the law of X is denoted

$$L_X(z) = \mathbb{E}e^{z.X}, \qquad \forall z \in Dom \subset \mathbb{C}^d$$

Remark 1.1.2 *First quote that the characteristic function always exists and $\phi_X(t) = L_X(it)$.*
If 0 is an point 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 and $\partial\phi(0)/\partial t_j = i\mathbb{E}X_j$.
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 iid rvs 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 rvs*: there exists a measurable function $f : E \to \mathbb{R}^+$ such that for each $A \in \mathcal{E}$:
$$P_X(A) = \int_A f(x)\,dx$$
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.
Examples are
  - $f = 1_{[0,1]}$ yields the uniform distribution, the Lebesgue measure on the unit interval.
  - $f(x) = \lambda e^{-\lambda x} 1_{x \ge 0}$ is the exponential law $\mathcal{E}(\lambda)$ with parameter λ.
  - The Normal law $\mathcal{N}(0,1)$ studied later on admits the density
  $$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

– Cauchy distribution is defined with

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

Clearly the mean of such Cauchy distributed rvs is not defined.

Example 1.1.1 *An example of probability space is* $\Omega = [0,1]^{\mathbb{Z}}$ *endowed with its product sigma-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* $A_n \neq [0,1]$ *only for only finitely many indices* n.
Then a sequence of random variables X_n *is defined as the* $n-$*th coordinate function*

$$X_n(\omega) = \omega_n, \quad \forall \omega = (\omega_n)_{n\in\mathbb{Z}}$$

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 1.1.1 (Hoeffding)

1. *Let* Z *be a 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*

$$Cov(X,Y) = \int_{-\infty}^\infty \int_{-\infty}^\infty (\mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)P(Y \geq t))\, ds\, dt$$

Proof.

1. Let $z \geq 0$ then

$$z = \int_0^\infty \mathbb{1}_{(z\geq t)}\, dt$$

The Fubini theorem is always valid in case of non-negative functions: apply it to the measure $\lambda \otimes \mathbb{P}_Z$ to conclude (here λ denotes Lebesgue measure).

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 rvs. The formula holds for each of them and $\mathbb{P}(X \geq s) = \mathbb{P}(X^+ \geq s)$ if $s \geq 0$ and $= 1 - \mathbb{P}(X^- > -t)$ else. Now for an arbitrary couple of real valued random variables $\mathrm{Cov}\,(X,Y)$ writes as a linear combination of such four integrals with coefficients ± 1. ∎

1.2 Independence

Definition 1.2.1 *Events $A, B \in \mathcal{A}$ are called independent in case*

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

Random variables $X_1, \ldots, X_n$ with values for instance in the same space E are called independent if for any continuous functions $g_1, \ldots, g_n : E \to \mathbb{R}$

$$\mathbb{E}\left(g_1(X_1) \cdots g_n(X_n)\right) = \mathbb{E}g_1(X_1) \times \cdots \times Eg_n(X_n)$$

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

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

For events $A_1, \ldots, A_n$ simple rvs are $X_k = 1_{A_k} \in \{0, 1\}$ and 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 $A_1, \ldots, A_n$ writes a bit differently, as:

$$\mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i), \qquad \forall I \subset \{1, \ldots, n\}$$

Pairwise independence should be carefully distinguished from independence.
Now quote that the example 1.1.1 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 rvs, then the calculation of generating functions implies $X_1 + \cdots + X_n \sim B(n, p)$ admits a Binomial distribution.

1.3 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 1.3.1 *The sequence X_n converges in distribution to X, which we denote*

$$X_n \to^{\mathcal{L}} X$$

if

$$\mathbb{E}g(X_n) \to_{n\to\infty} \mathbb{E}g(X)$$

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.
From now on, we shall restrict to the case $E = \mathbb{R}^d$. In this case,

Lemma 1.3.1 (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 sequential continuity of the probability $\mathbb{P}$ there exists n such that $\mathbb{P}(A_n^c < \epsilon)$. Then the ball with radius n is convenient, $K = B(0, n)$. ∎

Remark 1.3.1 *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 (from Heine theorem).*
- *The class of functions C_b^3 with third order continuous and bounded partial derivatives (from a convolution approximation).*
- *If $\phi_{X_n}(t) \to \phi_X(t)$ for each $t \in \mathbb{R}^d$ (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 $\|g\|_K = \sup_{x \in K} |g(x)|$).*
- *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).*

Lemma 1.3.2 (Lindeberg) *Assume that $U_1, \ldots, U_k$ are centered real valued random variables. We suppose that $V_1, \ldots, V_k$ are independent random variables, independent of $U_1, \ldots, U_k$ and such that $U_j \sim \mathcal{N}(0, \mathbb{E}U_j^2)$ then for each $g \in C_b^3$ Set $U = U_1 + \cdots + U_k$ and $V = V_1 + \cdots + V_k$ then*

$$|\mathbb{E}(g(U) - g(V))| \leq \sum_{i=1}^{k} \mathbb{E}\left(|U_i|^2 \left(\|g''\|_\infty \wedge (\|g'''\|_\infty |X_i|)\right)\right)$$

Proof of lemma 1.3.2. Set $Z_j = U_1 + \cdots + U_{j-1} + V_{j+1} + \cdots + V_k$ for $1 \leq j \leq k$ then

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

Order 3 Taylor formula (applied twice around Z_j) implies with the fact that terms up to order

$$|\mathbb{E}\delta_j| \leq \frac{\|g'''\|_\infty}{6} \left(|U_j|^3 + |V_j|^3\right)$$

The other term comes from a second order expansion which yields the bound

$$|\delta_j| \leq \frac{\|g''\|_\infty}{2} \left(|U_j|^2 + |V_j|^2\right)$$

Write $\delta_j = \widetilde{\delta}_j + (V_j - U_j)g'(W_j)$ (for some centered additional term). Use again Taylor formula up to orders 2 and 3 implies that $|\delta_k|$ is controled by the minimum of 2 bounds. To conclude note $\mathbb{E}|V_j|^2 = \mathbb{E}|U_j|^2$ and thus $\mathbb{E}|V_j|^3 = \mathbb{E}|Z|^3 \left(\mathbb{E}U_j^2\right)^{3/2}$ for a standard normal rv $Z \sim \mathcal{N}(0,1)$. Hölder inequality thus yields $\left(\mathbb{E}U_j^2\right)^{3/2} \le \mathbb{E}|U_j|^3$. An integration by parts implies $\mathbb{E}|Z|^3 = \frac{4}{\sqrt{2\pi}} < 2$. ∎

A simple consequence of the result is

Theorem 1.3.1 (Lindeberg) *Let $(\zeta_{n,k})_{k\in\mathbb{Z}}$ be iid sequences of centered random variable (for each n). Suppose*

- $\displaystyle\sum_k \mathbb{E}\zeta_{n,k}^2 \to_{n\to\infty} \sigma^2 > 0,$
- $\displaystyle\sum_k \mathbb{E}\zeta_{n,k}^2 1\!\!1_{\{|\zeta_{n,k}|>\epsilon\}} \to_{n\to\infty} 0$, *for each $\epsilon > 0$,*

Then

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

Example 1.3.1 *Some examples of the use for Lindeberg lemma now follow*

- ***CLT***

  Theorem 1.3.2 *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 iid sequences with finite variance.

  Proof. This follows from lemma 1.3.2 with $k = n$ and $U_j = X_j/\sqrt{n}$ (or from theorem 1.3.1). 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 and non decreasing on $\mathbb{R}^+$ ([3])).

- ***Empirical median***
  Suppose $n = 2N+1$ is even we consider an iid $n-$sample $Y_1, \ldots, Y_n$ 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 $N+1$.

[3]For each $k > 0$ there exists $M_k > 0$ that we may choose non decreasingf 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 $M_k \le |x| < M_{k+1}$.

If the cumulative repartition function F of Y_1 admits a derivative γ at point M then

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

Indeed 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 satisfy $Y_i \le M + x/\sqrt{n}$:

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

Set $p_n = \mathbb{P}(Y_1 \le M + x/\sqrt{n})$ et $X_{i,n} = \dfrac{1\!\!1_{\{Y_i \le M+x/\sqrt{n}\}} - p_n}{\sqrt{np_n(1-p_n)}}$ yields

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

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. ∎

- ***Gaussian approximation of binomial laws***
  Let $S_n \sim B(n,p)$ and fix some $\epsilon > 0$ then

$$\sup_{\theta\in]0,1-\epsilon]} \sup_{u\in\mathbb{R}} \Delta_{n,\theta}(u) = \mathcal{O}\Big((n\theta)^{-\frac{1}{8}}\Big)$$

with

$$\Delta_{n,\theta}(u) = \left|\mathbb{P}_\theta\Big(\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \le u\Big) - \Phi(u)\right|$$

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 $n\theta$ is large as a classical heuristiic tells us: in statistics $n\theta \ge 5$ is the condition used to use a Gaussian approximation of binomials.
Use lemma 1.3.2. Rewrite $S_n = b_1 + \cdots + b_n$ with iid $b_1, b_2, \ldots \sim b(\theta)$. Then $X_1, \ldots, X_n$ $X_i = (b_i - \theta)/\sqrt{n\theta(1-\theta)}$ is centered iid and $\mathbb{E}_\theta b_i^3 = \mathbb{E}_\theta b_i = \theta$.
From the convexity of $x \mapsto |x|^3$ we derive $\mathbb{E}_\theta|b_i - \theta|^3 \le 2^2\left(\mathbb{E}_\theta b_i^3 + \theta^3\right) \le 8\theta$.
Let $0 < \theta \le 1-\epsilon$ then for $f \in \mathcal{C}^3$ we get from the lemma 1.3.2, with some $Z \sim \mathcal{N}(0,1)$:

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

In order to conclude one needs to prove as an exercise that for each $\eta > 0$, $u \in \mathbb{R}$ there exists a function $f_{u,\eta} \in \mathcal{C}_b^3$ with $1\!\!1_{[u,\infty[} \le f_{u,\eta} \le 1\!\!1_{[u+\eta,\infty[}$ and $\|f'''_{u,\eta}\|_\infty = \mathcal{O}\left(\eta^{-3}\right)$ ([4]). Using $\mathbb{P}(Z \in [u, u+\eta]) \le \eta/\sqrt{2\pi}$ we then derive

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

Thus

$$\Delta_{n,\theta}(u) \le C\left(\frac{1}{\eta^3\sqrt{n\theta}} + \eta\right),$$

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

To conclude this section we present without proof a very simple dependent version of the Lindeberg lemma.

Lemma 1.3.3 (Dependent Lindeberg) *] 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

$$T(k) = \sum_{j=1}^k \left|Cov(e^{i<t,X_1+\cdots+X_{j-1}>}, e^{i<t,X_j>})\right|.$$

Then

$$\Delta_k \le T(k) + 6\|t\|^{2+\delta} A_k.$$

1.4 Pathwise convergence

Definition 1.4.1 *The sequence X_n converges in probability to X, which we denote*

$$X_n \to^{\mathbb{P}}_{n\to\infty} X$$

if for each $\epsilon > 0$ $\lim_{n\to\infty}\ P(|X_n - X| \ge \epsilon) = 0$.

Assume that convergence in probability holds then lemma 1.3.1 then we may assume that g is uniformly continuous in the definition of convergence in distribution. Let $\epsilon > 0$ we set $A = (|X_n - X| \ge \epsilon)$

$$\begin{aligned}|\mathbb{E}(g(X_n) - g(X))| &= |\mathbb{E}(g(X_n) - g(X))1_A + \mathbb{E}(g(X_n) - g(X))1_{A^c}| \\ &\le 2\|g\|_\infty \mathbb{P}(A_n) + \sup_{|x-y|<\epsilon} |g(x) - g(y)|.\end{aligned}$$

Uniform continuity of g yields convergence in law. □

[4] Choose it such that restriction of $f_{u,\eta}$ to the interval $[u, u+\eta]$ is a polynomial with degree 3. Coefficients are chosen in order that derivatives at the extremities of this interval.

Definition 1.4.2 *If $\mathbb{E}|X_n - X|^p \to_{n\to\infty} 0$ we say that the sequence X_n converges in $\mathbb{L}^p$.*

From Markov inequality applied to $V = |X_n - X|$ this is immediate to derive that $\mathbb{L}_p$ convergence implies convergence in probability..

Definition 1.4.3 *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.
For a sequence of events B_n we set $\overline{\lim}_n B_n = \bigcap_n \bigcup_{k\ge n} B_k$. Note that $A_n = \bigcup_{k\ge n} B_k$ is a decreasing sequence of events.

Lemma 1.4.1 (Borel-Cantelli) *If $(B_n)_{n\in\mathbb{N}}$ is a sequence of events such that $\sum_n \mathbb{P}(B_n) < \infty$ then $\mathbb{P}(\overline{\lim}_n B_n) = 0$.*

We now derive 2 moment inequalities respectively called Marcinkiewicz-Zygmund and Rosenthal moment inequalities.
Later in those notes alternative proofs of those results will be obtained.

Lemma 1.4.2 *Let X_n be a sequence of 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:*

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

And for $p = 2$:

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

Remark 1.4.1 *The second inequality also extends to arbitrary $p \ge 2$ There exists a constant C only depending on p such that*

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

Proof. Simple combinatory aguments yield:

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

Now from centering conditions we see that terms T vanish except for cases when $i_1 = i_2, \dots, i_{2p-1} = i_{2p}$, since else an index i would be isolated and the corresponding term vanishes by using independence. Among $A = \{i_2, i_4, \dots, i_{2p}\}$ which take precisely n^p values one needs to make summations according to 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 just use Hölder inequality to derive the first result. □

Corollary 1.4.1 *If* $\mathbb{E}X_0^4 < \infty$ *then*

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

Proof. quote that $\mathbb{P}(|\overline{X}| \geq \epsilon) \leq C\mathbb{E}X_0^4/(\epsilon n^2)$ is a convergent serie. Hence the a.s. is thus a a consequence of Borel-Cantelli lemma. ∎

Now in case $\mathbb{E}X_0^2 < \infty$ then Markov inequality yields $\mathbb{L}^2-$convergence of $\overline{X}$; indeed Var $(\overline{X})$ = Var $(X_0)/n$, thus convergence in probability also holds.
This allows to prove a first statistical fundamental result:

Theorem 1.4.1 *Let* (Y_n) *be a real valued and iid sequence such that* Y_0 *admits cumulative distribution function* $F(y) = \mathbb{P}(Y_0 \leq y)$ *on* $\mathbb{R}$. *Set*

$$F_n(y) = \frac{1}{n}\sum_{j=1}^n 1\!\!I_{\{Y_j \leq y\}}$$

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

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

Proof. The previous strong law of large (first result in Corollary 1.4.1) implies the convergence. Uniform convergence is implied by a variant of Dini theorem and is left as an exercise. It relies of the fact that function F_n are non-decreasing and that F is continuous. ∎

1.5 Kernel density estimation

We now introduce another standard tool of statistics. When a density is to be estimated, unfortunately the simple plug-in technique consisting to derive an estimator to fit a derivative does not work since F_n's derivative is 0 (a.s.).

Definition 1.5.1 *Let* (Y_n) *be a real valued and iid sequence such that* Y_0 *admits a density* f *on* $\mathbb{R}$. *If* $K : \mathbb{R} \to \mathbb{R}$ *denotes a function such that:*

$$\int_{\mathbb{R}} (|K(y)| + |K(y)|^2)dy < \infty, \qquad \int_{\mathbb{R}} K(y)dy = 1$$

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

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

A first result allows to bound the bias of such estimators

Lemma 1.5.1 *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 $\leq p$ such that $K = Pg$ is a kernel satisfying*

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

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 $A = (a_{i+j})_{0\leq i,j\leq n}$ with $a_k = \int_{\mathbb{R}} y^k g(y)dy$ with order $(p+1)$ is symmetric positive definite. Indeed if $x = (x_0, \ldots, x_p)' \in \mathbb{R}^{p+1}$

$$x'Ax = \int_{\mathbb{R}} \left(\sum_{j=0}^{p} x_j y^j\right)^2 g(y)dy \geq 0$$

If the previous expression vanishes the fact that $g \neq 0$ 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. ∎

A change of variable that

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

A simple application of Taylor formula is left as an exercise and proves

Proposition 1.5.1 *Assume that $h \to_{n\to\infty} 0$.*
Assume that the function f admits p continuous and bounded derivatives, then if K is a $p-$order kernel:

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

Remark 1.5.1 *Independence of (Y_k) is not necessary here, but only the fact that each of the Y_k $1 \leq k \leq n$) admits the same distribution.*
Uniformity over $\mathbb{R}$ may be suppressed. If K admits a compact support, then

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

for a neighborhood V of y.

