A Review on some weak dependence conditions

Benjamin Bobbia, Paul Doukhan, Xiequan Fan

A Review on some weak dependence conditions

Benjamin Bobbia[1], Paul Doukhan[1] and Xiequan Fan[2]

[1] *CY University, AGM UMR 8088, site Saint-Martin, 95000 Cergy-Pontoise, France*

[2] *Center for Applied Mathematics, Tianjin University, Tianjin 300072, China*

1. Introduction

This paper is a short survey over some dependence measures useful to deal with for time series analysis. It may be seen as a simple toolbox to deal with some dependence questions. Indeed to derive statistical bounds for the validity of procedures one needs more control of the dependence between past and future. The aim of weak dependence structure is to derive results analogue to those, classical obtained under independence. As this is generally assumed we mainly consider stationary models and under such assumptions a main assumption is ergodicity which entails the Strong Law of Large Numbers. The latter result may be regarded as the basic consistency result in statistics which proves the convergence of the basic empirical estimate for the mean

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n),$$

for a given stationary time series $(X_t)_{t\in\mathbb{Z}}$. Section 2 is devoted to a rapid tour on some dependence conditions suitable for time series analysis. The consistency of such estimates is considered through moment and probability inequalities in Section 3. A nice reference is the short monograph [35] which adresses several issues of non parametric estimation and [13] provides minimax viewpoints based on wavelets shrinkage. A related question is to know more precisely with quality of such estimates and Section 4 makes a rapid tour of Central Limit results necessary to deal to get asymptotic confidence bounds. The Donsker variants of the CLT which describe the behaviour of partial sums processes are useful for change point analysis. A last section 5 deals with functional central limit theorems usually necessary to deal with general contrast estimation techniques such as mean squares techniques, QMLE (quasi maximum likelihood) or Whittle estimates (periodogram based). Those techniques are mathematically heavy and we describe them in details to make this section useful for practitioners. Such results allow to deal with many types of estimates and the reference [9] gives a tour on such questions, see also [2], [1], and [14] for various applications.

Our idea is not to provide a reader with a complete survey but only to give some few hints of how to deal with dependence questions in the statistical context. This is why we restrained to some few of items and we insist a bit more on some complex issues.

2. Some weak dependence conditions

2.1. Mixing conditions

A natural idea to deal with dependence in a process is to assume that two events occurring far from each other in the process are almost independent. For a process $(X_t)_{t\in\mathbb{Z}}$ this roughly consist in assuming that, for fixed t, the behavior of X_t influence less and less X_{t+r} as long as r increase. This influence is measured in terms of mixing coefficients, of which the most famous are detailed in this section. The mixing conditions introduced by Rosenblatt [34] are weak dependence conditions stated in terms of σ algebra.

Definition 2.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A map $c : \mathcal{A}\times\mathcal{A} \to [0, +\infty]$ is called mixing coefficient if for every independent σ-algebra $\mathcal{U},\ \mathcal{V} \subset \mathcal{A}$, $c\,(\mathcal{U}, \mathcal{V}) = 0$.

Among all possible mixing coefficients, the following five are commonly used.

- Strong mixing coefficient [34]

$$\alpha(\mathcal{U}, \mathcal{V}) = \sup_{U\in\mathcal{U}, V\in\mathcal{V}} |\mathbb{P}(U \cap V) - \mathbb{P}(U)\mathbb{P}(V)|.$$

- Absolute regularity coefficient [39]

$$\beta(\mathcal{U}, \mathcal{V}) = \frac{1}{2}\sup_{I,\, J\in\mathbb{N}} \sup_{\substack{(V_j)_{1\le j\le J}\in\mathcal{V}^J \\ (U_i)_{1\le i\le I}\in\mathcal{U}^I,}} \sum_{i=1}^{I}\sum_{j=1}^{J} |\mathbb{P}(U_i \cap V_j) - \mathbb{P}(U_i)\mathbb{P}(V_j)|,$$

where $(U_i)_{1\le i\le I} \in \mathcal{U}^I$ and $(V_j)_{1\le j\le J} \in \mathcal{V}^J$ are partitions of Ω. This coefficient can be stated in a more compact way as

$$\beta(\mathcal{U},\mathcal{V}) = \|\mathbb{P}_{\mathcal{U}\otimes\mathcal{V}} - \mathbb{P}_{\mathcal{U}} \otimes \mathbb{P}_{\mathcal{V}}\|_{TV}.$$

- Maximal correlation coefficient [25]

$$\rho(\mathcal{U},\mathcal{V}) = \sup\{|corr(X,Y)| : \ X \in \mathbb{L}^2(\mathcal{U}),\ Y \in \mathbb{L}^2(\mathcal{V})\}.$$

- Uniform mixing coefficient [24]

$$\phi(\mathcal{U},\mathcal{V}) = \sup_{(U,V)\in\mathcal{S}} \left| \frac{\mathbb{P}(U\cap V)}{\mathbb{P}(U)} - \mathbb{P}(V) \right|,$$

  where $\mathcal{S} = \{(U,V) \in \mathcal{U}\times\mathcal{V} : \ \mathbb{P}(U) > 0\}$.
- ψ-mixing coefficient [6]

$$\psi(\mathcal{U},\mathcal{V}) = \sup_{(U,V)\in\mathcal{H}} \left| \frac{\mathbb{P}(U\cap V)}{\mathbb{P}(U)\mathbb{P}(V)} - 1 \right|, \tag{1}$$

  where $\mathcal{H} = \{(U,V) \in \mathcal{U}\times\mathcal{V} : \ \mathbb{P}(U) > 0, \mathbb{P}(V) > 0\}$.