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

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

Then:

Theorem 1.5.1 *Assume that $nh \to_{n\to\infty} \infty$:*

$$Var\,\widehat{f}(y) \sim_{n\to\infty} \frac{1}{nh} f(y) \int_{\mathbb{R}} K^2(u)du$$

$$\sqrt{nh}(\widehat{f}(y) - \mathbb{E}\widehat{f}(y)) \to^{\mathcal{L}}_{n\to\infty} \mathcal{N}(0, f(y) \int_{\mathbb{R}} K^2(u)du)$$

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

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

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 minimising this expression wrt $h = h_n$ or by equating the squred bias and variance of the estimator.
Now if $nh \to \infty$ using Rosenthal moment inequalities from remark 1.4.1 implies

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

This with Markov inequality and Borel-Cantelli implies the result

Proposition 1.5.2 *If f is continuous then*

$$\widehat{f}(y) \to f(y), \qquad a.s.$$

if

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

A.s. uniform convergence may also be derived. 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.

Chapter 2

Stationarity

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 probability space.

2.1 Stationarity

Definition 2.1.1 *A random sequence $(X_n)_{n\in\mathbb{Z}}$ is stationary if, for each $k \geq 0$, the law of the vector $(X_l, \ldots, X_{l+k})$ does not depend on $l \in \mathbb{Z}$.*
It 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 ofX_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 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 Gaussianity we will see that both notions coincide. Anyway this is not true in general.
We consider an iid sequence $(\xi_n)_{n\in\in\mathbb{Z}}$:

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

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

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

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

- a simple modification of this example yields a second order stationary 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 $\mathbb{E}X_nX_m = 0$ or 1 if $n \neq m$ or $n = m$. Non stationarity relies on the calculation of $\mathbb{E}X_nX_{n-1}X_{n-2}$ which depends on n.

- 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 $c_n^2 + s_n^2 = 1$). Then this sequence is always centered and orthogonal if $\mathbb{E}V_n^2 < \infty$. Also using independence $\mathbb{E}X_n^2X_{n-1} = \mathbb{E}V_nV_{n-1}^2\xi_{n-1}^2$. 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 proposed as an exercise.

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

$$\mathbb{E}X_n^4 = a\mathbb{E}V_n^4 = a\mathbb{E}(c_n\xi_{n-1}^2+s_n\xi_{n-2}^2)^2 = a(a(c_n^2+s_n^2)+2s_nc_n) = a(a+2s_nc_n)$$

  is not a constant in general.

2.2 Spectral representation

Quote the following property of covariances:
let $c_l \in \mathbb{C}$ for all $|l| \leq n$, then setting $c = (c_l)_{|l|\leq n}$ and $\Sigma_n = (r_{|i-j|})_{|i|,|j|\leq n}$ we obtain:

$$c^t\Sigma_n\overline{c} = \sum_{|i|,|j|\leq n} c_i\overline{c_j}r_{|i-j|} = \mathbb{E}\left|\sum_{|i|\leq n} c_iX_i\right|^2 \geq 0 \qquad (2.1)$$

Theorem 2.2.1 (Herglotz) *If a sequence $(r_n)_{n\in\mathbb{Z}}$ satisfies (2.1) then there exists a non-decreasing function G (essentiallyt 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 $\mu([-\pi,\lambda]) = G(\lambda)$, $\forall\lambda \in [-\pi,\pi]$.
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 (2.1) 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

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

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

Definition 2.2.1 *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 2.2.1
- *For an orthogonal sequence (i.e $\mathbb{E}X_kX_l = 0$ pour $k \neq l$) 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 has 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 by the relation*

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

[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 $G_{\phi(n)} = G_{\phi_n}(n)$.

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 centtered and iid 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 identifie g and the associated nonnegative measure.

Definition 2.2.2 (Random measure) *A random measure is defined with a random function $\Omega \times [-\pi, \pi] \to \mathbb{C}$, $(\omega, \lambda) \mapsto Z(\omega, \lambda)$, 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

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

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

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

Notice that

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

- 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 2.2.2 *A natural example of such a random measure is the Brownian measure. Namely we denote $W([a,b]) = W(b) - W(a)$ then such a 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.*

Theorem 2.2.2 (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.

Some relevant examples are reported as Examples 2.3.1.
Proof. The spectral function G of the process X_n is nondecreasing, hence its discontinuities are at most denumerable set denoted D_G. If $I = (a,b)$ is an interval with $a, b \notin D_G$, set

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

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

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

Note

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

(the truncated Fourier series of the indicator function $1\!\!\mathrm{I}_I$.).
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.
A limit allows to consider extremities in D_G.
To conclude, quote that

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

Hence,

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

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

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

Some relevant examples are reported as Examples 2.3.1.

2.3 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 function

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

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

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

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

Definition 2.3.1 *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 g is uniformly continuous and

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

Example 2.3.1

- *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 $g(\lambda) = \frac{\sigma^2}{2\pi}$, the sequence $\xi_n = \int_{-\pi}^{\pi} e^{in\lambda} Z(d\lambda)$ is a second order white noisewith variance σ^2: $\mathbb{E}\xi_n\xi_m = 0$ or σ^2 according that $n \neq m$ 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 iid 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 spectral density g_X of X writes

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

For this 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)),

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

are such that

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

g *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 2.3.1 *Let* (X_n) *be a centered second order stationary sequence and*

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

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

$$\begin{aligned} g_Y(\lambda) &= \left|\sum_{k=-\infty}^{\infty} c_k e^{ik\lambda}\right|^2 g_X(\lambda) \\ Z_Y(d\lambda) &= \left(\sum_{k=-\infty}^{\infty} c_k e^{ik\lambda}\right) Z_X(d\lambda) \end{aligned}$$

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. ∎

This definition of the range of a process is justified if X_n is centered :

$$\begin{aligned}\mathbb{E}\left|X_1+\cdots+X_n\right|^2 &= \sum_{s=1}^{n}\sum_{t=1}^{n}\mathbb{E}X_sX_t\\ &= \sum_{s=1}^{n}\sum_{t=1}^{n}r_{t-s}\\ &= \sum_{|k|<n}^{n}(n-|k|)r_k\end{aligned}$$

Hence

Proposition 2.3.2 *If X_n is SRD then*

$$\mathbb{E}\left|X_1+\cdots+X_n\right|^2 \sim ng(0)$$

Proof (This is a variant of Cesaro lemma).
This will be enough to proof 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

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

In fact the second order stationary processes write as infinite order moving average of a weak white noise:

Theorem 2.3.1 (Crámer Wold) *Let $(X_n)_{n\in\mathbb{Z}}$ be a second ordered stationary sequencewith 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

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

Chapter 3

Gaussian chaos

Gaussian play a special role in the field of probability since they appear as limit distributions from the CLT.
We develop here some aspects of the Gaussian chaos. A nice feature is that for such models all calculations seem to be possible.

3.1 Gaussian laws

3.1.1 Normal distribution

A standard normal random $N \sim \mathcal{N}(0,1)$ admits a density

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

wrt Lebesgue measure on $\mathbb{R}$. The norming facteur $\sqrt{2\pi}$ may be checked through the computation of a square as follows:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx\right)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\, dx\, dy \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}}\, r\, dr \\ &= 2\pi,\end{aligned}$$

For this use a change in variables with polar coordinates $(r,\theta) \mapsto (x,y) = (r\cos\theta, r\sin\theta)$, $\mathbb{R}^+ \times [0, 2\pi[$ dans $\mathbb{R}^2$. This is easy to check that the Jacobian is simply r in this case.

Lemma 3.1.1 *The characteristic function of this normal distribution writes*

$$\phi_N(s) = \mathbb{E}e^{isN} = e^{-\frac{s^2}{2}} \tag{3.1}$$

Proof. Indeed the Laplace $L_N(z) = \mathbb{E}e^{zN}$ 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

$$\phi_N(s) = L_N(is) = e^{-\frac{s^2}{2}}.$$ ∎

Equation (3.1) may also be rewritten:

$$\mathbb{E}e^{zN - \frac{z^2}{2}} = 1, \qquad \forall z \in \mathbb{C} \tag{3.2}$$

From the analycity of ϕ_N the distribution of a Normal rv is given from its characteristic function.

Definition 3.1.1 *A rv admits the law $Y \sim \mathcal{N}(m, \sigma^2)$ if it may be written $Y = m + \sigma N$ for $m, \sigma \in \mathbb{R}$ and for a Normal rv N.*

Density and characteristic functions of such distributions are derived from linear changes in variable:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-m)^2}{2\sigma^2}}, \qquad \phi_Y(t) = e^{itm} e^{-\frac{1}{2}\sigma^2 t^2}$$

A important property is that if rvs $Y_j \sim \mathcal{N}(m_j, \sigma_j^2)$ are independent for $j = 1, 2$, then $Y_1 + Y_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.
A converse of this result is that if Y_1, Y_2 are independent and with a same distribution μ Ifi $(Y_1 + Y_2)/\sqrt{2}\mu\mu$ then μ is a centered Gaussian distribution. This property follows from a property of characteristic functions. The characteristic function $\gamma(t) = \int e^{tx} \mu(dx)$ satisfies $\gamma(t) = \gamma^2(t/\sqrt{2})$ from independence. To prove that this characterizes Gaussians, it may be proved that the log-characteristic function is a second degree polynomial. With this formula recursion entails that $\log \gamma(t) = at^2$ for $t = k2^n$ with $k, n \in \mathbb{Z}$; continuity allows to conclude.

3.1.2 Multivariate Gaussians

Definition 3.1.2 *A random vector $Y \in \mathbb{Z}^k$ is Gaussien 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.

A Gaussian law only depends on its second order properties
Law of Gaussian Y rvs depend only on the mean and on the variance matrix. For $u \in \mathbb{R}^k$, $\Sigma = \mathbb{E}(Y-\mathbb{E}Y)(Y-\mathbb{E}Y)'$, we see 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 for 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 ([1]), 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 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). This solution may be proved to be unique in case Σ is definite, because eigen-spaces 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 iid thus $Z \sim \mathcal{N}_k(0, I_k)$.

Density.
For a change in variables if Σ is invertible, Y admits a density on $\mathbb{R}^k$:

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} e^{-\frac{1}{2}(y-\mathbb{E}Y)^t \Sigma^{-1}(y-\mathbb{E}Y)} \tag{3.3}$$

Characteristic function.
Even for Σ non invertible we may write $Y = \mathbb{E}Y + RZ$. Thus for each $s \in \mathbb{R}^k$:

$$\begin{aligned} \phi_Y(s) &= \mathbb{E}e^{is\cdot Y} \\ &= e^{is\cdot \mathbb{E}Y}\mathbb{E}e^{is\cdot RZ} \\ &= e^{it\cdot \mathbb{E}Y}\mathbb{E}e^{iZ\cdot Rs} \\ &= e^{is\cdot \mathbb{E}Y - \frac{1}{2}(Rs)\cdot(Rs)} \\ \phi_Y(s) &= e^{is\cdot \mathbb{E}Y - \frac{1}{2}(s^t\Sigma s)} \qquad (3.4) \end{aligned}$$

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

[1]Indeed $u'\Sigma u = \mathrm{Var}\,(Y \cdot u) \geq 0$ for each $u \in \mathbb{R}^k$,

a rectangular matrix C with order $a \times b$.Then $Z = Y - C'X$ est is orthogonal to X, from Gaussianness of this vector they are independent. Thus

$$\mathbb{E}(Y|X) = C'X.$$

3.1.3 Existence of Gaussian laws

If Σ is a $d \times d$ symmetric positive definite, we proved the existence of a symmetric positive definite matrix R with $R^2 = \Sigma$. For $Z = (Z_1, \ldots, Z_d)^t$ iid standard normal rvsand for each $m \in \mathbb{R}^d$:

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

As an application:

Proposition 3.1.1 *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. 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 (2.1) 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.

3.1.4 Fractional Brownian motion

The fractional Brownian motion (fBm see Taqqu, 2002 [10]) with exponent $H \in]0, 1]$ is a centered Gaussian process $(Z_t)_{t \in \mathbb{R}}$ with covariance $\Gamma(s, t) = \mathrm{Cov}\,(Z_s, Z_t)$ définie par

$$\Gamma(s, t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s - t|^{2H}) \tag{3.5}$$

Proposition 3.1.2 *n Γ in (3.5) pour $s, t \in \mathbb{R}$ is indeed the covariance of a centered Gaussian process $(B_H(t))_{t \in [0,1]}$.*

Proof. From proposition 3.1.1 we need to prove that for all $0 \leq t_1 < \cdots < t_n \leq 1$, and $u_1, \ldots, u_n \in \mathbb{C}$

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

- *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

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

Remarque. 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 $\mathbb{R}^+$ with $\Gamma(s,t) = s \wedge t$.

Lemma 3.1.2 *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 and of the calculation

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

Definition 3.1.3 *The process $(Z(t))_{t\in\mathbb{R}}$ is $H-$selfsimilar if for all $a>0$*

$$(Z(at))_{t\in\mathbb{R}} \overset{en\ loi}{=} (a^H Z(t))_{t\in\mathbb{R}}$$

This condition is equivalent to the stationarity of the proces $Y(t)=e^{-tH}Z(e^t)$ when it is indexed by $\mathbb{R}^+$. For this only check that finite dimensional repartitions of both processes coincide de deux processus.
Also quate that:

1. If Z is selfsimilar then $Z(0)=0$.
2. If Z is selfsimilar and its increments $(Z(t+s)-Z(t))_{t\in\mathbb{R}}$ are stationary for each s then: $\mathbb{E}Z(t)=0$ lorsque $H\neq 1$ because $\mathbb{E}Z(2t)=2^H\mathbb{E}Z(t)$ and $\mathbb{E}(Z(2t)-Z(t))=\mathbb{E}(Z(t)-Z(0))=\mathbb{E}Z(t)$ thus $(2^H-2)\mathbb{E}Z(t)=0$.
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 selfsimilarity: $\mathbb{E}Z^2(t)=|t|^{2H}$.
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 3.1.3 *B_H is Gaussian centered and $H-$selfsimilar with stationary increments.*

3.2 Gaussian chaos

A Gaussian family is a set such that each finite subset defines a Gaussian random vector (alternatively, if $(u_t)_{t\in\mathbb{T}}$ is such that $u_t\equiv 0$ excepted for a $t\in T'$ (a finite subset) then $\sum_{t\in\mathbb{T}} u_t X_t$ is a Gaussian random variable).
Let $X=(X_t)_{t\in\mathbb{T}}$ be a Gaussian family defined on some probability space $(\Omega,\mathcal{A},\mathbb{P})$, the Gaussian chos associated to X is the smallest complete vector sub-space $\mathbb{L}^2(\Omega,\mathcal{A},\mathbb{P})$ containing X_t $\forall t\in\mathbb{T}$), the constant 1, and which is stable under products (in more quick terms it is the adherence in $\mathbb{L}^2(\Omega,\mathcal{A},\mathbb{P})$ of the algebrac generated by X); its elements are thus $\mathbb{L}^2-$limits of polynomials

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

for some finite subset $T'\subset T$.
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 degree moments and higher order moments respectively.

3.2.1 Hermite polynomials

Definition 3.2.1 (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.

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

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \end{aligned}$$

Polynomial H_k is an orthogonal system wrt 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 $k = l$ it yields

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

This system is also complete (we admit this result proved eg. in [5]. Hence any measurable function g with $\mathbb{E}|g(N)|^2 < \infty$ admits the $\mathbb{L}^2$ representation

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

Definition 3.2.2 *Denote m or $m(g)$ this Hermite rank of a function g: this is the smallest index $k \geq 0$ with $g_k \neq 0$. .*

This orthonormal basis in $\mathbb{L}^2(\varphi(x)dx)$ also satisfies

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} H_k(x) = e^{zx - z^2/2} \tag{3.6}$$

The previous series converges (normaly) in $\mathbb{L}^2(\varphi(x)dx)$ because

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

We prove the identity $H_k' = kH_{k-1}$. As $d^\circ(H_k' - kH_{k-1}) < k-1$, this will follow from $\int (H_k'(x) - kH_{k-1}(x))H_l(x)\varphi(x)dx = 0$ for all $l < k$.
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 that

$$H_k' = kH_{k-1}.$$

Remark 3.2.1 *An alternative and more elementary proof of the previous relation begins with the relation $\varphi'(x) = x\varphi(x)$. 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 when 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 $g_z = \sum_k g_{z,k}H_k/k!$:

$$\begin{aligned} g_{z,k} &= \mathbb{E}g_z(N)H_k(N) \\ &= \int_{-\infty}^{\infty} H_k(x)e^{zx - z^2/2}\varphi(x)dx \\ &= \int_{-\infty}^{\infty} H_k(x)e^{-(z-x)^2/2}\frac{dx}{\sqrt{2\pi}} \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} H_k(t+z)\varphi(t)dt, \quad \text{from the change in variable } t = x - z \\
&= \sum_{l=0}^{k} \frac{z^l}{l!} \int_{-\infty}^{\infty} H_k^{(l)}(t)\varphi(t)dt, \quad \text{with a Taylor expansion} \\
&= \sum_{l=0}^{k} C_k^l z^l \int_{-\infty}^{\infty} H_{k-l}(t)\varphi(t)dt, \quad \text{because } H_k^{(l)} = \frac{k!}{(k-l)!} H_{k-l} \\
&= z^k
\end{aligned}$$

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

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} H_k(N) = e^{zN - z^2/2} \qquad \text{in} \qquad \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P}). \tag{3.7}$$

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 3.2.2 *If one knows how to prove that this series $x \mapsto g(x,z)$ converges and is derivable for each z, we would deduce $\partial g/\partial x(x,z) = zg(x,z)$ 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 (3.6) for all $x \in \mathbb{R}$, $z \in \mathbb{C}$. A alternative proof of (3.7) would follow.

3.2.2 Second order moments

Lemma 3.2.1 (Mehler formula) *Let $Y = (Y_1, Y_2)$ be a Gaussian random vector with law $\mathcal{N}_2\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right)$ a normalized random Gaussian vector, 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 $\sigma^2 = \mathrm{Var}\,(t_1 Y_1 + t_2 Y_2) = t_1^2 + t_2^2 + 2rt_1t_2$, then$t_1Y_1 + t_2Y_2 \sim \sigma N$; relation (3.2) implies

$$\mathbb{E}e^{t_1Y_1 + t_2Y_2 - \frac{t_1^2 + t_2^2}{2}} = e^{rt_1t_2}$$

From the $\mathbb{L}^2-$identity (3.7) 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 wrt to powers of t_1 et t_2 yields the conclusion since $\mathbb{E}H_k(Y_1) \neq 0$ only for the case $k = 0$. ∎

Remark 3.2.3 *Let $g : \mathbb{R} \to \mathbb{C}$ be measurable and $\mathbb{E}|g(N)|^2 < \infty$. Thus $g = \sum_k \frac{g_k}{k!} H_k$ with $g_k = \mathbb{E}H_k(N)\overline{g(N)}$ 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 $\mathbb{E}g(Y_0) = 0$ (for Hermite it means $m(g) \geq 1$):

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

Thus in case $\sum_l |r_l| < \infty$, then each series $R_k = \sum_l r_l^k$ converges (for $k \geq 1$) 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 $\sum_l |r_l|^{m(g)} < \infty$ then the previous claim still holds true.

Example 3.2.1 *For statistics the empirical cumulative function is of a first importance $F_n(x) = \frac{1}{n}\sum_{k=1}^n 1\!\!1_{\{Y_k \leq x\}}$. This is an unbiased estimator of the cumulative function $\mathbb{E}F_n(x) = F(x)$. The expression of its variance relies on the previous identity for $g(u) = 1\!\!1_{\{u\leq x\}}$. Here*

$$\begin{aligned} g_k &= \mathbb{E}H_k(N) 1\!\!1_{\{N \leq x\}} \\ &= \int_{-\infty}^x H_k(u)\varphi(u)du \end{aligned}$$

$$= (-1)^k \int_{-\infty}^{x} \varphi^{(k)}(u)du$$
$$= \begin{cases} \Phi(x), & (a\ primitive\ of\ \varphi) \quad for \quad k=0 \\ \varphi(x)H_{k-1}(x), & \qquad if \quad k \neq 0 \end{cases}$$

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 $\mathcal{O}(\frac{1}{n})$ *as* $n \to \infty$ *if* $\sum_l |r_l| < \infty$*. If now* $\sum_l |r_l| = \infty$ *its order of magnitude is* $\mathcal{O}(\frac{1}{n}\sum_{|l|<n}\left(1 - \frac{|l|}{n}\right) r_l$ *which is more that* $\frac{1}{n}$*. Anyway this expression converges 0 if the sequence* r_l *does converge to 0.*

3.2.3 Higher order moments

The technique used to derive Mehler formula suggest an extension for an arbitrary number of factors $H_{l_j}(Y_j)$.
Thus let $Y = (Y_1, \dots, Y_p) \sim \mathcal{N}_p(0, R)$ for a symmetric matrix $R = (r_{i,j})_{1\le i,j\le p}$ with $r_{i,i} = 1$. If $(t_1, \dots, t_p) \in \mathbb{R}^p$ 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 (3.2) proves

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

in case an $\mathbb{L}^p-$analogue of (3.7)exists then

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

Any argument allowing inversion of sums and integrals would allow the identification of such moments.
Anyways such convergences are not simple and our technique to derive such moments will rely on an argument due to Slepian (1972) [25]. Use $\phi_Y(s) = e^{-\frac{1}{2}s^t\Sigma s}$ the characteristic function of a centered Gaussian rv $Y = (Y_1, \dots, Y_k)$ with covariance Σ 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 $r^n=\prod_{i<j} r_{i,j}^{n_{i,j}}$ and $n! = \prod_{i<j} n_{i,j}!$. Also set $n_{i,j}=n_{j,i}$ if $i>j$ and $s_{n,i}=\sum_{j\neq i} n_{i,j}$ then denoting by $f(y,I_k)=\prod_{i=1}^k \varphi(y_i)$ we get

$$\begin{aligned} f(y,\Sigma) &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \frac{\partial^{\left(\sum_{i<j} n_{i,j}\right)} f(y,I_k)}{\prod_{i<j} \partial r_{i,j}^{n_{i,j}}} \\ &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \frac{\partial^{s_{n,i}} f(y,I_k)}{\prod_{i<j} \partial y_i^{n_{i,j}} \partial y_j^{n_{i,j}}} \\ &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^k \frac{\partial^{s_{n,i}} \varphi(y_i)}{\partial y_i^{s_{n,i}}} \\ &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^k \varphi^{(s_{n,i})}(y_i) \\ f(y,\Sigma) &= \sum_{n=(n_{i,j})} \frac{r^n}{n!} \prod_{i=1}^k H_{s_{n,i}}(y_i)\cdot\phi(y) \qquad (3.8) \end{aligned}$$

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 (3.8) 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 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 multi-indices n with $s_{n,i}=s_i$ pour $1\le i\le k$.
If $k=2$ the sum in $n$is a simple sum on integers $\mathbb{N}$ because $i<j$ i implies $i=1$ and $j=2$. Thus $\sum_{n\in N(s_1,s_2)}$ corresponds the only value $n_{1,2}=s_1=s_2$: it is again Mehler formula.
For $k\ge 2$ the previous formul is called the diagram formula. The $n_{i,j}$s correspond to partitions of the array such that

x_1	$\dots$	x_1	appears s_1 times
x_2	$\dots$	x_2	rappears s_2 times
$\dots$	$\dots$	$\dots$	$\dots$
x_k	$\dots$	x_k	appears s_k ftimes

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 3.2.4 *Numerous uses of this formula are known. Breuer and Major (1983) (*[2]*) prove*

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

if $\sum_k |r_k|^m < \infty$ *and* m *denotes the Hermite 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 in Taqqu (2002) [10]. That inequality is extended in [26] and further by Bardet and Surgailis . Other developpments are also reported in the book by Rosenblatt (1985).
Finally the method of diagrams is now less used because of the recents developments by Peccati and coauthors see eg. [21].

[2] P. Breuer and P. Major (1983) "Central limit theorems for nonlinear functionals of Gaussian fields". Journal of Multivariate Analysis 13, pages $425 - 441$.

Chapter 4

Linear processes

Stationary sequences generated through iid $(\xi_n)_{n\in\mathbb{Z}}$ are considered hereafter. Such models are natural in signal theory since they appear through linear filtering of a white noise
Let $(c_n)_{n\in\mathbb{Z}}$ a sequence of real numbers, we set

$$X_n = \sum_{k=-\infty}^{\infty} c_k \xi_{n-k}$$

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 $\mathbb{E}|\xi_0| < \infty$ $(m=1)$ this series converges from Markov inequality

$$\mathbb{P}\left(\sum_k |c_{n-k}||\xi_k| > A\right) \leq \frac{1}{A^m}\mathbb{E}\left(\sum_k |c_k||\xi_{n-k}|\right)^m \leq \frac{1}{A^m}\mathbb{E}|\xi_|^m \sum_k |c_k|^m |n-k|$$

(use relation $(a+b)^m \leq a^m + b^m$).
If $\mathbb{E}\xi_0^2 < \infty$ $(m=2)$ et $\mathbb{E}\xi_0 = 0$, the weaker condition for statiionarity and existence in $\mathbb{L}^2$ holds

$$\sum_k |c_k|^2 < \infty$$

If $c_k = 0$ for $k < 0$ the processus is *causal*.
This process admits the covariance:

$$r_k = \operatorname{Cov}(X_0, X_k) = \sum_l c_l c_{l+k} = c \star \widetilde{c}_k$$

denoting $\widetilde{c} = (\widetilde{c}_k)_{k\in\mathbb{Z}}$ with $\widetilde{c}_k = c_{-k}$.
Remark

$$\sum_k |r_k| \leq \left(\sum_k |c_k|\right)^2$$

thus this series converges at the same time as $\sum_k |c_k|$.

4.1 FARIMA$(0, d, 0)-$processes

Set $\Delta = I - B$ with B the Backward operator defined as $BX_t = X_{t-1}$ 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 $X_t = \xi_1 + \cdots + \xi_t$, 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 $\Delta^d X_t = \xi_t$ pour $d \in \mathbb{N}$.
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

$$(1 - z)^{-d} = \sum_{j=0}^{\infty} b_j z^j, \quad b_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} = \frac{1}{\Gamma(d)} \prod_{k=1}^{j} \frac{k - 1 + d}{k}$$

Stirling formula

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \qquad n \to \infty$$

implies indeed $b_j \sim \frac{1}{\Gamma(d)} j^{d-1}$ as $j \to \infty$.
Hencee $\sum_j b_j^2 < \infty$ si $d < \frac{1}{2}$ and the series $X_t = \sum_{j=0}^{\infty} b_j \xi_{t-j}$ converge in $\mathbb{L}^2$.
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

$$\Delta^{d-1} X_t = \xi_t \Rightarrow \Delta^d X_t = \Delta\xi_t = \xi_t - \xi_{t-1}$$

It may be proved that

$$\begin{aligned} r(0) &= \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)}, \qquad \sigma^2 = \mathbb{E}\xi_0^2 \\ r(k) &= \sigma^2 \frac{\Gamma(k + d)\Gamma(1 - 2d)}{\Gamma(k - d + 1)} \sim \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma(d)\Gamma(1 - d)} |k|^{2d-1} \quad (|k| \to \infty) \end{aligned}$$

Thus $\sum_k |r(k)| = \infty$ if and only if $d \in]0, \frac{1}{2}[$ (set $H = d + \frac{1}{2}$).
Set Z for the random spectral measure associated to the white noise ξ_t.
Then

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}(1 - e^{-i\lambda})^{-d} Z(d\lambda)$$

and

$$g_X(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{-i\lambda}|^{-2d} = \frac{\sigma^2}{2\pi}\left(4\sin^2\frac{\lambda}{2}\right)^{-2d}.$$

4.2 ARMA(p,q) processes

Relation

$$X_t - \sum_{j=1}^{p} a_j X_{t-j} = \xi_t - \sum_{k=1}^{q} b_k \xi_{t-k}$$

is written

$$\alpha(B)X_t = \beta(B)\xi_t$$

and

$$X_t = \sum_{j=0}^{\infty} c_j \xi_{t-j}$$

with

$$\sum_{j=0}^{\infty} \frac{\beta(z)}{\alpha(z)}$$

and

$$\alpha(z) = 1 - a_1 z - \cdots - a_p z^p = \left(1 - \frac{z}{r_1}\right) \cdots \left(1 - \frac{z}{r_p}\right)$$

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 on the 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 $|c_k| \leq Ce^{-\gamma k}$.
Thus $|r_k| \leq c\rho^k$ for some $0 \leq \rho < 1$.

4.3 FARIMA(p,d,q) processes

Those models fit the equation

$$\alpha(B)(I-B)^d X_t = \beta(B)\xi_t$$

If $d < \frac{1}{2}$ the processus is causal and well defined in case the roots of α are not inside the unit disk. It is inversible if $d > -\frac{1}{2}$ and the roots of α are out of the unit disk. Indeed in this case $\xi_t = \gamma(B)X_t$ pfor 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

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}(1-e^{-i\lambda})^{-d} \frac{\beta(e^{i\lambda})}{\alpha(e^{i\lambda})} Z(d\lambda)$$

Thus

$$g_X(\lambda) = \frac{\sigma^2}{2\pi} |1-e^{-i\lambda}|^{-2d} \left| \frac{\beta(e^{i\lambda})}{\alpha(e^{i\lambda})} \right|^2.$$

Clearly any meromorphic function $\gamma : \mathbb{C} \to \mathbb{C}$ without singularities on $D(0,1)$ and with finitely many singularities on the unit circle allows to define a process

$$X_t = \gamma(B)\xi_t$$

In case $1/\gamma$ satisfies the same assumptions then the process is reversible (zeros remplace singularities). Singularities $\neq 1$ on the unit circle cercle are called periodic long range singularities.

Chapter 5

Non linear processes

This chapter aims at describing stationary sequences generated from iid samples $(\xi_n)_{n\in\mathbb{Z}}$. Many usual models of statistics are seen to be written this way. This organization follows the order from natural extensions of linearity to more general settings.

5.1 Discret chaos

This section aims at introducing some basing tools for algebraic extensions of linear to polynomial models.

5.1.1 Volterra expansions

Set

$$X_n = \sum_k X_n^{(k)},$$

with $X_n^{(0)} = c^{(0)}$ some constant and consider arrays $(c_j^{(k)})_{j\in\mathbb{Z}^k}$ of constants and a sequence of iid random variables $\left(\left(\xi_n^{(k,j)}\right)_{1\le j\le k}\right)_{n\in\mathbb{Z}}$ such that

$$X_n^{(k)} = \sum_{j_1<j_2<\cdots<j_k} c^{(k)}_{j_1,\dots,j_k}\xi^{(k,1)}_{n-j_1}\cdots\xi^{(k,k)}_{n-j_k}$$

Example 5.1.1 *In order to understand why the previous definition involves arrays of iid rvs $((\xi_n^{(k,j)})_{1\le j\le k})_{n\in\mathbb{Z}}$, 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 chaoses 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.

Suppose (without loss of generality) that $\mathbb{E}\left|\xi_n^{(k,j)}\right|^2=1$ then

$$\begin{aligned} \mathbb{E}X_0^{(k)}X_n^{(l)} &= 0, \qquad \text{si } k\neq l \\ \mathbb{E}X_0^{(k)}X_0^{(k)} &= \sum_{j_1<j_2<\cdots<j_k}\left|c_{j_1,\dots,j_k}^{(k)}\right|^2, \\ \mathbb{E}X_0^{(k)}X_n^{(k)} &= \sum_{j_1<j_2<\cdots<j_k} c_{j_1,\dots,j_k}^{(k)}c_{n+j_1,\dots,n+j_k}^{(k)}. \end{aligned}$$

Those calculations yield explicit expressions for the covariance of the process $(X_n)_{n\in\mathbb{Z}}$.

5.1.2 Appell polynomials

As for the special case of the Gaussian laws (whields the construction of Hermite chaos) one mais 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 $A_0,\dots,A_m$ are defined recurrsively with $A_0(x)=1$ and

$$A_k'(x)=kA_{k-1}(x), \qquad \sum_{j=0}^k \mathbb{E}\xi_0^j A_j(0)=0, \qquad 1\le k\le m$$

Hence

$$\begin{aligned} A_0(x) &= 1 \\ A_1(x) &= x-\mathbb{E}\xi_0 \\ A_2(x) &= x^2-2\mathbb{E}\xi_0 x+2(E\xi_0)^2-\mathbb{E}\xi_0^2 \\ \cdots &\ \cdots\ \cdots \\ A_k(x) &= x^k+\cdots \end{aligned}$$

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).
Here

$$c_k = \mathbb{E}P^{(k)}(\xi_0) = (-1)^k \int_{-\infty}^{\infty} P(x) f^{(k)}(x) dx$$

lf 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.
An important property of those Appell polynomials writes

$$\mathbb{E}A_k(\xi_0)P(\xi_0) = 0, \qquad \text{if} \qquad d^\circ P < k.$$

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 integrations by parts alow to conclude.
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 5.1.1 *Moreover Kazmin (1969) proves that if the function $x \mapsto A(z) = 1/\mathbb{E}e^{z\xi}$ is analytic and does not vanish on the disk $\{z \in \mathbb{C}/|z| < \sigma\}$ then each function $g \in E_\tau$ (set of analytic function on a disk $\{z \in \mathbb{C}/|z| < \tau\}$) admits a representation*

$$g(z) = \sum_{n=0}^{\infty} \frac{c_n}{n!} A_n(z), \qquad \lim\sup_{n\to\infty} |c_n|^{1/n} < \tau$$

for series which converge uniformly over compact subsets of the disk $\{z \in \mathbb{C}/|z| < \tau\}$.
Conversely for a sequence such that $\limsup_{n\to\infty} |c_n|^{1/n} < \tau$, the function g defined this way is est analytic on $\{z \in \mathbb{C}/|z| < \tau\}$.
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.