The strong mixing coefficient for a random process $X = (X_t)_{t\in\mathbb{Z}}$ is defined by, for $r > 0$,

$$\alpha(r) := \alpha_X(r) = \sup_{i\in\mathbb{Z}} \alpha\left(\sigma(X_t, t\le i), \sigma(X_t, t \ge i+r)\right). \tag{2}$$

The random process $(X_t)_{t\in\mathbb{Z}}$ is called α-mixing if $\alpha(r) \underset{r\to\infty}{\longrightarrow} 0$. Accordingly, we shall use below the notations $\beta(r), \rho(r), \phi(r)$ and $\psi(r)$, which mean sequences defined by (2) with α replaced by β, ρ, ϕ and ψ, respectively.

Those conditions are related to the following diagram

$$\psi\text{-}mixing \Rightarrow \phi\text{-mixing} \Rightarrow \begin{array}{c}\rho\text{-mixing}\\ \beta\text{-mixing}\end{array} \Rightarrow \alpha\text{-mixing}.$$

The relationships between those mixing conditions are studied together with examples of models in [14] and converse implications always fail to hold. Moreover examples of such mixing models may be found in the same reference.

2.2. Weak dependence conditions

Even if the mixing coefficient previously introduced are useful to derive asymptotic results for dependent sequences, in practice it is sometime difficult to establish that a given model satisfies a mixing condition. That is why a more tractable weak dependence conditions were introduced. The first family of conditions (Definitions 2.2 - 2.6) also relies on the fact that the dependence between the "past" and the "future" of a time series decreases with the distance. However, it is focused on covariance between past and future rather than corresponding σ-algebra. Whereas the second family (Definitions 2.7 and 2.8) is concerned by the difference between the distribution of the future and the distribution of the future given the past.

Definition 2.2. ([15]) The sequence $(X_t)_{t\in\mathbb{Z}}$ is $(\psi_0, \mathcal{F}, \mathcal{G}, \epsilon_r)$-weakly dependent if there exists a sequence $\epsilon_r := (\epsilon_r)_{r\in\mathbb{N}}$ converging to zero at infinity and a function ψ_0 with arguments $(f,g) \in \mathcal{F}\times\mathcal{G}$, $f : \mathbb{R}^u \to \mathbb{R}$, $g : \mathbb{R}^v \to \mathbb{R}$ such that for any $(i_1,\dots,i_u)$ and $(j_1,\dots,j_v)$ with $i_1 \le \cdots \le i_u < j_1 \le \cdots \le j_v$ and $j_1 - i_u \ge r$, one has

$$|cov\left(f(X_{i_1},\dots,X_{i_u}), g(X_{j_1},\dots,X_{j_v})\right)| \le \psi_0(f,g)\epsilon_r.$$

In this definition, it is required that any function $f \in \mathcal{F}\cup\mathcal{G}$ is a form from a finite dimensional vector space. Is that, there exists $d_f \in \mathbb{N}^*$ such $f : \mathbb{R}^{d_f} \to \mathbb{R}$. Note that the dimension d_f depend on f and two functions on $\mathcal{F}$ may be defined on vector spaces with different dimensions. Note also that $\mathcal{F}$ and $\mathcal{G}$ are classes of functions without any structure assumptions.

A natural way to describe different types of weak dependence is specifying such classes $\mathcal{F}$ and $\mathcal{G}$ but also the application ψ_0.

First consider that $\mathcal{F} = \mathcal{G}$ be the space of bounded Lipschitz functions with uniform norm bounded by one, namely

$$\mathcal{F} = \Lambda^{(1)} := \{f \in L^\infty : \ Lip(h) < 1,\ \|f\|_\infty \le 1\}.$$

This class is used together with function ψ_0, defined on $(\Lambda^{(1)})^2$, such that

$$\psi_0(f,g) := \nu(d_f, d_g)\mu(Lip(f), Lip(g))$$

with ν and μ respectively defined on $\mathbb{N}^2$ and $\mathbb{R}_+^2$. We will see in this section that the choice of ψ_0 is of particular interest.

Definition 2.3 (η-dependence condition)**.** Consider the function

$$\psi_1(f,g) = d_f Lip(f) + d_g Lip(g).$$

The random process $(X_n)_{n\in\mathbb{Z}}$ is η-weakly dependent (or just η-dependent) if it is $(\psi_1, \Lambda^{(1)}, \Lambda^{(1)}, \epsilon_r)$-weakly dependent. In this case, we denote the sequence ϵ_r by η_r.

Definition 2.4 (κ-dependence condition)**.** Consider the function

$$\psi_2(f,g) = d_f d_g Lip(f) Lip(g).$$

The random process $(X_n)_{n\in\mathbb{Z}}$ is κ-weakly dependent (or just κ-dependent) if it is $(\psi_2, \Lambda^{(1)}, \Lambda^{(1)}, \epsilon_r)$-weakly dependent. In this case, we denote the sequence ϵ_r by κ_r.

Those definitions can be extended to for any integer $j > 2$ setting

$$\psi_j(f,g) = \left(d_f Lip(f) + d_g Lip(g)\right)^j.$$

Secondly we may consider the classes of functions $\mathcal{F}$ of bounded functions with respect to the uniform norm and $\mathcal{G} = \Lambda^{(1)}$ or the class of 1-bounded Lipschitz functions.

Definition 2.5 (θ-dependence condition)**.** Consider the function

$$\psi'(f,g) = d_g Lip(g).$$

The random process $(X_n)_{n\in\mathbb{Z}}$ is θ-weakly dependent (or just θ-dependent) if it is $(\psi', \mathcal{F}, \mathcal{G}, \epsilon_r)$-weakly dependent. In this case, we denote the sequence ϵ_r by θ_r.

This definition is a particular case of a more general definition holding in a causal case. We can refer to section 2.3 of [9] for more details about the general definition of the coefficient θ_r.
In this framework, we can define a notion of weak convergence which include the cases η and κ.

Definition 2.6 (λ-dependence conditions)**.** Consider the function

$$\tilde{\psi}(f,g) = d_f d_g Lip(f) Lip(g) + d_f Lip(f) + d_g Lip(g).$$

The random process $(X_n)_{n\in\mathbb{Z}}$ is λ-weakly dependent (or just λ-dependent) if it is $(\tilde{\psi}, \Lambda^{(1)}, \Lambda^{(1)}, \epsilon_r)$-weakly dependent. In this case, we denote the sequence ϵ_r by λ_r.