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

$$\begin{aligned} \frac{\partial}{\partial x_i} A_{n_1,\ldots,n_k}(x_1, \ldots, x_k) &= n_i A_{n_1,\ldots,n_k}(x_1, \ldots, x_k), \qquad 1 \le i \le k \\ \mathbb{E}A_{n_1,\ldots,n_k}(\xi) &= 1 \text{ lorsque } n_1 + \cdots + n_k = 0 \text{ et } 0 \text{ sinon.} \end{aligned}$$

[1] The proof is quite analogue to those for the Gaussian chaos.

If random variables $\xi_1, \ldots, \xi_k$ are inpendent and admit respective distributions $\nu_1, \ldots, \nu_k$ then

$$A_{n_1,\ldots,n_k}(x_1, \ldots, x_k) = A_{n_1}^{(\nu_1)}(x_1) \cdots A_{n_k}^{(\nu_k)}(x_k)$$

5.2 Chaotic memory models

The section aims at considering some models whose chaotic expansions may be explicitely determined.

5.2.1 Bilinear models

For simplicity we first consider the Markov model

$$X_n = (a + b\xi_n)X_{n-1} + \xi_n, \qquad \alpha = \mathbb{E}|a + b\xi_0| < 1$$

The stationary solutionof this Markov recursion writes

$$X_n = \sum_{k=0}^{\infty} \xi_{n-k} \prod_{j=1}^{k} (a + b\xi_{n-j}).$$

for such models a recursion is also available for the sequence of covariances.
One variant for this model writes

$$X_n = h(\xi_n)X_{n-1} + \xi_n$$

and a stationary solution still writes

$$X_n = \sum_{k=0}^{\infty} \xi_{n-k} \prod_{j<k} h(\xi_{n-j})$$

in case $\mathbb{E}|h(\xi_0)| < 1$.
Notice that analogue expressions may be provided under more complicated assumptions for models like

$$X_n = h(\xi_n, \xi_{n-1}, \ldots, \xi_{n-p+1})X_{n-1} + \xi_n$$

in this case $\xi_n, \xi_{n-1}, \ldots, \xi_{n-p+1}$ and X_{n-1} are not intependent anymore which needs additional moment conditions.

5.2.2 LARCH(∞)−models

Consider the general non-Markov model

$$X_n = \left(b_0 + \sum_{j=1}^{\infty} b_j X_{n-j} \right) \xi_n$$

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

$$(\mathbb{E}|\xi_0|^p)^{\frac{1}{p}} \sum_{k=1}^{\infty} |b_k| < 1.$$

If now the variables ξ_n are centered and admit a finite variance the previous representation still holds in $\mathbb{L}^2$ if

$$\mathbb{E}\xi_0^2 \sum_{k=1}^{\infty} b_k^2 < 1$$

A vector valued variant of this model as well as a random field variant 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 iid random variables (ξ_j)

$$Y_n = \left(b_0 + \sum_{j=1}^{J} b_j Y_{n-j}^2 \right)^{\frac{1}{2}} \xi_n$$

In this case the vector valued model $Y_n = (X_n, \ldots, X_{n-J})$ is Markov for large enough J.

5.3 Stable Markov chains

Any Markov chain (homogeneous in time) X_t may be represented as the solution of a recursion

$$X_t = M(X_{t-1}, \xi_t) \tag{5.1}$$

with $M(u, z)$ a (measurable) kernel and $\{\xi_t\}$ an iid sequence (see Proposition 7.6 in [20]).
The objective is to determine simple conditions for such iterative models to admit a stationary solution. Further we will see that such solutiions write as Bernoulli schemes (5.6).
Suppose $X_n \in \mathbb{R}^d$ and $\xi_n \in E$ for $d \geq 1$ and for some measurable space $(E, \mathcal{E})$. As in Duflo (1990) such a model is Lipschitz if the kernel $M(u, z)$ fits the condition

$$\mathbb{E}\|M(u, \xi_0) - M(v, \xi_0)\|^p \leq a\|u - v\|^p \tag{5.2}$$

for all $u, v \in \mathbb{R}^d$, $p \geq 1$ and for some fixed $a < 1$ (we denote by $\|\cdot\|$ any norm onr $\mathbb{R}^d$). Suppose (ξ_t) is iid with values in E. 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 $e = 0$).
Define $(U_t^{(n)})_{t\in\mathbb{Z}}$ a Markov chain such that

$$U_t^{(n)} = \begin{cases} u_0 & \text{if } t \leq -n, \\ M(U_{t-1}^{(n)}, \xi_t) & \text{if } t > -n. \end{cases}$$

Lipschitz condition implies with independence of inputs:

$$\mathbb{E}\left\|U_0^{(n)} - U_0^{(n+1)}\right\|^p \leq a\mathbb{E}\left\|U_0^{(n-1)} - U_0^{(n)}\right\|^p.$$

From a recursion

$$\mathbb{E}\left\|U_0^{(n)} - U_0^{(n+1)}\right\|^p \leq a^n\mathbb{E}\left\|M(u_0, \zeta_0) - u_0\right\|^p$$

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 $\{\xi_t,\ t \leq 0\}$ 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 $(U_t^{(0)})_t$ et $(U_t^{(1)})_t$, 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 $\gamma_n = \mathbb{E}^{1/p}\left\|U_n^{(0)} - U_n^{(1)}\right\|^p \leq a^{1/p}\gamma_{n-1}$ and

$$\gamma_n \leq a^{n/p}\gamma_0 = a^{n/p}\mathbb{E}^{1/p}\left\|M(u_0, \zeta_0) - u_0\right\|^p$$

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

$$\exists u_0, \quad \mathbb{E}|M(u_0, \xi_0)|^p < \infty \tag{5.3}$$

Only set $U_{-n}^{(n)} = M(u_0, \xi_{-n})$. In fact we obtain the following:

Theorem 5.3.1 *Assume that conditions (5.3) and (5.2) hold for some $p \geq 1$. The equation (5.1) admits a stationary condition in $\mathbb{L}^p$ such that for each $t \in \mathbb{Z}$, X_t is measurable wrt to the sigma-algebra $\mathcal{F}_t = \sigma(\xi_s, s \leq t)$.*

Diaconis and Friedmann (1999) provide series of examples for which the previous technique applies. See also Doukhan (1994, 2002).

5.3.1 AR-ARCH models

Let $d = 1$, $E = \mathbb{R}$ and $m = 2$ set

$$M(u, z) = A(u) + B(u)z \tag{5.4}$$

for Lipschitz functions $A(u), B(u), u \in \mathbb{R}$.
If

$$\mathrm{Lip}(A) = \sup_{u \neq v} \frac{|A(u) - A(v)|}{|u - v|}$$

then Duflo condition hold if $\mathbb{E}\xi_t = 0$ with

$$a = (\mathrm{Lip}(A))^2 + \mathbb{E}\xi_0^2 (\mathrm{Lip}(B))^2 < 1.$$

Examples of such models follow:

- 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 quation

$$X_n = B(X_{n-1})\xi_n$$

- $AR - ARCH(1)$ classical model 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 initially defined conditionally wrt to its pas 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.

Remark 5.3.1 *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 (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: $\mathbb{E}X_t^2 = (\beta + \gamma^2 \mathbb{E}X_{t-1}^2)\|\xi_t\|_2^2 = \beta\|\xi_t\|_2^2 + \mathbb{E}X_{t-1}^2$ (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 5.3.1 *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 $\mathbb{E}Z \leq \|Z\|_r$. As in the proof of (1.1) 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 $u \neq z$ 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. ∎

5.3.2 Branching models

Here $d = 1$ and $E = \mathbb{R}^D$ for some $D \geq 2$ and we choose again $m = 2$. Let $\xi_t = \left(\xi_t^{(1)}, \ldots, \xi_t^{(D)}\right)$ be such that $\mathbb{E}\xi_t^{(i)}\xi_t^{(j)} = 0$ if $i \neq j$.
If the functions $A_1, \ldots, A_D$ are Lipschitz on $\mathbb{R}$ we set

$$M\left(u, \left(z^{(1)}, \ldots, z^{(D)}\right)\right) = \sum_{j=1}^{D} A_j(u) z^{(j)}, \quad \text{for} \quad (z^{(1)}, \ldots, z^{(D)}) \in \mathbb{R}^D.$$

[2] This result is used by Zbigniew Szewcsak to derive a central limit theorem with unusual $\sqrt{n \log n}$ rate in this case.

The previous contraction assumption writes with the Euclidean norm $\|\cdot\|$ if

$$a=\sum_{j=1}^{D}\left(\mathrm{Lip}(A_j)\right)^2\mathbb{E}\left|\xi_0^{(j)}\right|^2<1.$$

For example,

- if $D=2$ and $\xi_t^{(1)}\sim b(p)$ are independent and Bernoulli distributed and independent of the centered and iid real valued sequence $\xi_t^{(2)}\in\mathbb{L}^2$ and $A_1(u)=u$, $A_2(u)=1$ the previous relation holds if $p<1$.
  This model is defined through the equation

$$X_n=\begin{cases} X_{n-1}+\xi_n^{(2)} & \text{if} \quad \xi_n^{(1)}=1\\ \xi_n^{(2)} & \text{if} \quad \xi_n^{(1)}=0\end{cases}$$

- 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

$$X_n=\begin{cases} A_1(X_{n-1})+\xi_n^{(3)} & \text{if} \quad \xi_n^{(1)}=1\\ A_2(X_{n-1})+\xi_n^{(3)} & \text{if} \quad \xi_n^{(1)}=0\end{cases}$$

5.3.3 Integer valued models

Definition 5.3.1 *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*

$$a\circ x=\sum_{i=1}^{x}Y_i,\qquad \text{for } x\ge 1 \text{ and } 0 \text{ else}$$

for a sequence of iid random variables with $Y_i\sim\mathbf{P}(a)$. The rvs Y_i are also assumed to be context free, ie. independent of any past history.

For example Galton Watson processes with immigration fit the simple equation $X_t=a\circ X_{t-1}+\zeta_t$ for another iid and integer valued sequence (ζ_t) independent of this operator. This means that for an iid triangular array $(Y_{t,i})_{t\in\mathbb{Z},i\in\mathbb{N}}$ we have

$$X_t=\sum_{i=1}^{X_{t-1}}Y_{t,i}+\zeta_t$$

Hence we again write

$$X_t = M(X_{t-1}, \xi_t) \qquad \text{with iid} \qquad \xi_t = ((Y_{t,i})_{i\ge 1}, \zeta_t)$$

Here $M(0,\xi_0) = \zeta_0$ hence $\|M(0,x_0)\|_p = \|\zeta_0\|_p$. Now for $y > x$ and $p \ge 1$ we derive $M(y,\xi_0) - M(x,\xi_0) = \sum_{x+1}^{y} Y_i$; thus $\|M(y,\xi_0) - M(x,\xi_0)\|_p \le a|y-x|$. 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.

Another way to produce big classes of integer valued models follows the same lines as for AR-ARCH models.

$$X_t \,|\, \mathcal{F}_{t-1} \;\sim\; \mathcal{P}(\lambda_t), \qquad \lambda_t = f(X_{t-1}, \lambda_{t-1})$$

A simple solution of this equation write as a recursive system

$$X_t = \mathcal{P}_t(\lambda_t), \qquad \lambda_t = f(X_{t-1}, \lambda_{t-1})$$

for some iid sequence $\mathcal{P}_t$ of Poisson processes. Note that X_t is not Markov and that either (λ_t) or (X_t, λ_t) are Markov processes (or iterative systems). As an exercise on may check the existence of L^1 solutions of those processes. A main point relies on the fact that for any Poisson process

$$|\mathcal{P}(u) - \mathcal{P}(v)| \sim \mathcal{P}(|u-v|)$$

5.3.4 Non linear AR(p)-models

The (real valued) non linear auto-regressive model with order d writes:

$$X_t = r(X_{t-1}, \ldots, X_{t-d}) + \xi_t, \tag{5.5}$$

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 $A(u_1, \ldots, u_d) = \big(r(u_1, \ldots, u_d), u_1, \ldots, u_{d-1}\big)$.
Assume $\mathbb{E}|\xi_0|^m < \infty$ et

$$|r(u_1, \ldots, u_d) - r(v_1, \ldots, v_d)| \;\le\; \sum_{i=1}^{d} a_i |u_i - v_i|$$

for $a_1, \ldots, a_d \ge 0$ such that

$$\alpha = \left(\sum_{i=1}^{d} a_i\right)^{1/d} < 1.$$

Define a norm on $\mathbb{R}^d$ by

$$\|(u_1,\ldots,u_d)\| = \max\{|u_1|, \alpha|u_2|, \ldots, \alpha^{d-1}|u_d|\}.$$

Let $u=(u_1,\ldots,u_d)$, $v=(v_1,\ldots,v_d)\in\mathbb{R}^d$ we set $w_j=|u_j-v_j|$ for $j=1,\ldots,d$ then

$$\|A(u)-A(v)\| \le \max\left\{\alpha^d \max\{w_1,\ldots,w_d\}, \alpha w_1, \ldots, \alpha^{d-1}w_{d-1}\right\} \le \alpha\|u-v\|.$$

Duflo condition thus holds with $a=\alpha^m<1$.

5.4 Bernoulli schemes

The foliowing 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.

5.4.1 Definitions

Definition 5.4.1 (Unformal definition) *Such models write*

$$X_n = H(\xi^{(n)}), \qquad \text{with} \qquad \xi^{(n)} = (\xi_{n-t})_{t\in\mathbb{Z}} \tag{5.6}$$

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 iid 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 abefore 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

$$\sum_{n=1}^{\infty} \zeta_n < \infty$$

with

$$\zeta_n^p = \mathbb{E}\left|H\left(\widehat{\xi}(n)\right) - H\left(\widehat{\xi}(n-1)\right)\right|^p \tag{5.7}$$

Proposition 5.4.1 *Let $p\ge 1$ be such that $\sum_{n=1}^{\infty}\zeta_n<\infty$ then the sequence $(X_n)_{n\in\mathbb{Z}}$ defined this way is stationary and $\mathbb{L}^p-$valued.*

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 $Z_n = (X_{n+1},\dots,X_{n+s}) \in \mathbb{R}^s$.
This is thus the limit of of a sequence of $\mathbb{R}^s-$valued random variable with a distribution independent of n. ∎

Example 5.4.1

- *Let $H : \mathbb{R}^m \to \mathbb{R}$ the process $X_n = H(\xi_n,\dots,\xi_{n-m+1})$ is an $m-$dependent sequence i.e. $\sigma\{X_j, j < a\}$ and $\sigma\{X_j, j > a+m\}$ are independente $\sigma-$algebras.*

- *Stochastic volatility model. Let $Y_n = H(\xi_n,\xi_{n-1},\dots)$ be a causal Bernoulli scheme such that the iid innovations $\xi_n \in \mathbb{L}^2$ are centered.*
  Set
  $$X_n = \xi_n Y_{n-1} = \xi_n H(\xi_{n-1},\xi_{n-2},\dots)$$
  The sequence X_n is orthogonal and $Var(X_n|\mathcal{F}_{n-1}) = Y_{n-1}^2$. This property incates possible brusk changes adapted to model stock exchange..

- *All the previous sections of the present chapter provide us with a series of examples of this situation.*

The previous definition 5.4.1 is really adapted to deal with the previous chaotic examples for which tails may be bounded.
A more general setting is adapted to prove existence of stationary processes.

Definition 5.4.2 (Formal definition) *Let μ denote a distribution on a measurable $(E,\mathcal{E})$.*
Consider an iid 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 5.4.1 *The spaces $\mathbb{L}^p(\nu)$ and $\mathbb{L}^p(\nu^+)$ are Banach spaces (complete normed vector spaces) equipped respectively with the norms*

$$\begin{aligned}\|H\|_p &= \left(\mathbb{E}|H\left((\xi_n)_{n\in\mathbb{Z}}\right)|^p\right)^{\frac{1}{p}}\\ &= \left(\mathbb{E}|H\left((\xi_n)_{n\in\mathbb{N}}\right)|^p\right)^{\frac{1}{p}}\end{aligned}$$

The definition of Bernouilli schemes is as in the unformal definition5.4.1 and holds non causal or causal schemes for elements $H \in \mathbb{L}^p(\nu)$ or $\mathbb{L}^p(\nu^+)$ respectively. Morevor condition

$$\sum_{n=1}^{\infty} \zeta_n < \infty \tag{5.8}$$

before eqn. (5.7) imply with proposition 5.4.1, a simple sufficient condition for for functions of infinitely many variables to exist in those huge spaces.
Next subsection also proves that those assumptions are relevant to prove short range conditions.

Proof of theorem 5.3.1. 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 station solution of eqn. (5.1) writes

$$X_t = H(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots)$$

Tho 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 (5.3) and (5.2) 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 wrt $\xi_1, \xi_2, \ldots$ implies

$$\|K - K'\|_p \le a\|H - H'\|_p$$

The classical Picard fixed point theorem thus implies that Φ admits a unique fixed point H. ∎

5.4.2 Weak dependence of Bernoulli schemes

The following results also give an idea for the interest of such models. Those ideas are widely developed later.
Let $(\xi'_k)_{k\in\mathbb{Z}}$ another iid 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:

$$\delta_n^p = \mathbb{E}\left|H\left(\widetilde{\xi}(n)\right) - H(\xi)\right|^p$$

Remark 5.4.2 *Replace* $\widetilde{\xi}(n)$ *by*

$$\widehat{\xi}(n)_k = \begin{cases} \xi_k & \text{if} \quad |k| \neq n \\ \xi'_k & \text{if} \quad |k| = n \end{cases}$$

leads to the fruitful physical measure of dependence by Wei Biao Wu.
Another alternative is to set

$$\xi'(n)_k = \begin{cases} \xi_k & \text{if} \quad |k| \leq n \\ 0 & \text{if} \quad |k| > n \end{cases}$$

which is essentially the same as $\widetilde{\xi}(n)$ *and makes easy to define functions of infinitely many variable as limits of functions of finitely many variables in the Banach space* $\mathbb{L}^p(\nu^+)$ *from the fact act that* $\xi'(n)$ *is a Cauchy sequence see eqn. (5.8).*

and:

Proposition 5.4.2 *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::*

$$\begin{aligned} |Cov(X_0, X_k)| &\leq 2(\mathbb{E}|X_0|^p)^{1/p}\delta_{[k/2]} \\ | &\leq 2\|H\|_\infty \delta^2_{[k/2]} \end{aligned}$$

If the Bernoulli scheme is causal the previous inequalities write:

$$\begin{aligned} |Cov(X_0, X_k)| &\leq (\mathbb{E}|X_0|^p)^{1/p}\delta_k, \\ &\leq \|H\|_\infty \delta^2_k \end{aligned}$$

Remark 5.4.3 *Such results imply short range dependence of the process* X *in the sense of definition 2.3.1, in case the above covariances are summable.*

Proof. Use Hölder 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 $X_{k,l} = H\left(\widetilde{\xi}(l)^{(k)}\right)$.
Remember $\widetilde{\xi}(l)^{(k)}$ 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

$$\mathrm{Cov}\,(X_0, X_k) = \mathrm{Cov}\,(X_0, X_k - X_{k,k}).$$ ∎

A question is the heredity of such quantities through images $Y_k = g(X_k)$. Denote the corresponding expressions by $\delta_{k,Y}$ $\delta_{k,X}$.
Then:

Lemma 5.4.1 *Assume that* $m \geq 1$ *and* $g : \mathbb{R} \to \mathbb{R}$ *satisfies*

$$|g(x) - g(y)| \leq L|x - y|, \qquad \forall x, y \in \mathbb{R}$$

for some constant $L > 0$.
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 but indicators $g_x(u) = \mathbb{1}_{\{u \leq x\}}$ are convenient to derive bounds for the empirical process. We obtain:

Lemma 5.4.2 *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 proces defined by par* $Y_{x,n} = \mathbb{1}_{\{X_n \leq x\}}$ *satisfies:*

$$\delta_{k,Y_x} \leq 2(2C)^{2/(c+2)} \delta_{k,X}^{\frac{c}{c+2}}$$

Proof. Consider: $Y_{x,\epsilon,n} = g_{x,\epsilon}(X_n)$ with the continuous function $g_{x,\epsilon} = 1$ if $u \leq x - \epsilon$, =0 pour $u \geq x$ and linear else.
Then $|g_{x,\epsilon}(u) - g_{x,\epsilon}(v)| \leq |u - v|/\epsilon$ and $\delta_{k,Y_{x,\epsilon}} \leq \delta_{k,X}/\epsilon$.
Moreover $|\delta^2_{k,Y_{x,\epsilon}} - \delta^2_{k,Y_x}| \leq 2\mathbb{P}(X_0 \in [x - \epsilon, x]) \leq 2C\epsilon^c$.
So $\delta^2_{k,Y_{x,\epsilon}} \leq \frac{\delta^2_{k,X}}{\epsilon^2} + 2C\epsilon^c$. 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 5.4.4 *A control for the cumulative empirical function follows:*

$$\mathit{Var}\, F_n(x) = \mathcal{O}\left(\frac{1}{n}\right) \qquad \text{if} \qquad \sum_{k=0}^{\infty} \delta_{k,X}^{c/(c+2)} < \infty.$$

In case $c = 1$ *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.*

Chapter 6

Associated processes

6.1 Association

Definition 6.1.1 *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 6.1.2 *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 6.1.1 *Covariances of an associated process are non negative if they are well defined.*

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 6.1.1 *Independent vectors are associated.*
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 6.1.1 *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 $x \in \mathbb{R}^p$. 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 Schwartz inequality one derive $F(X), G(X) \in \mathbb{L}^2$. ∎

Heredity of association is very important to handle applications.

Example 6.1.1 *The following example inherit association properties*

- *A LARCH(∞) with nonnegative coefficients ,*
- *An autoregressive process solution of an equation*

$$X_t = r(X_{t-1}) + \xi_t$$

  if the function $r : \mathbb{R} \to \mathbb{R}$ is a non decreasing function,

- *a non-decreasing image of an associated sequence.*

6.2 A main inequality

A new concept is needed

Definition 6.2.1 *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 6.2.1 *Assume that the function f satisfies*

$$|f(y) - f(x)| \leq a_1|y_1 - x_1| + \cdots + a_p|y_p - x_p|$$

for all vectors $x = (x_1, \ldots, x_p), y = (y_1, \ldots, y_p) \in \mathbb{R}^p$ 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:

$$-a_1(y_1 - x_1) + \cdots - a_p(y_p - x_p) \leq f(y) - f(x) \leq a_1(y_1 - x_1) + \cdots + a_p(y_p - x_p).$$

An essential inequality follows:

Lemma 6.2.1 *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:*

$$|Cov(f(X), g(X))| \leq Cov(f_1(X), g_1(X))$$

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 $\text{Cov}\,(f(X)+af_1(X), g(X)+bg_1(X))$ are non negative if $a, b = 0$ or 1, then add them 2 by 2 yields the result. ∎

Let now $(Y, Z) \in \mathbb{R}^u \times \mathbb{R}^v$ be an associated vector in $\mathbb{L}^2$. If the functions f et g satisfy

$$|f(y) - f(y')| \le \sum_{i=1}^{u} a_i |y_i - y_i'|, \;\; |g(z) - g(z')| \le \sum_{j=1}^{v} b_j |z_j - z_j'|$$

the previous lemma implies with Example 6.2.1:

$$|\text{Cov}\,(f(Y), g(Z))| \le \sum_{i=1}^{u} \sum_{j=1}^{v} a_i b_j \text{Cov}\,(Y_i, Z_j) \tag{6.1}$$

Thus eg. If the vectors Y, Z admit pairwise orthogonal components then they are stochastically independent as for the Gaussian case.
Also

$$\begin{aligned} |\text{Cov}\,(f(X), g(X))| &\le \text{Lip}\, f \cdot \text{Lip}\, g \sum_{i=1}^{u} \sum_{j=1}^{v} \text{Cov}\,(Y_i, Z_j) \\ &\le uv \text{Lip}\, f \cdot \text{Lip}\, g \max_{1 \le i \le u} \max_{1 \le j \le v} \text{Cov}\,(Y_i, Z_j) \end{aligned}$$

this implies the κ-weak dependence for such models (see chapter 7).

6.3 Limit theory

If the condition

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \text{Cov}\,(X_0, X_n) < \infty,$$

holds for the stationary and associated process $(X_n)_{n\in\mathbb{Z}}$ then the following weak invariance principle holds:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} X_k \to \sigma W_t \quad \text{ in the Skohorod space } D[0,1].$$

Chapter 7

Dependence

The chapter begins with a central result saying that the strong law of large numbers (SLLN) works for most of the previously introduced models. The question of convergence rates in this results is solved in the forthcoming dependence types for stationary sequences. Anyway those will also give access to goodness of fit tests through associated limit theorems which give asymptotic confidence bounds..

7.1 Ergodic theorem

This section is due to Jérôme Dedecker.

Definition 7.1.1 *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, with a measurable inverse and if $\mathbb{P}(T(A)) = \mathbb{P}(A)$ 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 7.1.1 (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 $\left(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}})\right)$. The sigma-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$ is generated by elementary events $A = \prod_{k\in\mathbb{Z}} A_k$ with $A_k = \mathbb{R}$ excepted for finitely many indices k.
The transformation T defined by $T(x)_i = (x_{i+1})$ for $x = (x_i)_{i\in\mathbb{Z}}$ satisfies $T(\prod_{k\in\mathbb{Z}} A_k) = \prod_{k\in\mathbb{Z}} A_{k+1}$. If 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 est ergodic (i.e. $\mathbb{P}(A) = 0$ or 1 if $A \in \mathcal{J}$) the process $X = (X_n)_{n\in\mathbb{Z}}$ is ergodic.

Proposition 7.1.1 *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 7.1.1. Let C denote the closure (in $\mathbb{L}^2(\Omega,\mathcal{A},\mathbb{P})$) of the convex hull $\mathcal{C}$ of $E=\{f\circ T^k\ /\ k\in\mathbb{Z}\}$ ([1]). From orthogonal projection (e.g. see Doukhan, Sifre, théorème 3.81, page 124, volume 1, 2001) there exists a unique $\overline{f}\in C$ with $\|\overline{f}\|_2=\inf\{\|g\|_2\ /\ g\in C\}$. If one prove $\|R_n(f)\|_2\to_{n\to\infty}\|\overline{f}\|_2$ 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

$$\|\overline{f}\circ T-\overline{f}\|_2\le\|\overline{f}\circ T-R_{n-1}(f)\circ T\|_2+\frac{1}{n}\|f\|_2+\|R_n(f)-\overline{f}\|_2.$$

$\mathbb{P}$-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 $R_n(f)\to\overline{f}$ in $\mathbb{L}^2$ 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 $g=\sum_{|j|\le k}a_jf\circ T^j\in\mathcal{C}$ with $\|g\|_2\le\|\overline{f}\|_2+\epsilon$. With the invariance of T we derive $\|R_n(g)\|_2\le\|g\|_2\le\|\overline{f}\|_2+\epsilon$.
From another hand

$$\|R_n(f-g)\|_2=\|\sum_{j=-k}^k a_j(R_n(f)-R_n(f\circ T^j))\|_2\le\sum_{j=-k}^k a_j\|R_n(f)-R_n(f\circ T^j)\|_2$$

and using again T's invariance,

$$\|R_n(f)-R_n(f\circ T^j)\|_2\le\frac{1}{n}\sum_{i=k+1}^{k+j}(\|f\circ T^j\|_2+\|f\circ T^{-j}\|_2)\le\frac{2j}{n}\|f\|_2 \quad (7.1)$$

Thus $\|R_n(f-g)\|_2\le\sum_{|j|\le k}\frac{2ja_j}{n}\|f\|_2\le\frac{2k}{n}\|f\|_2\to_{n\to\infty}0$.
Hence $\|\overline{f}\|_2\le\limsup_n\|R_n(f)\|_2\le\|\overline{f}\|_2+\epsilon$ yielding the result. ∎

Corollary 7.1.1 *If we only assume $\mathbb{E}|f|<\infty$ then*

$$R_n(f)\xrightarrow{\mathbb{L}^1}_{n\to\infty}\mathbb{E}^{\mathcal{I}}f.$$

[1] $C=\overline{\mathcal{C}}$ with

$$\mathcal{C}=\left\{\sum_{i=1}^I a_ix_i;\ a_i\ge 0, x_i\in E,\sum_{i=1}^I a_i=1, I\ge I\right\}$$

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

$$\begin{aligned}\|R_n(f) - \mathbb{E}^{\mathcal{I}} f\|_1 &\le \|R_n(f - g_m)\|_1 + \|R_n(g_m) - \mathbb{E}^{\mathcal{I}}(g_m)\|_1 + \|\mathbb{E}^{\mathcal{I}}(g_m - f)\|_1 \\ &\le 2\|f - g_m\|_1 + \|R_n(g_m) - \mathbb{E}^{\mathcal{I}}(g_m)\|_1.\end{aligned}$$

The previous proposition implies $\limsup_n \|R_n(f) - \mathbb{E}^{\mathcal{I}} f\|_1 \le 2\|f - g_m\|_1$. 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 7.1.1 (Hopf maximal inequality) *Let T be a bijective bi-measurable and $\mathbb{P}$-invariant transformation. For $f \in \mathbb{L}^1$ set $S_0(f) = 0$ et si $k \ge 1$, $S_k(f) = \sum_{j=1}^k f \circ T^j$ then $S_n^+(f) = \max_{0\le k\le n} S_k(f)$.*
Thus $\mathbb{E}\left(f \circ T \cdot 1\!\!1_{S_n^+(f)>0}\right) \ge 0$.

Proof of lemma 7.1.1. If $1 \le k \le n+1$ then $S_k(f) \le f \circ T + S_n^+(f) \circ T$.
Moreover if $S_n^+(f) > 0$ then $S_n^+(f) = \max_{1\le k\le n} S_k(f)$.
Thus

$$S_n^+(f) 1\!\!1_{S_n^+(f)>0} \le f \circ T 1\!\!1_{S_n^+(f)>0} + S_n^+(f) \circ T 1\!\!1_{S_n^+(f)>0}.$$

This entails $f \circ T 1\!\!1_{S_n^+(f)>0} \ge (S_n^+(f) - S_n^+(f) \circ T) 1\!\!1_{S_n^+(f)>0}$.
Now $\mathbb{E} f \circ T 1\!\!1_{S_n^+(f)>0} \ge \mathbb{E} S_n^+(f) - \mathbb{E} S_n^+(f) \circ T = 0$. ∎

Corollary 7.1.2 *Assume that assumptions in lemma 7.1.1 hold then*

$$\mathbb{P}\left(\sup_{n\ge1} |R_n(f)| > c\right) \le \frac{\mathbb{E}|f|}{c}, \qquad \forall c > 0.$$

Proof. Apply lemma 7.1.1 to $f - c$: $\mathbb{E}(f - c) \circ T 1\!\!1_{S_n^+(f-c)>0} \ge 0$.
Hence $\dfrac{\mathbb{E} f \vee 0}{c} \ge \dfrac{\mathbb{E} f \circ T 1\!\!1_{S_n^+(f-c)>0}}{c} \ge \mathbb{P}(S_n^+(f - c) > 0)$.
Thus $S_n^+(f - c) = 0 \vee \max_{1\le k\le n} k\,(R_k(f) - c) \ge \max_{1\le k\le n}(R_k(f) - c)$.
Hence $\dfrac{\mathbb{E} f \vee 0}{c} \ge \mathbb{P}\left(\max_{1\le k\le n}(R_k(f) - c) > 0\right)$.
Replace f by $-f$ one proves analogously:

$$\frac{-(\mathbb{E} f \wedge 0)}{c} \ge \mathbb{P}(S_n^+(f + c) < 0) \ge \mathbb{P}\left(\max_{1\le k\le n}(R_k(f) + c) < 0\right).$$

The result follows from suming up the previous inqualities and making $n \to \infty$. 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 7.1.1 (Ergodic theorem) *Let T bijective bi-measurable and $\mathbb{P}$-invariant. Let $f \in \mathbb{L}^1$ then*

$$R_n(f) \to_{n\to\infty} \mathbb{E}^{\mathcal{I}} f, \quad a.s.$$

Proof of theorem 7.1.1. Assume first that g is bounded. If $n, m \geq 1$ then

$$\left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| \leq |R_n(g - R_m(g))| + \left|R_n(R_m(g) - \mathbb{E}^{\mathcal{I}} g)\right|.$$

Using the same idea as to derive inequality (7.1) we obtain

$$\|R_n(g) - R_n(g \circ T^j)\|_\infty \leq 2j\|g\|_\infty/n$$

Hence

$$|R_n(g - R_m(g))| \leq \frac{\|g\|_\infty}{nm} \sum_{j=1}^{m} 2j = \frac{(m+1)\|g\|_\infty}{n}.$$

Thus

$$\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| \leq \sup_{n \geq 1} \left|R_n(R_m(g) - \mathbb{E}^{\mathcal{I}} g)\right| \leq \left|R_n(R_m(g) - \mathbb{E}^{\mathcal{I}} g)\right|, \text{ a.s.}$$

With corollary 7.1.2 we derive

$$\mathbb{P}\left(\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| > c\right) \leq \frac{1}{c}\mathbb{E}\left|R_m(g) - \mathbb{E}^{\mathcal{I}} g\right| \to_{m\to\infty} 0.$$

Ainsi $\mathbb{P}\left(\limsup_n \left|R_n(g) - \mathbb{E}^{\mathcal{I}} g\right| = 0\right) = 1$.
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 $\limsup_n \left|R_n(f) - \mathbb{E}^{\mathcal{I}} f\right| \leq \sup_{n\geq 1} |R_n(f - g_m)| + \left|\mathbb{E}^{\mathcal{I}}(g_m - f)\right|$ *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 7.1.1:

$$\mathbb{P}(A_m > c) \leq \frac{1}{c}\|f - g_m\|_1$$

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 7.1.3 *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*

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\to_{n\to\infty}\mathbb{E}^{\mathcal{J}}f(X)\quad a.s.\qquad and\ in\ \mathbb{L}^1.$$

If $\mathbb{E}f^2(X)<\infty$ the convergence also holds in $\mathbb{L}^2$.