The τ-dependence coefficients were introduced in Dedecker and Prieur [11].

Definition 2.7 (τ-dependence conditions)**.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. We define the coefficient τ_p, for $d > 1$ and a random variable X, by:

$$\tau_p(\mathcal{M}, X) = \left\| \sup_{g\in\Lambda^{(1)}} \left(\int g(x)\mathbb{P}_{X|\mathcal{M}}(\mathrm{d}x) - \int g(x)\mathbb{P}_X(\mathrm{d}x) \right) \right\|_p,$$

where $\mathbb{P}_X$ and $\mathbb{P}_{X|\mathcal{M}}$ denote respectively the distribution of X and the conditional distribution of X given $\mathcal{M}$. In practice we consider the σ-algebras $\mathcal{M}_i = \sigma\left(X_j,\ j \le i\right)$ in order to introduce the coefficient $\tau_{p,k}(r)$ define by

$$\tau_{p,k}(r) = \max_{1\le l\le k} \frac{1}{l} \sup_{i+r\le j_1<\ldots,<j_l} \tau_p\left(\mathcal{M}_i, (X_{j_1}, \ldots, X_{j_l})\right).$$

We also recall Wu [40] and Wu and Shao [41]'s dependence structures.

Definition 2.8 (Physical dependence)**.** Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be i.i.d. random variables, and denote $\mathcal{F}_i = (..., \varepsilon_{i-1}, \varepsilon_i)$. Let $(\varepsilon'_i)_{i\in\mathbb{Z}}$ be an independent copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$ and $\mathcal{F}'_i = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, ..., \varepsilon_i)$ the coupled version of $\mathcal{F}_i$. Assume

$$X_i = g(..., \varepsilon_{i-1}, \varepsilon_i) \in L^p, \quad p > 0, \tag{3}$$

where g is a measurable function such that X_i is well-defined. Define

$$\theta_p(i) = \|X_i - X_i'\|_p \qquad \text{and} \qquad \Theta_{m,p} = \sum_{i=m}^{\infty} \theta_p(i),$$

where $X_i' = g(\mathcal{F}_i')$. Denote $\mathcal{F}_i' = (\ldots, \varepsilon_{i-1}', \varepsilon_i')$. Assume $X_i \in L^p, p > 2$, and define

$$\Delta_p(n) = \sup_i \|X_i - g(\mathcal{F}_{i-n}', \varepsilon_{i-n+1}, ..., \varepsilon_i)\|_p,$$

where g is a measurable function such that X_i is well-defined.

The relationships between those mixing conditions are studied together with examples of models in [9] and the relationships between all those coefficients are studied in details in this monograph. Moreover examples of weakly dependent models may be found in the same reference.

3. Moment and probability inequalities

We present some moment and probability inequalities for weak dependence sequences. In this section, let $(X_i)_{i\in\mathbb{Z}}$ be a sequence of centered real valued random variables, and let $S_n = \sum_{i=1}^n X_i, n \geq 1$, be the partial sum of $(X_i)_{i\in\mathbb{Z}}$.

Denote $\tau(x) = \sup_{k\geq 1} \tau_{1,k}(\lfloor x \rfloor)$, where $\lfloor x \rfloor$ denote the greatest integer lower than x. Assume that there exist three positive constants γ_1, a and c such that

$$\tau(x) \leq a \exp\{-c\, x^{\gamma_1}\}, \qquad x \geq 1, \tag{4}$$

and that, for two constants $\gamma_2 \in (0, \infty]$ and $b \in (0, \infty)$, the following tail condition is satisfied:

$$\sup_{k>0} \mathbb{P}(|X_k| > t) \leq \exp\{1 - (t/b)^{\gamma_2}\}, \qquad t > 0. \tag{5}$$

Notice that when $\gamma_2 = \infty$, $(X_k)_{k>0}$ are uniformly bounded. Suppose furthermore that $\gamma < 1$ and it is defined by

$$1/\gamma = 1/\gamma_1 + 1/\gamma_2. \tag{6}$$

Merlevède, Peligrad and Rio [27] have established the following Bernstein type inequality for τ-mixing sequences.

Theorem 3.1 ([27]). *Let*

$$V = \sup_{M>0} \sup_{i>0} \Big(var(\varphi_M(X_i)) + 2 \sum_{j>i} |cov(\varphi_M(X_i), \varphi_M(X_j))| \Big),$$

where $\varphi_M(x) = (x \wedge M) \vee (-M)$. Assume the conditions (4), (5) and (6). Then V is finite and, for any $n \geq 4$, there exist positive constants C_1, C_2, C_3 and C_4 depending only on c, γ and γ_1 such that it holds for any $x > 0$,

$$\begin{aligned} \mathbb{P}\Big(\sup_{1\leq j\leq n} |S_j| \geq x \Big) \quad &\leq \quad n \exp\left\{ -\frac{x^\gamma}{C_1} \right\} + \exp\left\{ -\frac{x^2}{C_2(1+nV)} \right\} \\ &\quad + \exp\left\{ -\frac{x^2}{C_3 n} \exp\left\{ \frac{x^{\gamma(1-\gamma)}}{C_4(\log x)^\gamma} \right\} \right\}. \end{aligned}$$

Liu, Xiao and Wu [26] obtained the following Rosenthal type inequality for physical dependence random variables.