Proof. The only point to notice is that $\mathbb{E}^{\mathcal{J}}f(X)=\mathbb{E}^{\mathcal{I}}_{\mathbb{P}_X}f$. ∎

Remark 7.1.2 *If the process X is ergodic*

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\longrightarrow\mathbb{E}f(X)\quad if\ \mathbb{E}|f(X)|<\infty.$$

Ergodicity may also be omitted if $\mathbb{E}f^2(X)<\infty$ and

$$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\to_{n\to\infty}\mathbb{E}f(X)\ a.s.\Leftrightarrow\frac{1}{n}\sum_{k=1}^{n}\ Cov(f(X),f\circ T^k(X))\to 0$$

After those remarks we derive examples of ergodic processes.

Example 7.1.1 (ergodic processes)

- *An i.i.d. sequence is also a stationary and ergodic sequence.Use Kolmogorov law of $0-1$.*
- *Hence Bernoulli schemes are also ergodic. Indeed if $X=(X_i)_{i\in\mathbb{Z}}$ is defined from an iid sequence $\xi=(\xi_i)_{i\in\mathbb{Z}}$ and a function H through equation (5.6) 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 $\mathbb{R}^{\mathbb{Z}}$ dans $\mathbb{R}$ which entails the ergodicity of X.
- *If relation $Cov(f(X_0),f(X_n))\to 0$ as $n\to\infty$ for $f\in\mathcal{F}$ (a class of functions which generates linearly a dense vector subspace in $\mathbb{L}^1$) . Indeed this relation implies with Cesaro lemma that*

  $$\frac{1}{n}\sum_{k=1}^{n} f\circ T^k(X)\to_{n\to\infty}\mathbb{E}f(X)\qquad dans\ \mathbb{L}^1.$$

  The result still holds for each bounded function from a density argument. Now corollary 7.1.3 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* $r_n \to 0$ *(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* $[-1,1]$ *with* $G(1) = \mathbb{E}f^2(X_0)$ *and* $G(0) = 0$*. Ergodicity follows.*

- *Weakly dependent sequences (see the definition (7.5)) are mostly ergodic according to the same proof.*

- *A last example is that of stationary associated sequences such that* $r_n \to 0$*. For this use inequality (6.1).*

Remark 7.1.3 *Let* $(X_n)_{n\in\mathbb{Z}}$ *be a stationary and ergodic centered sequence in* $\mathbb{L}^2$*. Then* $\widehat{c_{n,p}} = \frac{1}{n-|p|}\sum_{k=|p|+1}^{n} X_k X_{k-|p|}$ *fits* $c_p = \mathbb{E}X_0X_p$ *without bias (i.e.* $\mathbb{E}\widehat{c_{n,p}} = c_p$*) and* $\widehat{c_{n,p}} \to c_p$ *a.s. and in* $\mathbb{L}^1$ *(it is consistant). 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

$$\widehat{a_n} = \sum_{k=2}^{n} X_n X_{n-1} / \sum_{k=1}^{n} X_n^2 \longrightarrow a$$

a.s. if $\mathbb{E}\xi_0 = 0$ *and* $\mathbb{E}\xi_0\xi_p = 0$ *as* $p \to \infty$*.*

7.2 Range

We aim at providing some ideas yielding definitions for the range of a process. The proof of proposition 2.3.2 provides an expression of the square of a convergence rate in $\mathbb{L}^2$ in the ergodic theorem

$$v_n^2 = \mathbb{E}(S_n - n\mathbb{E}X_0)^2 = \mathrm{Var}\left(\sum_{k=1}^{n} g(X_k)\right)$$

Definition 7.2.1 *The classical definition of the long/short range dependence for second order stationary sequences is based on the convergence of series of covariances*

$$\sum_k r_k$$

In case this series is absolutely convergent the process is short range dependent (SRD) and if the series diverges the process is long range dependent (LRD).

Based on this definition the partial sums

$$S_n = \sum_{k=1}^{n} X_k$$

admit variances with order n or $>> n$. A phenomenon of very short range corresponds to $g_X(0) = 0$; in this case $\operatorname{Var} S_n << n$.
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

$$\sum_{k=-n}^{n} r_k \sim n^{\alpha} L_1(n) \tag{7.2}$$

$$r_k \sim k^{-\beta} L_2(n) \tag{7.3}$$

$$g_X(\lambda) \sim |\lambda|^{-\gamma} L_3\left(\frac{1}{|\lambda|}\right) \qquad (\lambda \to 0) \tag{7.4}$$

One may prove (Taqqu, 2002, [10])

Theorem 7.2.1 (Tauber) *If r_k is monotonous for $k \geq k_o$ then relations (7.2), (7.3) and (7.4) 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 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 $\operatorname{Var} S_n = n \operatorname{Var} X_0$ does not imply an asymptotically Gaussian behavior.

Example 7.2.1 *Let (ξ_n) be an iid 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.*

An attractive definition is thus based on limit theorems relative to the sequence

$$S_n = X_1 + \cdots + X_n$$

- if $\frac{1}{\sqrt{n}} S_n$ is asymptotically Gaussian one expects short range dependence. Precisely we may suppose that $\operatorname{Var} S_n \sim cn$ (as $n \to \infty$) for some constant $c > 0$.
  Assume that the sequence of processes

  $$t \in [0,1] \mapsto \frac{1}{\sqrt{\operatorname{Var} S_n}} S_{[nt]}$$

  converges toward a Brownian motion

- if the sequence of processes

$$t \in [0,1] \mapsto Z_n(t) = \frac{1}{\sqrt{\operatorname{Var} S_n}} S_{[nt]}$$

  does not converge toward a Brownian motion it would be long range dependence

An alternative is that this limit admits independent increments or not. This nice proposal is that of Herold Dehling and allows to aggregate cases of heavy tail processes and Lévy processes.

7.3 Long range dependence

The most elementary example is that of Gaussian processes. We follow the presentation in Rosenblatt (1985) [22] who discovered the long range dependent behaviors. He considered models of instantaneous functions of a Gaussian process.

7.3.1 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

$$r_k \sim ck^{-a} \qquad \text{as} \qquad k \to \infty$$

for $c > 0$, $a > 0[$ (theorem 2.2.1 proves that the sequence $r_k = (1+k^2)^{-a/2}$ is indeed the sequence of covariances of a stationary Gaussian process: hence there exist such sequences). Tauber theorem 7.2.1 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 $a > 1$, $\operatorname{Var} S_n \sim n\sigma^2$ 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.

- If now $a > 1$ the series of covariances diverges $\mathrm{Var}\, S_n \sim n^{2-a}$ if $r_k \sim ck^{-a}$. Hence
$$Z_n(t) \to \mathcal{N}\left(0, ct^{2-a}\right)$$
does not converge to the Brownian motion; indeed contrary to the Brownian motion the previous variance does not increase linearly with t.

7.3.2 Gaussian polynomials

Generally if the process (X_n) is standard stationary Gaussian ($\mathbb{E}X_0 = 0$, $\mathrm{Var}\, X_0 = 1$) with $r_k \sim ck^{-a}$ and the function g is such that $\mathbb{E}|g(X_0)|^2 < \infty$ then

- $$\mathrm{Var}\left(\sum_{k=1}^{n} g(X_k)\right) = \mathcal{O}(n)$$
if $a \cdot m(g) > 1$ and $m(g)$ denote the Hermite rank of g.
In this first case the diagram formula (§ 3.2.3) allows to prove the convergence in distribution (Breuer et Major, 1984, [3])
$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} g(X_k) \to_{n\to\infty} \sigma W_t, \qquad \text{in} \quad D[0,1]$$
The result is also proved in a shorter way in [21]

- Else say if $a \cdot m(g) < 1$ then
$$\mathrm{Var}\left(\sum_{k=1}^{n} g(X_k)\right) = \mathcal{O}\left(n^{2-m(g)\alpha}\right)$$
Here $1 - \frac{m(g)\alpha}{2} > \frac{1}{2}$ and the convergence still holds
$$\frac{1}{n^{1-\frac{m(g)\alpha}{2}}} \sum_{k=1}^{n} g(X_k) \to_{n\to\infty} Z_r$$
to some non Gaussian distribution in case the rank is > 1 (Taqqu (1975), Dobrushin & Major (1979) [14]). 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.

7.3.3 Rosenblatt process

The previous non Gaussian asymptotic may be proved elementary "à la main" in the case enlighted in Murray Rosenblatt (1961) (see [22]). Set $Y_n = X_n^2 - 1$ then Mehler formula implies that the covariance $\mathrm{Cov}\,(Y_0, Y_k)$ equals $2r_k^2 \sim 2c^2k^{-2a}$.

The series of those covariances is divergent in case $a < \frac{1}{2}$.
In this case we aim at proving that

$$U_n = n^{a-1} \sum_{k=1}^{n} Y_k$$

converges toward a non Gaussian limit.
Set R_n for the covariance matrix of the vector $(X_1, \ldots, X_n)$, then for t small enough:

$$\begin{aligned}
\mathbb{E}e^{tU_n} &= \mathbb{E}e^{tn^{a-1}\sum_{k=1}^{n}(X_k^2-1)} \\
&= e^{-tn^a} \int_{\mathbb{R}^n} e^{-x^t(R_n^{-1}-2tn^{\beta-1}I_n)x/2} \frac{dx}{(2\pi)^{n/2}\sqrt{\text{dét} R_n}} \\
&= e^{-tn^a} \int_{\mathbb{R}^n} e^{-y^t(I_n-2tn^{\beta-1}R_n)y/2} \frac{dy}{(2\pi)^{n/2}} \\
&= e^{-tn^a} \det{}^{-\frac{1}{2}} \left(I_n - 2tn^{a-1}R_n\right)
\end{aligned}$$

Indeed through a linear change in variable for each symmetric definite positive matrix A with order n:

$$\int_{\mathbb{R}^n} e^{-y^tAy/2} \frac{dy}{(2\pi)^{n/2}} = \frac{1}{\sqrt{\det(A)}}$$

Now denote $(\lambda_{i,n})_{1\le i\le n}$ the eigenvalues (≥ 0) of the symmetric and non-negative matrix R_n (diagonalizable) then

$$\begin{aligned}
\det{}^{-\frac{1}{2}} \left(I_n - 2tn^{a-1}R_n\right) &= \prod_{i=1}^{n}(1 - 2tn^{a-1}\lambda_{i,n})^{-1/2} \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \log\left(1 - 2tn^{a-1}\lambda_{i,n}\right)\right)
\end{aligned}$$

Use the following analytic expansion (valid for $|z| < 1$)

$$\log(1-z) + z = -\sum_{k=2}^{\infty} \frac{z^k}{k}$$

nous obtenons donc

$$\begin{aligned}
\mathbb{E}e^{tU_n} &= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\log\left(1 - 2tn^{a-1}\lambda_{i,n}\right) + 2tn^{a-1}\right)\right) \\
&= \exp\left(\frac{1}{2}\sum_{k=2}^{\infty} \frac{1}{k}(2tn^{a-1})^k \text{ trace } R_n^k\right)
\end{aligned}$$

if we prove that

$$e^{-tn^a} \sim \exp\left(-\sum_{i=1}^{n}(2tn^{a-1})\lambda_{i,n}\right) = \exp\left(-2tn^{a-1} \text{ trace } R_n\right)$$

Quote thus

$$\begin{aligned} n^{k(a-1)} \text{ trace } R_n^k &= n^{k(a-1)} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} r_{i_1-i_2} r_{i_2-i_3} \cdots r_{i_{k-1}-i_k} r_{i_k-i_1} \\ &\sim \frac{c^k}{n^k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \left| \frac{i_1}{n} - \frac{i_2}{n} \right|^{-a} \cdots \left| \frac{i_{k-1}}{n} - \frac{i_k}{n} \right|^{-a} \left| \frac{i_k}{n} - \frac{i_1}{n} \right|^{-a} \end{aligned}$$

Hence through a discretization of a multiple integral by Riemann sums we derive ([2]):

$$n^{k(a-1)} \text{ trace } R_n^k \to_{n\to\infty} c_k > 0$$

with

$$c_k = c^k \int_0^1 \cdots \int_0^1 |x_1 - x_2|^{-a} |x_2 - x_3|^{-a} \cdots |x_{k-1} - x_k|^{-a} |x_k - x_1|^{-a} dx_1 \cdots dx_k$$

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 (Rosenblatt distribution). Indeed the logarithm of its Laplace transform is not a polynomial of order 2. ∎

This technique does not extend to polynomials with degree > 2 since its Laplace transform is not analytic and thus the method of moments does not apply to prove convergence in law (see theorem 8.1.1). Pour montrer des théorèmes de limite non centrale, Dobrushin and Major (1979) introduce alternative convergences of sequences of multiple Ito integrals in order to prove such theorems.

7.3.4 Linear processes

Other models admit analogue behaviors. Linear processes

$$X_n = \sum_{k=0}^{\infty} c_k \xi_{n-k}$$

for which $c_k \sim ck^{-\beta}$ with $\frac{1}{2} < \beta < 1$ satisfy

$$r_k = \sum_l c_l c_{l+k} \sim ck^{1-2\beta} \int_0^{\infty} \frac{ds}{(s(s+1))^{\beta}}$$

[2]Polya & Szégö (Problems and Theorems in Analysis, volume 1) prove validity of such approximations for generalized integrals, in case of functions monotonic around their singularity as it is the case here.

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). It follows from the following simple result in [6]:

Theorem 7.3.1 (Davydov, 1970) *Let (X_n) be a linear process. Set $S_n = X_1 + \cdots + X_n$. If $Var\,S_n = n^{2H}L(n)$ $(n \to \infty)$ for a slowly varying function L and $0 < H < 1$ then*

$$\frac{1}{n^H L(n)} \sum_{k=1}^{[nt]} X_k \stackrel{in\ law}{\longrightarrow}_{n\to\infty} B_H(t)$$

This result also relies on the Lindeberg theorem 1.3.1.
The behavior of the empirical process is considered eg. in Doukhan, Lang and Surgailis (2002, Ann. I. H. P.).
A martingales based technique was introduced by Ho and Hsing [18] (1996) for the extension of such behaviors as previously considered for the Gaussian case

7.4 Short range

Conditions on time series such that the standard limit theorems obtained for iid sequences are investigated. Looking for asymptotic independence it seems natural to consider conditions such as this in [12]:

$$|\mathrm{Cov}\ (f(X_{i_1}, \ldots, X_{i_u}), g(X_{j_1}, \ldots, X_{j_u}))| \le \epsilon_r \psi(u, v, f, g) \qquad (7.5)$$

for functions f, g belonging respectively to classes $\mathcal{F}, \mathcal{G}$ and a sequence $\epsilon_r \downarrow 0$ as $r \uparrow \infty$, and

$$i_1 \le \cdots i_u \le j_1 - r \le j_1 \le \cdots \le j_v$$

the function ψ depends on f, g and on the number of their arguments u, v.