Theorem 3.2 ([26]). *Let X_i be defined by (3). Assume $\mathbb{E}|X_1|^p < \infty, p > 2$. Then*

$$\begin{aligned} \|\max_{1\leq j\leq n} S_j\|_p \quad &\leq \quad n^{1/2} \left[\frac{87p}{\log p} \sum_{j=1}^n \theta_2(j) + 3(p-1)^{1/2} \sum_{j=n+1}^{\infty} \theta_p(j) + \frac{29p}{\log p} \|X_1\|_2 \right] \\ &\quad + n^{1/p} \left[\frac{87p(p-1)^{1/2}}{\log p} \sum_{j=1}^n j^{1/2-1/p} \theta_p(j) + \frac{29p}{\log p} \|X_1\|_p \right]. \end{aligned}$$

In particular, it implies that

$$\|\max_{1\leq j\leq n} S_j\|_p \quad \leq \quad c_p n^{1/2} (\Theta_{1,2} + \|X_1\|_2) + c_p n^{1/p} \left[\sum_{j=1}^{\infty} \min(j,n)^{1/2-1/p} \theta_p(j) + \|X_1\|_p \right].$$

Denote by $G_q(y)$ the following Gaussian-like tail function

$$G_q(y) = \sum_{j=1}^{\infty} \exp\{-j^q y^2\}, \qquad y > 0,\ q > 0.$$

Note that $\sup_{y \geq 1} G_q(y) e^{y^2} = G_q(1)e$. Hence if $y \geq 1, G_q(y) \leq G_q(1)e^{1-y^2}$. With the notation of $G_q(y)$, Liu, Xiao and Wu [26] proved the following Nagaev type inequalities.

Theorem 3.3 ([26]). *Let X_i be defined by (3).*

(i) *Assume that*

$$v := \sum_{j=1}^{\infty} \mu_j < \infty, \qquad \textit{where} \quad \mu_j = (j^{p/2-1}\theta_p^p(j))^{1/(p+1)}.$$

Then for any $x > 0$,

$$\begin{aligned}\mathbb{P}\Big(\sup_{1 \leq j \leq n} |S_j| \geq x \Big) \quad &\leq \quad c_p \frac{n}{x^p}\Big(v^{p+1} + \|X_1\|_p^p\Big) + 4 \sum_{j=1}^{\infty} \exp\Big\{ - \frac{c_p \mu_j^2 x^2}{n v^2 \theta_2^2(j)} \Big\} \\ &\quad + 2 \exp\Big\{ - \frac{c_p x^2}{n \|X_1\|_2^2} \Big\}.\end{aligned}$$

(ii) *Assume that $\Theta_{m,p} = O(m^{-\alpha}), \alpha > 1/2 - 1/p$. Then there exist positive constants C_1, C_2 such that for any $x > 0$,*

$$\mathbb{P}\Big(\sup_{1 \leq j \leq n} |S_j| \geq x \Big) \quad \leq \quad \frac{C_1 \Theta_{0,p}^p n}{x^p} + +4G_{1-2/p}\Big(\frac{C_2 x}{\sqrt{n}\Theta_{0,p}} \Big).$$

(iii) *If $\Theta_{m,p} = O(m^{-\alpha}), \alpha < 1/2 - 1/p$, then for any $x > 0$,*

$$\mathbb{P}\Big(\sup_{1 \leq j \leq n} |S_j| \geq x \Big) \quad \leq \quad \frac{C_1 \Theta_{0,p}^p n^{p(1/2-\alpha)}}{x^p} + 4G_{(p-2)/(p+1)}\Big(\frac{C_2 x}{n^{(2p-1-2\alpha p)/(2+2p)}\Theta_{0,p}} \Big).$$

Let $p \in [1, 2]$. For any $x > 1$, let $r_x > 0$ be the solution to the equation

$$x = (1 + r_x)^{\nu(p)} \exp\Big\{ \frac{r_x^2}{2} \Big\}, \qquad \text{where } \nu(p) = \begin{cases} 2p+1 & \text{if } p \in (1, \frac{3}{2}]; \\ 6p-3 & \text{if } p \in (\frac{3}{2}, 2]. \end{cases}$$

One says that U_n satisfies Cramér moderate deviations (CMD) with rate t_n and exponent $p > 0$ if, for every $a > 0$, there exists a constant $C = C_{a,p}$, does not depend on x or n, such that

$$\Big| \frac{\mathbb{P}(U_n \geq r_x)}{1 - \Phi(r_x)} - 1 \Big| \leq C(\frac{x}{t_n})^{\frac{1}{1+2p}} \quad \text{and} \quad \Big| \frac{\mathbb{P}(U_n \leq -r_x)}{\Phi(-r_x)} - 1 \Big| \leq C(\frac{x}{t_n})^{\frac{1}{1+2p}}$$

hold uniformly in $x \in [1, at_n]$, where $\Phi(x)$ is the standard normal distribution function. Wu and Zhao [42] showed that physical dependence sequences satisfy CMD.

Theorem 3.4 ([42]). *Let X_i be defined by (3). Assume $X_0 \in L^{2p}, p \in (1, 2]$ and $\Theta_{0,2p} < \infty$. Then the limit $\sigma = \lim_{n\to\infty} \|S_n\|_2/\sqrt{n}$ exists and is finite. Assume that $\sigma > 0$, and that there exist $0 < \alpha \leq \beta \leq \alpha + \frac{1}{2}$ such that the following conditions hold:*

$$\Theta_{m,2p} = O(m^{-\alpha}) \qquad \textit{and} \qquad \sum_{i=m}^{\infty} \theta_{2p}^2(i) = O(m^{-2\beta}).$$

Let $\eta = \alpha\beta/(1+\alpha)$. Then $S_n/(\sigma\sqrt{n})$ satisfies CMD with rate $t_n = n^{p-1}$, or $t_n = n^{p-1}/\log^p n$, or $t_n = n^{p\eta}$, under $\eta > 1 - 1/p$, or $\eta = 1 - 1/p$, or $\eta < 1 - 1/p$, respectively, and exponent p.

Set $\alpha \in (0, 1)$. Let $m = \lfloor n^\alpha \rfloor$ and $k = \lfloor n/(2m) \rfloor$. Denote

$$S_{l,s} = \sum_{i=l+1}^{l+s} X_i$$

the block sums of X_i for $l+1 \le i \le l+s$, and $Y_j = S_{2m(j-1),\,m}$. Set

$$S_k^o = \sum_{j=1}^k Y_j \qquad \text{and} \qquad [S^o]_k = \sum_{j=1}^k (Y_j)^2.$$

Denote the *interlacing self-normalized sums* as follows

$$W_n^o = S_k^o / \sqrt{[S^o]_k}. \tag{7}$$

The following self-normalized Cramér moderate deviation result holds under a geometric moment contraction condition.

Theorem 3.5 ([7]). *Let X_i be defined by (3). Assume that $\mathbb{E}|X_i|^q \le c_1^q$ for all i and a constant $q \in (2,3]$ and that*

$$\mathbb{E}S_{l,s}^2 \ge c_2^2 s \quad \text{for all } l \ge 0 \text{ and } s \ge 1.$$

Assume also that there exist three positive constants a_1, a_2 and $\tau \in (0,1]$ such that

$$\Delta_q(n) \le a_1 e^{-a_2 n^\tau}.$$

For any $0 < \alpha < 1$, then there exists a positive constant C, depending only on $c_1/c_2, a_1, a_2, \alpha, q$ and τ such that

$$\left| \log \frac{\mathbb{P}(W_n^o \ge x)}{1 - \Phi(x)} \right| \le C \left(\frac{(1+x)^q}{n^{(1-\alpha)(q/2-1)}} \right) \tag{8}$$

for all $0 \le x \le c_0 \min(n^{(1-\alpha)/2}, n^{\alpha\tau/2})$. In particular, it implies that

$$\frac{\mathbb{P}(W_n^o \ge x)}{1 - \Phi(x)} = 1 + o(1) \tag{9}$$

for all $0 \le x = o\big(\min(n^{(1-\alpha)(q-2)/2q}, n^{\alpha\tau/2})\big)$.

We also have the following self-normalized Cramér moderate deviations for β-mixing sequences.

Theorem 3.6 ([7]). *Assume that $\mathbb{E}|X_i|^{2+\nu} \le c_0^{2+\nu}$ for all i and a constant $\nu \in (0,1]$, and that*

$$\mathbb{E}S_{l,s}^2 \ge c_1^2 s \quad \text{for all } l \ge 0 \text{ and } s \ge 1.$$

Assume also that there exist three positive constants a_1, a_2 and $\tau \in (0,1]$ such that

$$\beta(n) \le a_1 e^{-a_2 n^\tau}.$$

Then for any positive constant $\rho < \nu$,

$$\left| \log \frac{\mathbb{P}(W_n^o \ge x)}{1 - \Phi(x)} \right| \le c_\rho \left(\frac{(1+x)^{2+\rho}}{n^{(1-\alpha)\rho/2}} \right) \tag{10}$$

uniform for $0 \le x = o(\min\{n^{(1-\alpha)/2}, n^{\alpha\tau/2}\})$, where c_ρ depends only on c_0, c_1, ρ, a_1, a_2 and τ. In particular, it implies that

$$\frac{\mathbb{P}(W_n^o \ge x)}{1 - \Phi(x)} = 1 + o(1) \tag{11}$$

uniformly for $0 \le x = o(\min\{n^{(1-\alpha)\rho/(4+2\rho)}, n^{\alpha\tau/2}\})$.

For ψ-mixing sequences, we have the following self-normalized Cramér moderate deviations.

Theorem 3.7 ([21]). *Assume that there exists a constant $\rho \in (0,1]$ such that*

$$\mathbb{E}|S_{l,s}|^{2+\rho} \le s^{1+\rho/2} c_2^{2+\rho} \tag{12}$$

and that

$$\mathbb{E}S_{l,s}^2 \ge c_1^2 s \quad \text{for all } l \ge 0 \text{ and } s \ge 1. \tag{13}$$

Assume also that for some $\alpha \in (0,1)$, *it holds*

$$\psi(n) = O\big(n^{-(1+\rho)/\alpha}\big), \quad n \to \infty.$$

Then for any $\rho \in (0,1]$, *the following equality holds*

$$\frac{\mathbb{P}(W_n^o > x)}{1-\Phi(x)} = 1 + o(1) \tag{14}$$

uniformly for $0 \leq x = o(n^{(1-\alpha)\rho/(4+2\rho)})$.

Similar results for stationary or ϕ-mixing sequences, we refer to Fan [20].

4. Central limit theorems

In this section, we consider a random process $(X_n)_{n\in\mathbb{Z}}$ with finite expectation. We set, for all $n \in \mathbb{N}$,

$$S_n = \sum_{i=1}^{n}(X_i - \mathbb{E}X_i),$$

and, for all $t \in [0,1]$,

$$W_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor}(X_i - \mathbb{E}X_i),$$

where $\lfloor t \rfloor$ denote the greatest integer lower than t. The aim of this section is to summarize some conditions under which the asymptotic behaviors of $n^{-1/2}S_n$ and $(W_n(t))_{t\in[0,1]}$ are known.

4.1. Central limit theorems for sequences

This section is devoted to results about asymptotic normality of $n^{-1/2}S_n$. Since we have introduce two type of dependence measures, mixing and weak dependence, the additional assumptions differ depending on which kind of dependence structure is fulfilled. Nevertheless, the general idea is, in each case the series of corresponding coefficients should converge. The first result of asymptotic normality relies on a hypothesis about the sum of covariances.

Theorem 4.1 ([33])**.** *Suppose* $(X_n)_{n\in\mathbb{N}}$ *be stationary and ergodic random variables with finite second order moment. If for* $\mathcal{F}_0 = \sigma(X_i,\ i \leq 0)$ *and* $n \in \mathbb{N}$,

$$\sum_{k>0} cov\left(\mathbb{E}(X_n|\mathcal{F}_0), X_k\right) < \infty$$

and

$$\lim_{n\to+\infty}\sup_{K>0}\Big|\sum_{k>K} cov\left(\mathbb{E}(X_n|\mathcal{F}_0), X_k\right)\Big| = 0,$$

then $n^{-1}var(S_n)$ *converges to a finite* $\sigma^2 > 0$ *and* $n^{-1/2}S_n$ *converges in distribution to a normal distribution* $\mathcal{N}(0,\sigma^2)$.