7.4.1 Strong mixing

A special case is strong mixing for which

$$\mathcal{F} = \mathcal{G} = \mathbb{L}^\infty \text{ and } \psi(u, v, f, g) = 4\|f\|_\infty \|g\|_\infty.$$

Examples of strongly mixing are given in [9].
The sup bound of such ϵ_r satisfying this inequality writes

$$\alpha_r = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\,;\ A \in \sigma(X_i, i \le 0), B \in \sigma(X_j, j \ge r)\}$$

Anyway this condition does not hold for some models, even very simple ones like

$$X_n = \frac{1}{2}(X_{n-1} + \xi_n)$$

where the iid inputs (ξ_n) admit a Bernoulli distribution with parameter $\frac{1}{2}$. Indeed 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 mean the inclusion of sigma algebras generated by marginals of such processes. 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 marginals of this process are easily proved to be uniformly distributed on $[0, 1]$.
Such examples prove that strong mixing type notions are not enough to consider a wide class of statistical models.

7.4.2 Weak dependence

Doukhan and Louhichi [12] (1999) introduce more easy condition in terms of Lipschitz classes which avoid such problems.
Set $\mathcal{L}$ the class of functions $g : \mathbb{R}^v \to \mathbb{R}$ for an integer $v \geq 1$, with

$$\text{Lip}(g) = \sup_{(x_1,\ldots,x_v)\neq(y_1,\ldots,y_v)} \left| \frac{g(x_1,\ldots,x_v) - g(y_1,\ldots,y_v)}{|x_1 - y_1| + \cdots + |x_v - y_v|} \right|$$

A simple weak dependence condition corresponds to $\mathcal{G} = \mathcal{L}$ and respectively $\mathcal{F} = \mathcal{L}$ (non causal case) or $\mathcal{F} = \mathbb{B}_\infty = \{f \text{ measurables, } \|f\|_\infty \leq 1\}$ (causal case). Here respectively

$$\begin{aligned} \psi(u, v, f, g) &= \psi_1(u, v, f, g) = u\text{Lip}(f) + v\text{Lip}(g), \\ &= \psi_1'(u, v, f, g) = v\text{Lip}(g) \end{aligned}$$

Most of the previous models satisfy such conditions as for Bernoulli schemes. Here

$$\epsilon_r \geq \mathbb{E}\left|H\left((\xi_i)_{i\in\mathbb{Z}}\right) - H\left((\xi_i)_{|i|\leq r}\right)\right|$$

(the sequence $(\xi_i)_{|i|\leq r}$ is obtained by setting 0 for indices with $|i| > r$) and the previous conditions ψ_1' or ψ_1 apply according the fact the the Bernoulli is causal or not.

- Hence $\epsilon_r = \mathbb{E}|\xi_0| \sum_{|i|>2r} |a_i|$ for non causal linear processes $X_n = \sum_i a_i\xi_{n-i}$.
- Respectively $\epsilon_r = \mathbb{E}|\xi_0| \sum_{i>r} |a_i|$ for causal cases, $a_i = 0$ if $i < 0$.

Various applications are considered in [8].
The function $g(x_1, \ldots, x_u) = x_1 \cdots x_u$ is considered in the next chapter .

7.4.3 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 from chapter 8 yield good controls $\mathbb{E}|S_n|^p$ and a central limit theorem may be derived by using the previous dependent Lindeberg CLT 1.3.3. In case the series of covariance is summable we already quoted that $\mathbb{E}S_m^2 \sim \sigma^2 m$ for large values of m.
The diea is to calculate

$$\Delta_n = \mathbb{E}\left(f\left(\frac{S_n}{\sqrt{n}}\right) - f(\sigma N)\right)$$

for function in the $\mathcal{C}^3$ class.
We need to derive $\Delta_n \to 0$ as $n \to \infty$.
To this end consider sequences

$$q = q(n) << p = p(n) << n \qquad \text{las} \qquad n \uparrow \infty$$

then set

$$\frac{S_n}{\sqrt{n}} = U_1 + \cdots + U_k + V$$

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 remaindier $\|V\|_2 \to 0$ because its cardinality is bounded above by $kq+p$. Because of the $1/\sqrt{n}$ normalization $\mathbb{E}V^2 \le C((p+kq)/n) \le C(p/n + q/p) \to 0$.
The variables U are almost independent since they are all distant at least q so that lemme 1.3.3 may be applied. ∎

To conclude cite a result adapted to causal cases and proved differently

Theorem 7.4.1 (Dedecker & Rio, 1998) *Let $(X_n)_{n\in\mathbb{Z}}$ be an ergodic stationary sequence with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$ and such that the following random series converges in $\mathbb{L}^1$:*

$$\sum_{n=0}^{\infty} X_0 \mathbb{E}\left(X_n | \sigma(X_k / k \le 0)\right)$$

Set $S_n = X_1 + \cdots + X_n$
Then the sequence $\mathbb{E}(X_0^2 + 2X_0 S_n)$ admits a limit σ^2 such that

$$\frac{1}{\sqrt{n}} S_{[nt]} \to_{n\to\infty} \sigma W_t, \qquad \text{in distribution in the Skorohod space } D([0,1]).$$

This result yields analogue CLT such as in [7] for the case of the function ψ_1'. Conditions as previously introduced are to take in mind as possible ones for all standard models in a statistician's mind.

Chapter 8

Moments and cumulants

The chapter is devoted to this tool 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

$$g(\lambda_2, \dots, \lambda_p) = \sum_{k_2=-\infty}^{\infty} \cdots \sum_{k_p=-\infty}^{\infty} \kappa(X_0, X_{k_2}, \dots, X_{k_p}) e^{-i(k_2\lambda_2+\cdots+k_p\lambda_p)}$$

Gaussian laws are characterized by the fact that cumulants with order > 2 vanish.

8.1 Method of moments

Recall that the method of moments yields limit theorems:

Theorem 8.1.1 (Feller) *If the sequence of real valued random variables U_n is such that*

$$\mathbb{E}U_n^p \to_{n\to\infty} \mathbb{E}U^p, \qquad \textit{for each integer } p \geq 0$$

then

$$U_n \to_{n\to\infty} U, \qquad \textit{in law}$$

If moreover U admits an analytic Laplace transform around 0. This holds if there exists $\alpha > 0$ with $\mathbb{E}e^{\alpha|U|} < \infty$.

Indeed from analycity the *analytic continuation theorem* implies that U's distribution is characterized through its moments.

8.2 Definitions

Let $Y = (Y_1, \ldots, Y_k) \in \mathbb{R}^k$ be a random vector we set

$$\begin{aligned}
\phi_Y(t) &= \mathbb{E}e^{it\cdot Y} = \mathbb{E}\exp\left(i\sum_{j=1}^k t_j Y_j\right), \\
m_p(Y) &= \mathbb{E}Y_1^{p_1}\cdots Y_k^{p_k}, \qquad \text{lorsque} \\
p &= (p_1, \ldots, p_k), \qquad t = (t_1, \ldots, t_k) \in \mathbb{R}^k, \\
|p| &= p_1 + \cdots + p_k = r, \qquad \mathbb{E}(|Y_1|^r + \cdots + |Y_k|^r) < \infty.
\end{aligned}$$

Notons $p! = p_1!\cdots p_k!$, $t^p = t_1^{p_1}\cdots t_k^{p_k}$ si $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$ et $p = (p_1, \ldots, p_k)$. In case the previous condition holds for some integer $r \in \mathbb{N}^*$ $t \mapsto \log\phi_Y(t)$ admits a Taylor expansion

$$\log\phi_Y(t) = \sum_{|p|\le r}\frac{i^{|p|}}{p!}\kappa_p(Y)t^p + o(|t|^r), \qquad \text{lorsque } t \to 0$$

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, $\ldots$, p_k repetitions for Y_k allows to consider $p = (1, \ldots, 1)$ and we set $\kappa_{(1,\ldots,1)}(Y) = \kappa(Y)$. If $\mu = \{i_1, \ldots, i_u\} \subset \{1, \ldots, k\}$

$$\kappa_\mu(Y) = \kappa(Y_{i_1}, \ldots, Y_{i_u}), \qquad m_\mu(Y) = m(Y_{i_1}, \ldots, Y_{i_u}).$$

Identifying Taylor expansions Leonov et Shyraev (1959) (voir Rosenblatt, 1985, page 33-34 [22]) derive the relations

$$\kappa(Y) = \sum_{u=1}^k (-1)^{u-1}(u-1)!\sum_{\mu_1,\ldots,\mu_u}\prod_{j=1}^u m_{\mu_j}(Y) \tag{8.1}$$

$$m(Y) = \sum_{u=1}^k \sum_{\mu_1,\ldots,\mu_u}\prod_{j=1}^u \kappa_{\mu_j}(Y) \tag{8.2}$$

Previous sums are taken over all the partitions $\mu_1, \ldots, \mu_u$ of the set $\{1, \ldots, k\}$. As $t \to 0$ Taylor formula with the expansion of $s \mapsto \log(1+s)$ at the origine successively yield

$$\phi_Y(t) = 1 + \sum_{0<|p|\le r}\frac{i^{|p|}}{p!}m_p(Y)t^p + o(|t|^r),$$

$$\log \phi_Y(t) = \sum_{u=1}^{r} \frac{(-1)^{u-1}}{u} \left(\sum_{0<|p|\le r} \frac{i^{|p|}}{p!} m_p(Y) t^p \right)^u + o(|t|^r)$$
$$= \sum_{u=1}^{r} \frac{(-1)^{u-1}}{u} \sum_{0<|p|\le r, p_1+\cdots+p_u=p} \frac{(it)^{|p|}}{p!} \prod_{j=1}^{u} m_{p_j}(Y) + o(|t|^r)$$

hence identify the coefficient corresponding to $p = (1, \ldots, 1)$ les $u-$tuples such that $p_1+\cdots+p_u = p$ (also given by partitions of $\{1, \ldots, k\}$ up to a conbinatorical coefficient $u!$ corresponding to the number of their permutations; choose $r = k$ to derive relation (8.1). Indeed then $|p| = k, p! = 1$ and $(it)^p = i^k t^k$.

Recall now some notions from Saulis et Statulevicius (1991)'s book.

Definition 8.2.1 *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*

$$c(\xi_j, \xi_{j-1}, \ldots, \xi_1) = \xi_j \overbrace{c(\xi_{j-1}, \ldots, \xi_1)} = \xi_j \left(c(\xi_{j-1}, \ldots, \xi_1) - \mathbb{E} c(\xi_{j-1}, \ldots, \xi_1) \right)$$

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 result from Saulis and Statulevicius (1991) explains the nature ofumulants. This is a representation in terms of centered moments.

Theorem 8.2.1 (Saulis, 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 partition satisfy*

$$N(k, u) = \sum_{\mu_1, \ldots, \mu_u} N_u(\mu_1, \ldots, \mu_u) = \sum_{j=1}^{u-1} C_k^j (u-j)^{k-1},$$

$$\sum_{u=1}^{k} N(k, u) = (k-1)!.$$

The following bound is a simple consequence of this result

Lemma 8.2.1 *Let $Y_1,\dots,Y_k\in\mathbb{R}$ 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*

$$\begin{aligned} |\kappa(Y_1,\dots,Y_k)| &\leq M_k, && (8.3)\\ M_kM_l &\leq M_{k+l},\ \textit{for } k,l\geq 2. && (8.4)\end{aligned}$$

Hence:

$$\prod_{i=1}^{u}|\kappa_p(Y_1,\dots,Y_{p_u})|\leq M_{p_1+\cdots+p_u} \qquad (8.5)$$

Proof of lemma 8.2.1. 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 8.2.1 and of the forthcoming lemma

Lemma 8.2.2 *For each $j,p\geq 1$ and for all the real valued random variables*

$$\|c(\xi_j,\xi_{j-1},\dots,\xi_1)\|_p\leq 2^j\max_{1\leq i\leq j}\|\xi_i\|_{pj}^j \qquad (\textit{with } \|\xi\|_q=\mathbb{E}^{1/q}|\xi|^q)$$

Proof of lemma 8.2.2. Jensen inequality leads to

$$\|c(\xi_1)\|_p\leq\|\xi_1\|_p+|\mathbb{E}\xi_1|\leq 2\|\xi_1\|_p,$$

Set $Z_j=c(\xi_j,\xi_{j-1},\dots,\xi_1)$ then $Z_j=\xi_j(Z_{j-1}-\mathbb{E}Z_{j-1})$ and from Hölder inequality

$$\|\xi_jZ_{j-1}\|_p^p\leq\|\xi_j\|_{pj}^p\|Z_{j-1}\|_{pj/(j-1)}^p$$

Thus using the recursion assumption for the pair $(q,j-1)$ where $q=pj/(j-1)$ Minkowski and Hölder inequalities give

$$\begin{aligned}\|Z_j\|_p &\leq \|\xi_jZ_{j-1}\|_p+\|\xi_j\|_p|\mathbb{E}Z_{j-1}|\\ &\leq 2\|\xi_j\|_{pj}\|Z_{j-1}\|_q, \qquad q=p\frac{j}{j-1}\\ &\leq 2^j\|\xi_j\|_{pj}\max_{0\leq i<j}\|\xi_i\|_{q(j-1)}^{j-1}\\ &\leq 2^j\max_{0\leq i\leq j}\|\xi_i\|_{pj}^j\end{aligned}$$

because $q(j-1)=pj$ which concludes. ∎

Proof of lemma 8.2.1. Omit suprema by replacing $\max_{j\leq J}\|Y_j\|_p$ by $\|Y_0\|_p$ for the sake of simplicity. Lemma 8.2.2 yields $|\ \widehat{\mathbb{E}}\ Y_\mu|\leq 2^{l-1}\|Y_0\|_l^l$ with $l=\#\mu$. Indeed write $Z=c(Y_2,\dots,Y_l)$ and define p through the identity $\frac{1}{p}+\frac{1}{l}=1$. Then

$$\left|\widehat{\mathbb{E}}\ (Y_1,\dots,Y_l)\right|=|\mathbb{E}Y_1Z|\leq\|Y_0\|_l\|Z\|_p\leq 2^{l-1}\|Y_0\|_l^l$$

since $p(l-1)=l$. Theorem 8.2.1 implies

$$|\kappa(Y)| \leq \sum_{u=1}^{k}\sum_{\mu_1,\dots,\mu_u}N_u(\mu_1,\dots,\mu_u)\prod_{i=1}^{u}2^{\#\mu_i-1}\|Y_0\|_{\#\mu_i}^{\#\mu_i}$$

$$\begin{aligned} &\le \sum_{u=1}^{k} 2^{k-u} N(k,u) \|Y_0\|_k^k \\ &\le 2^{k-1} \|Y_0\|_k^k \sum_{u=1}^{k} N(k,u) \\ &= 2^{k-1}(k-1)! \|Y_0\|_k^k. \qquad \blacksquare \end{aligned}$$

8.3 Dependence and cumulants

The following lemmas are essentially proved in Doukhan & León (1989) [11] for sequences of real valued random variables $(X_n)_{n\in\mathbb{Z}}$.
Consider a stationary real valued sequence $(X_n)_{n\in\mathbb{Z}}$ as Doukhan & Louhichi (1999) [12] set

$$c_{X,q}(r) = \max_{1\le l<q} \sup_{\substack{t_1 \le \cdots \le t_q \\ t_{l+1} - t_l \ge r}} \left|\mathrm{Cov}\left(X_{t_1}\cdots X_{t_l}, X_{t_{l+1}}\cdots X_{t_q}\right)\right| \tag{8.6}$$

Example 8.3.1 *Under weak dependence faible (7.5) associated with the functional ψ_1 and the classes of function $\mathcal{F} = \mathcal{G} = \mathcal{L}$. If $Y_i = h(X_i)$ Lipschitz function h bounded by M we get obtient $c_{Y,q}(r) \le M^{q-1} Lip(h)\theta_r$.*

The followinge coefficients are also useful

$$c^{\star}_{X,q}(r) = \max_{1\le l\le q} c_{X,l}(r)\mu_{q-l}, \qquad \text{with } \mu_t = \mathbb{E}|X_0|^t. \tag{8.7}$$

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 weak dependence 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 8.3.1 *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 another dependence coefficient*

$$\kappa_p(r) = \max_{\substack{t_1 \le \cdots \le t_p \\ r(t_1,\ldots,t_p) \ge r}} \left|\kappa_p\left(X_{t_1},\ldots,X_{t_p}\right)\right| \tag{8.8}$$

Lemma 8.3.1 *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 8.2.1 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 8.3.1. 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 $\nu_\mu^- = [1, l] \cap \nu_\mu \ne \emptyset$ and $\nu_\mu^+ = [l+1, Q] \cap \nu_\mu \ne \emptyset$. From formula (8.2) we obtain with $\eta = \{1, \ldots, l\}$,

$$\kappa(X_{k_1}, \ldots, X_{k_Q}) = \mathrm{Cov}\,(X_{\eta(k)}, X_{\overline{\eta}(k)}) - \sum_u \sum_{\{\mu\}} \kappa_{\nu_\mu(k)} K_{\mu,k}, \tag{8.9}$$

with $K_{\mu,k} = \prod_{\mu_i \ne \nu_u} \kappa_{\mu_i(k)}$ and the 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

$$r\big(\nu_\mu(k)\big) \ge r(k)$$

derive $|\kappa_{\nu_\mu(k)}| \le \kappa_{X,\#\nu_\mu}(r)$.
This operation let the size of lags augment. With lemma 8.2.1 we arrive to $|M_\mu| \le M_{Q-\#\mu_\nu}$ as in (8.5) 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 serie with an integral. ∎

Write lemma 8.3.1 as

$$\kappa_{X,Q}(r) \le c_{X,Q}(r) + \sum_{s=2}^{Q-2} B_{Q,s}\kappa_{X,s}(r),$$

then

$$\begin{aligned}
\kappa_{X,2}(r) &\le c_{X,2}(r), \\
\kappa_{X,3}(r) &\le c_{X,3}(r), \\
\kappa_{X,4}(r) &\le c_{X,4}(r) + B_{4,2}\kappa_{X,2}(r) \\
&\le c_{X,4}(r) + B_{4,2}c_{X,2}(r),
\end{aligned}$$

$$\begin{aligned}
\kappa_{X,5}(r) &\le c_{X,5}(r) + B_{5,3}\kappa_{X,3}(r) + B_{5,2}\kappa_{X,2}(r)\\
&\le c_{X,5}(r) + B_{5,3}c_{X,3}(r) + B_{5,2}c_{X,2}(r),\\
\kappa_{X,6}(r) &\le c_{X,6}(r) + B_{6,4}\kappa_{X,4}(r) + B_{6,3}\kappa_{X,3}(r) + B_{6,2}\kappa_{X,2}(r)\\
&\le c_{X,6}(r) + B_{6,4}\left(c_{X,4}(r) + B_{4,2}c_{X,2}(r)\right) + B_{6,3}c_{X,3}(r) + B_{6,2}c_{X,2}(r)\\
&\le c_{X,6}(r) + B_{6,4}c_{X,4}(r) + B_{6,3}c_{X,3}(r) + (B_{6,2} + B_{6,4}B_{4,2})c_{X,2}(r).
\end{aligned}$$

Lemma 8.3.1 impies the following important corollary derived from a recursion with the previous inequalities.