The extra assumption due to dependence in this theorem may be interpreted as a condition of summability about coefficients ϵ_r in 2.2. In the case of mixing processes we can weaken this condition.

Theorem 4.2 ([33, 16])**.** *Suppose* $(X_n)_{n\in\mathbb{N}}$ *be stationary and ergodic random variables with finite second order moment. If the process* $(X_n)_{n\in\mathbb{N}}$ *fulfills one of the following two conditions, then* $n^{-1}var(S_n)$ *converges to a finite* $\sigma^2 > 0$ *and* $n^{-1/2}S_n$ *converges in distribution to a normal distribution* $\mathcal{N}(0,\sigma^2)$.

1. *The process* $(X_n)_{n\in\mathbb{N}}$ *is* α*-mixing and satisfies*

$$\int_0^1 \alpha^{-1}(u)[Q_0(u)]^2 \mathrm{d}u < \infty,$$

where $Q_0(u) = \inf\{\, t\ :\ \mathbb{P}(|X_0| > t) \leq u\}$ *and*

$$\alpha^{-1}(u) = \inf\{n \in \mathbb{N}\ :\ \sup_{k\in\mathbb{Z}} \alpha(\mathcal{F}_k, X_{k+n}) \leq u\}.$$

2. *The sequence* $(X_n)_{n\in\mathbb{N}}$ *has a common distribution and is* β*-mixing (resp* α*-mixing) with mixing coefficients satisfying*

$$\sum_{n\geq 1}\beta(n)<\infty\left(\textit{resp}\sum_{n\geq 1}\alpha(n)<\infty\right).$$

There exist an extension of this theorem, when it is stated in terms of condition 2, to the case of transformation of the sequence $(X_n)_{n\in\mathbb{N}}$, we refer to [33] for more details.

4.2. Donsker-type theorems

Donsker's theorem always involve Wiener's measure on $[0,1]$ denoted by $\mathbb{W}$ and the Skorohod space $D([0,1])$ which is roughly the space of *càdlàg* functions. We refer to the monography of Billingsley [5] for those classical definitions.

Theorem 4.3 ([33]). *If the real valued strictly stationary process* $(X_n)_{n\in\mathbb{N}}$ *satisfies the assumption 1 of Theorem 4.2, then* $n^{-1}var(S_n)\to\sigma^2>0$ *and*

$$W_n \underset{n\to+\infty}{\longrightarrow} \sigma\mathbb{W} \quad \textit{in} \quad D([0,1]).$$

Apart of mixing conditions, in several different types of weak convergence it is possible to describe explicitly the rate of convergence required for $(\epsilon_r)_{r\in\mathbb{N}}$ in order to have Donsker's theorems.

Theorem 4.4 ([15]). *Consider the real valued stationary process* $(X_n)_{n\in\mathbb{N}}$ *with zero mean such that*

$$\mathbb{E}|X_0|^m<+\infty \quad \textit{for a real number } m>2.$$

Then

$$\sigma^2:=\sum_{t\in\mathbb{Z}}cov(X_0,X_t)<\infty$$

and it holds

$$W_n \underset{n\to+\infty}{\longrightarrow} \sigma\mathbb{W} \quad \textit{in} \quad D([0,1]),$$

if one of the following assumptions is fulfilled:

- κ***-dependence :*** *The process is* κ*-weakly dependent and satisfies* $\kappa_r=O(r^{-\kappa}), r\to\infty$*, for some* $\kappa>2+1/(m-2)$.
- λ***-dependence :*** *The process is* λ*-weakly dependent and satisfies* $\lambda_r=O(r^{-\lambda}), r\to\infty$*, for some* $\lambda>4+2/(m-2)$.
- θ***-dependence :*** *The process is* θ*-weakly dependent and satisfies* $\theta_r=O(r^{-\theta}), r\to\infty$*, for some* $\theta>1+1/(m-2)$.

This theorem is a concatenation of the results of Dedecker and Doukhan [8] (for θ-dependence) and Doukhan and Wintenberger [19] (for κ and λ-dependence).

4.3. Triangular arrays

In the context or triangular schemes conditions for asymptotic normality are stated in terms close to the notion of θ-weak dependence and rely on Lindeberg's method. Triangular arrays appear as natural issues in functional estimation which require windowing or thresholding as mentioned in [35].

Theorem 4.5 ([29]). *Suppose that* $(X_{n,k})_{1\leq k\leq n}$*,* $n\in\mathbb{N}$*, is a triangular scheme of stationary random variables with* $\mathbb{E}X_{n,k}=0$ *and* $\sum_{k=1}^n\mathbb{E}X_{n,k}^2\leq C$ *for all* n,k *and some* $C<\infty$*. Furthermore, we assume that*

$$\sum_{k=1}^n\mathbb{E}\left(X_{n,k}^2\mathbb{1}_{\{|X_{n,k}|>\varepsilon\}}\right)\underset{n\to\infty}{\longrightarrow}0$$

holds for all $\varepsilon>0$ *and that*

$$var(X_{n,1}+\cdots+X_{n,n})\underset{n\to\infty}{\longrightarrow}\sigma^2\in[0,\infty).$$

For all n large enough, there exists a monotonously nonincreasing and summable sequence $(\gamma(r))_{r\in\mathbb{N}}$ *such that, for all indices* $s_1<\cdots<s_u<s_u+r=t_1\leq t_2$*, the following upper bounds for covariances hold true:*

- *For all measurable and square integrable function $g : \mathbb{R}^u \to \mathbb{R}$,*

$$|cov\,(g(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1}))\,| \le \frac{1}{\sqrt{n}}\sqrt{\mathbb{E}(g^2(X_{n,s_1}, \dots, X_{n,s_u}))}\gamma(r).$$

- *For all measurable and bounded functions $g : \mathbb{R}^u \to \mathbb{R}$,*

$$|cov\,(g(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1}X_{n,t_2}))\,| \le \frac{1}{n}\|g\|_\infty\gamma(r).$$

Then

$$\sum_{k=1}^n X_{n,k} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2),$$

where $\xrightarrow{d}$ stands for convergence in distribution.