Corollary 8.3.1 *For each $Q \ge 2$ there exists a constante $A_Q \ge 0$ only depending on Q and such that*

$$\kappa_{X,Q}(r) \le A_Q c^*_{X,Q}(r).$$

Remark 8.3.1

- *This lemme prove the equivalence between coefficients $c_{X,Q}(r)$ and $\kappa_Q(r)$ Precise upper bounds follow from theorem 8.2.1. Decompose for it the sums corresponding to centered moments in 2 terms among which one explicitly depends on the maximal lag.*

  Formula (8.9) implies with $B_{Q,Q} = 1$,

  $$c_{X,Q}(r) \le \sum_{s=2}^{Q} B_{Q,s}\kappa_{X,s}(r)$$

  Thus there exists a constante $\widetilde{A}_Q$ with

  $$c_{X,Q}(r) \le \widetilde{A}_Q \kappa^*_{X,Q}(r), \quad \kappa^*_{X,Q}(r) = \max_{2\le l\le Q} \kappa^*_{X,l}(r)\mu_{Q-l}$$

  Hence constants $a_Q, A_Q > 0$ satisfy

  $$a_Q c^*_{X,Q}(r) \le \kappa^*_{X,Q}(r) \le A_Q c^*_{X,Q}(r)$$

  For fixes Q those coefficients are equivalent.

- *The previous formula (8.9) 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 $\alpha \subset \{1, \ldots, l\}$, $\beta \subset \{l+1, \ldots, Q\}$ 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 (8.9) 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.

Examples. *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 8.3.1 to bound high order cumulants.

8.3.1 Sums of cumulants

The previous bounds yield

Lemma 8.3.2 *Let*

$$\kappa_Q = \sum_{k_2=0}^{\infty} \cdots \sum_{k_Q=0}^{\infty} \left|\kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right)\right|, \tag{8.10}$$

with notation (8.7) for each $Q \ge 2$ there exists a constant B_Q such that

$$\kappa_Q \le B_Q \sum_{r=0}^{\infty} (r+1)^{Q-2} C^*_{X,Q}(r).$$

Proof of lemma 8.3.2. Decompose sums

$$\begin{aligned}
\kappa_Q &\le (Q-1)! \sum_{k_2 \le \cdots \le k_Q} \left|\kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right)\right|\\
&\equiv (Q-1)!\, \widetilde{\kappa}_Q
\end{aligned}$$

considering the following partition of the index set

$$E = \{k = (k_2, \ldots, k_Q) \in \mathbb{N}^{Q-1} / k_2 \le \cdots \le k_Q\}$$

as $E_r = \{k \in E / r(k) = r\}$ for $r \ge 0$ (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

$$\sum_{k \in E_r} \left|\kappa\left(X_0, X_{k_2}, \ldots, X_{k_Q}\right)\right| \le A_Q \# E_r C^*_{X,Q}(r)$$

for a constant $A_Q > 0$ and the elementary bound $\# E_r \le (Q-1)(r+1)^{Q-2}$ yields the result. ∎

8.3.2 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

$$\frac{1}{\sqrt{n}}\,(X_1+\cdots+X_n) \longrightarrow_{n\to\infty} \mathcal{N}(0,\sigma^2), \qquad \text{en loi}$$

The behavior of moments in $\mathbb{L}^p-$norm is of importance. Cumulants allows an elementary approach of such expressions.

Lemma 8.3.3 *If the series (8.10) are summable for each $Q \le p$ set $q = [p/2]$ ($q = p/2$ for p even and $q = (p-1)/2$ else then*

$$\left|\mathbb{E}\left(\sum_{j=1}^n X_j\right)^p\right| \le \sum_{u=1}^q n^u \gamma_u, \qquad \textit{with} \tag{8.11}$$

$$\gamma_u = \sum_{v=1}^{2q} \sum_{p_1+\cdots+p_u=p} \frac{p!}{p_1!\cdots p_u!}\kappa_{p_1}\cdots\kappa_{p_u}$$

Proof. As in Doukhan & Louhichi (1999) [12] bound

$$\begin{aligned}|\mathbb{E}(X_1+\cdots+X_n)^p| &= \left|\sum_{1\le k_1,\dots,k_p\le n} \mathbb{E}X_{k_1}\cdots X_{k_p}\right| \\ &\le p!A_{p,n} \equiv p! \sum_{1\le k_1,\dots,k_p\le n} \left|\mathbb{E}X_{k_1}\cdots X_{k_p}\right|\end{aligned}$$

Let also $\mu = \{i_1,\dots,i_v\} \subset \{1,\dots,p\}$ and $k=(k_1,\dots,k_p)$ set

$$\mu(k) = (k_{i_1},\dots,k_{i_v}) \in \mathbb{N}^v \tag{8.12}$$

To enumerate the terms with their multiplicity this is simpler to consider multi-indices than partitions. Cumulants et moments are defined analogously. As in Doukhan & León (1989) [11] with formula (8.2) and partitions $\mu_1,\dots,\mu_u$ of $\{1,\dots,p\}$ with exactly $1\le u\le p$ elements,

$$\begin{aligned}A_{p,n} &= \sum_{1\le k_1,\dots,k_p\le n} \sum_{u=1}^p \sum_{\mu_1,\dots,\mu_u} \prod_{j=1}^u \kappa_{\mu_j(k)}(X) \\ &= \sum_{u=1}^p \sum_{\mu_1,\dots,\mu_u} \sum_{1\le k_1,\dots,k_p\le n} \prod_{j=1}^u \kappa_{\mu_j(k)}(X) \\ &= \sum_{r=1}^p \sum_{p_1+\cdots+p_r=p} \frac{p!}{p_1!\cdots p_r!} \prod_{u=1}^r \sum_{1\le k_1,\dots,k_{p_u}\le n} \kappa_{p_u}(X_{k_1},\dots,X_{k_{p_u}}) \qquad (8.13) \\ |A_{p,n}| &\le \sum_{u=1}^q n^u \sum_{p_1+\cdots+p_u=p} \frac{p!}{p_1!\cdots p_u!} \prod_{j=1}^u \kappa_{p_j} \qquad (8.14)\end{aligned}$$

Identity (8.13) 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 t $p_1,\ldots,p_u\ge 2$ et donc $2u\le p$ hence $u\le q$ and we get (8.14). ∎

Remark 8.3.2 *If their exists $C>0$ with $\kappa_s\le C^s$ for each $s\le p$ then the bound (8.14) simply writes $C^p\sum_{u=1}^{q}u^pn^u$ due to the multinomial identity.*

8.3.3 Rosenthal inequality

Again as in Doukhan et Louhichi (1999) [12] 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 [12].
Iterating the previous relation yields

$$\begin{aligned}
A_{2,n}&\le nC_{2,n}\\
A_{3,n}&\le 2nC_{3,n}\\
A_{4,n}&\le 3nC_{4,n}+A_{2,n}^2\le 3nC_{4,n}+n^2C_{2,n}^2\\
A_{5,n}&\le 4nC_{5,n}+2A_{2,n}A_{3,n}\le 5nC_{5,n}+4n^2C_{2,n}C_{3,n}\\
A_{6,n}&\le 5nC_{6,n}+2A_{2,n}A_{4,n}+A_{3,n}^2\le 5nC_{6,n}+2n^2\left(2C_{3,n}^2+3C_{2,n}C_{4,n}\right)+8n^3C_{2,n}^3
\end{aligned}$$

We denote (for a fixed q)

$$C_{m,n} = \sum_{k=0}^{n-1} (r+1)^{m-2} c_{X,q}(r)$$

Generally if $p = 2q$ or $p = 2q+1$ we obtain

$$A_{p,n} \leq \sum_{j=1}^{q} c_{j,n} n^j$$

with $c_{j,n}$ a polynomial wrt expressions $C_{i,n}$ for $i \leq j$.
Precisely this is a linear combination of expressions $\prod_{s=1}^{t} C_{i_s,n}$ with $i_1 + \cdots + i_t = j$. Hence for $c_{X,p}(r) = \mathcal{O}(r^{-q})$ one deduces Marcinkiewicz-Zygmund inequality

$$|\mathbb{E}(X_1 + \cdots + X_n)^p| = \mathcal{O}(n^q)$$

Rosenthal type inequalities yields sharp bonds for centered moments of kernel density estimators or for the empirical process.
If the marginals of the stationary process (X_n) admit a density f. 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. ψ_1-weakly dependent.
This is easy to prove that

$$\begin{aligned} \operatorname{Lip} h &\leq l2^{p-1}\frac{\operatorname{Lip} K}{h_n}, \qquad \text{if we denote:} \\ h(t_1, \ldots, t_l) &= \prod_{j=1}^{l} \left(\frac{t_j - x}{h_n}\right) - \mathbb{E}K\left(\frac{X_j - x}{h_n}\right) \end{aligned}$$

Suppose now that for each $n > 0$ the joint density $f_n(x,y)$ of the couple (X_0, X_n) exists and

$$\|f_n(\cdot,\cdot)\|_\infty \leq M$$

From an integration one then get $\|f(\cdot)\|_\infty \leq M$ and

$$c_{U,p}(0) \leq 2^p f(x) \int K^2(s)ds$$

A direct calculation coupled with a weak dependence inequality yield 2 distinct controls of $c_{U,p}(r)$ pour $r > 0$

$$c_{U,p}(r) \leq 2^{p-1}\left(p\operatorname{Lip} K \frac{\theta_r}{h_n}\right) \wedge (2Mh_n^2)$$

thus there exists a constant $C > 0$ with

$$C_{p,n} = Ch_n \left(1 + \sum_{k=1}^{n-1} (r+1)^{p-2} \left(h_n \wedge \frac{\theta_r}{h_n^2} \right) \right)$$

Exercise 1 *As an exercise prove that*

$$C_{p,n} = \mathcal{O}\left(h_n\right)$$

if

$$\sum_{r=0}^{\infty} (r+1)^{p-2} \theta_r^{1/3} < \infty \tag{8.15}$$

using inequality $\theta/h^2 \wedge h \leq \theta^{1/3}$.

From recursion and using assumption .(8.15) and exercise 1 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 $(nh_n)^{-(p-1)/2}$ else. This agrees with the underlying CLT

$$\sqrt{nh_n}(\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)) \to_{n\to\infty} \mathcal{N}\left(0, f(x) \int K^2(t)dt\right)$$

Almost sure convergence of such estimates also follow from Markov inequality and Borel Cantelli lemma

$$\sum_{n=1}^{\infty} \frac{1}{(nh_n)^{p/2}} < \infty$$

for some even integer $p > 2$.

Exercise 2 *Extend this whole section to the case of associated processes. In particular precise how eqn.(8.15) should be modified in this case.*

Bibliography

[1] Ango Nze, P., Bühlmann, P., Doukhan, P. (2002) Weak dependence beyond mixing and asymptotics for non parametric regression. Ann. Statist. 30-2, 397-430.

[2] Azencott, R., Dacunha-Castelle, D. (1984) **Séries d'observations irrégulières; modélisation et prévisions.** Masson.

[3] Breuer, P., Major, P.(1983) Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13, 425-441.

[4] Brockwell, P. J., Davis, R. A. (1991) **Time Series: Theory and Methods** (second edition). Springer Series in Statistics.

[5] Choquet, G.(1973) **Topologie**, volume II, Masson ed.

[6] Davydov, Yu. (1970) Invariance principle for stationary processes. *Theor. Probab. Appl.* 15-3, 498-509.

[7] Dedecker, J., Doukhan, P. (2003) A new covariance inequality and applications, *Stoch. Proc. Appl.* 106, 63-80.

[8] Dedecker, J., Doukhan, P., León, J., Louhichi, S., Prieur, C. (2007) **Weak dependence, models and applications.** Lecture Notes in Statistics 190, Springer Verlag.

[9] Doukhan, P. (1994) **Mixing, properties and examples.** Lecture Notes in Statistics 85, Springer Verlag.

[10] Doukhan, P., Oppenheim, G., Taqqu, M. (ed.) (2002) **Theory and Applications of Long-Range Dependence.** Birkhaüser, 713 pages.
see eg. papers:
Doukhan, P. "Limit theorems for stationary sequences" and,
Taqqu "Fractional Brownian motion and long range dependence".

[11] Doukhan, P., León, J.R. (1989) Cumulants for mixing sequences and applications to empirical spectral density. Probab. Math. Stat. 10.1, 11-26.

[12] Doukhan, P., Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. Stoch. Proc. Appl. 84, 313-342.

[13] Doukhan, P., Lang, G. Surgailis, D. (2002) Functional CLTs for short or long memory linear sequences. Festschrift D. Dacunha-Castelle, J. Bretagnolle and I. Ibragimov. Ann. I.H.P. B. special issue 38-6, 879-896.

[14] Dobrushin, R. L., Major, P. (1979). Non-central limit theorems for non-linear functions of Gaussian fields. *Z. Wahrsch. und verw. Gebiete*, 50, 27-52.

[15] Duflo, M. (1990) **Méthodes récursives aléatoires.** Masson, Paris. (Edition anglaise réactualisée: Springer-Verlag, 1996)

[16] Giraitis, L. , Surgailis D. (1989). A central limit theorem for polynomial forms I. *Liet. matem. rink.* 29(2), 266-289.

[17] Giraitis L., Koul H. L., Surgailis D. (1996) Asymptotic normality of regression estimators with long memory errors *Statist. Probab. Letters* 29 , 317-335.

[18] Ho, H.-C., Hsing, T. (1996) On the asymptotic expansion of the empirical process of long memory moving averages. *Ann. Stat.* **24**, 992–1024.

[19] Hsieh, M., Hurvich, C. M., Soulier, T. (2007) Asymptotic properties for duration driven long range dependent processes. *To appear in Journal of Econometrics.*

[20] Kallenberg, O. (1997) **Foundations of modern probability.** Springer-Verlag.

[21] Nourdin, I., Peccati, G., Podolskij, M. (2011) Quantitative Breuer-Major theorems. Stochastic Processes and their Applications 121-4, pages 793-812.

[22] Rosenblatt, M. (1985) **Stationary sequences and random fields.** Birkhäuser.

[23] Rosenblatt, M. (1991) **Stochastic curve estimation.** NSF-CBMS Regional Conference Series in Probability and Statistics Volume3.

[24] Saulis L., Statulevicius, V.A. (1991) **Limit theorems for large deviations.** Birkhäuser.

[25] 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. pages 606–616.

[26] Soulier (2001) Moment bounds and central limit theorem for functions of Gaussian vectors. Statistics & Probability Letters 54-2, 193-203.

[27] van der Vaart, A., Wellner, J. A. (1998). **Weak convergence and empirical processes.** Springer Verlag series in statistics.