Theorem 4.5 can be extended to the case where $X_{n,k} \in \mathbb{R}^d$ with $d > 1$. Such result can be found in [4]. For triangular arrays, we also have the following result similar to Theorem 4.5.

Theorem 4.6 ([28])**.** *Suppose that $(X_{n,k})_{1\le k\le n}$, $n \in \mathbb{N}$, is a triangular scheme of stationary random variables with $\mathbb{E}X_{n,k} = 0$ and $\sum_{k=1}^n \mathbb{E}X_{n,k}^2 \le C$ for all n, k and some $C < \infty$. Furthermore, assume that*

$$\sum_{k=1}^n \mathbb{E}\left(X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}|>\varepsilon\}}\right) \underset{n\to\infty}{\longrightarrow} 0$$

holds for all $\varepsilon > 0$ and that

$$var(X_{n,1} + \cdots + X_{n,n}) \underset{n\to\infty}{\longrightarrow} \sigma^2 \in [0, \infty).$$

For all n large enough, there exists a summable sequence $(\gamma(r))_{r\in\mathbb{N}}$ such that, for all indices $s_1 < \cdots < s_u < s_u + r = t_1 \le t_2$, the following upper bounds for covariances hold true: for all measurable functions $g : \mathbb{R}^u \to \mathbb{R}$,

$$|cov\,(g(X_{n,s_1}, \dots, X_{n,s_u})X_{n,s_u}, X_{n,t_1}))\,| \le \left(\mathbb{E}X_{n,s_u}^2 + \mathbb{E}X_{n,t_1}^2 + n^{-1}\right)\gamma(r)$$

and

$$|cov\,(g(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1}X_{n,t_2}))\,| \le \left(\mathbb{E}X_{n,t_1}^2 + \mathbb{E}X_{n,t_2}^2 + n^{-1}\right)\gamma(r).$$

Then

$$\sum_{k=1}^n X_{n,k} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2).$$

5. Functional central limit theorems

This section is concerned with a weak convergence of empirical measures. The empirical measure of a sample $X_1, \dots, X_n$ of random variables, taken their values in a space $\mathcal{X}$, is defined as

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}.$$

Considering a class of functions $\mathcal{F}$ we defined the $\mathcal{F}$-indexed empirical process $\mathbb{G}_n$ by

$$f \in \mathcal{F} \mapsto \mathbb{G}_n f = \sqrt{n}\left(\mathbb{P}_n f - \mathbb{P}f\right),$$

where $\mathbb{P}$ is the common distribution of the $(X_i)_{i=1}^n$ and $\mathbb{Q}f$ denote, for any measure $\mathbb{Q}$, $\int_\mathcal{X} f\mathrm{d}\mathbb{Q}$. A functional central limit theorem can be seen as a uniform version of the central limit theorem $\mathbb{G}_n f \xrightarrow{d} \mathcal{N}(0, \mathbb{P}(f - \mathbb{P}f)^2)$. In that sense we investigated in this section, under which conditions there exists a Gaussian process $\mathbb{G}$ such that

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G}, \quad l^\infty(\mathcal{F}).$$

Here, $l^\infty(\mathcal{F})$ denotes the space of $\mathcal{F}$-indexed process endowed with uniform metric. Considering a stationary sequence $(X_t)_{t\in\mathbb{Z}}$, the candidate of the limit Gaussian process is entirely determined by finite dimensional convergences, also known as fi-di convergences. Indeed, for any $p \in \mathbb{N}^*$ and any vector $(f_1, \dots, f_p) \subset \mathcal{F}$ we have

$$(\mathbb{G}_n f_1, \dots, \mathbb{G}_n f_p) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right), \tag{15}$$

where $\Sigma_{i,j} := \mathbb{P}\big((f_i - \mathbb{P}f_i)(f_j - \mathbb{P}f_j)\big)$. Consequently, under weak dependence assumption for which previous convergences holds, only the question of the existence of a such Gaussian limit remains. The following theorem gives a baseline of a general strategy to solve this problem. In the two following subsections, we present results adapted for particular type of classes $\mathcal{F}$.

Lemma 5.1. *Let ρ be a metric on $\mathcal{F}$. Assume the following two conditions:*

1. *The convergence* (15) *hold for any $p \in \mathbb{N}^*$.*
2. *For every $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} \mathbb{P}^* \left(\sup_{\substack{(f,g)\in\mathcal{F}^2 \\ \rho(f,g)<\delta}} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| > \eta \right) < \varepsilon.$$

Then there exists a tight Gaussian process $\mathbb{G}$ with margins given by (15) *such that $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$. Conversely, if $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$ where $\mathbb{G}$ is a tight Gaussian process then conditions* 1 *and* 2 *are fulfilled.*

In Lemma 5.1, the notation $\mathbb{P}^*$ denotes the outer probability, which is the probability of the lowest measurable set containing our set of interest. This precaution is necessary since, in the empirical processes theory, non measurable map might easily appears (see e.g. Example 3 p. 48 in [37]). Note that, even if those details are hidden here, the weak convergence is defined in terms of outer expectation to avoid problems of measurability [23]. The condition 2 is commonly named asymptotic ρ-equicontinuity and is closely related with the asymptotic tightness of $\mathbb{G}$. For the i.i.d. case, we refer to the monographs of Van der Vaart and Wellner [37] or Pollard [31].

5.1. Empirical cumulative distribution functions

We first consider the case, may be the more natural, of empirical cumulative distribution function (cdf) of real valued random variable. This correspond to the case where the class $\mathcal{F}$ is the class of indicators of the half real line, that is

$$\mathcal{F} = \{\mathbb{1}_{\{.\le x\}},\ x \in \mathbb{R}\}.$$

This definition can be easily extended to the case of random variable with value in $\mathbb{R}^d$ considering the class of indicators of quadrant understanding "$\le$" components wise.

For the class of indicators, the condition 2 of Lemma 5.1 is easy to check, this leads functional CLT which can be clearly stated.

The next theorem presents some functional CLTs for univariate cdf.

Theorem 5.2 ([33, 36, 18])**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary sequence of real random variables with common continuous cumulative distribution function F. There exists a tight Gaussian process $\mathbb{G}$ such that*

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G}, \quad \textit{in } l^\infty(\mathbb{R}),$$

if one of the following assertion are fulfilled:

1. *The sequence $(\alpha(r))_{r\ge 0}$ of strong mixing coefficients satisfies*

$$\alpha(r) \le c\, r^{-\alpha}, \quad \textit{with } \alpha > 1 \textit{ and } c \ge 1.$$

2. *The sequence $(\rho(r))_{r\ge 0}$ of maximal correlation coefficients satisfies*

$$\sum_{n=1}^{\infty} \rho(2^n) < \infty.$$

3. *The sequence $(X_t)_{t\in\mathbb{Z}}$ is η-weakly dependent with dependence coefficients satisfying*

$$\eta_r = O\big(r^{-15/2-\nu}\big), \quad \textit{with } \nu > 0.$$

4. *The sequence $(X_t)_{t\in\mathbb{Z}}$ is κ-weakly with dependence coefficients satisfying*

$$\kappa_r = O\left(r^{-5-\nu}\right), \quad \textit{with } \nu > 0.$$

Under α and β-mixing conditions, we have the following functional CLT for multivariate cdf.

Theorem 5.3 ([33, 3])**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary sequence of random variables with values in $\mathbb{R}^d$. We assume that the univariate cdf of the margins are continuous. There exists a tight Gaussian process $\mathbb{G}$ such that*

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G}, \quad \textit{in } l^\infty(\mathbb{R}^d),$$

if one of the following assertion is fulfilled:

1. *The sequence $(\alpha(r))_{r\geq 0}$ of strong mixing coefficients satisfies*
$$\alpha(r) \leq c\, r^{-\alpha}, \quad \textit{with } \alpha > 1 \textit{ and } c \geq 1.$$
2. *The sequence $(\beta(r))_{r\geq 0}$ of absolute regularity coefficients satisfies*
$$\sum_{r=1}^{\infty} \beta(r) < \infty.$$

In order to derive a multivariate central limit theorem, it is convenient to introduce a particular case of the coefficient τ_p, which is more adapted to deal with empirical processes. Consider, for $p \in [1, \infty]$, the coefficient

$$\tilde{\beta}_p(\mathcal{M}, X) = \left\| \sup_{(t_i)_{i=1}^n \in (\mathbb{R}^d)^n} \left(\int \prod_{i=1}^n g_{t_i,i}(x_i) \mathbb{P}_{X|\mathcal{M}}(\mathrm{d}x) - \int \prod_{i=1}^n g_{t_i,i}(x_i) \mathbb{P}_X(\mathrm{d}x) \right) \right\|_p,$$

where $g_{t_i,i} = \mathbb{1}_{\{x \leq t_i\}} - \mathbb{P}(X_i \leq t_i)$. We define quantities $\tilde{\beta}_{p,k}(r)$ in the same way as Definition 2.7.

The following theorem gives a functional CLT for multivariate cdf under weak dependence.

Theorem 5.4 ([12])**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary sequence of random variables taking their values on $\mathbb{R}^d$. There exists a tight Gaussian process $\mathbb{G}$ such that*

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G}, \quad \textit{in } l^\infty(\mathbb{R}^d),$$

if one of the following assertion are fulfilled:

1. *There exists $\varepsilon \in (0, 1]$ and $p' > d(2 + \varepsilon)/(2\varepsilon)$ such that $\tilde{\beta}_{2,p'}(r) = O(r^{-1-\varepsilon})$.*
2. *There exists $\varepsilon > 0$ such that*
$$\sum_{r=1}^{\infty} r \tilde{\beta}_{2,d+\varepsilon}(r) < \infty.$$
3. *There exists $\varepsilon > 0$ such that $\tilde{\beta}_{2,\infty}(r) = O(r^{-1-\varepsilon})$.*
4. *There exists $\varepsilon > 0$ such that $\tilde{\beta}_{2,1}(r) = O(r^{-2d-\varepsilon})$.*
5. *Each component of X_1 has a bounded density and there exists $\varepsilon > 0$ such that $\tau_{2,\infty}(r) = O(r^{-2-\varepsilon})$.*
6. *Each component of X_1 has a bounded density and there exists $\varepsilon > 0$ such that $\tau_{2,1}(r) = O(r^{-4d-\varepsilon})$.*

Note that the conditions in this theorem are linked following the diagrams $5 \Rightarrow 3 \Rightarrow 1$ and $6 \Rightarrow 4 \Rightarrow 2$.

5.2. *Classes of functions with finite entropy*

The aim of this section is to pick out conditions on the class $\mathcal{F}$ under which the point 2) of Lemma 5.1 is fulfilled. Since this condition is concerned in equicontinuity, we can investigate from the side of class of regular functions. Following this way, if we take $\mathcal{F}$ as a ball in the space of Lipschitz square integrable functions endowed with a suitable metric, then Lemma 5.1 holds under assumption of sumability of strong mixing coefficients. We refer to section 8.2 in [33] to the construction of such balls and more specially to Theorem 8.1 to the result.

However, the fact that a class $\mathcal{F}$, fulfilling the equicontinuity conditions, is related to its "size". The size is defined in term of entropy with brackets (also call bracket entropy or bracketing number) or metric entropy number.

5.2.1. Bracketting entropy numbers

Definition 5.1. Consider $\mathcal{H}$ a vector space of functions and $\mathcal{F} \subset \mathcal{H}$.

1. Let $f,\ g \in \mathcal{F}$, such that $f \leq g$ (pointwise). We define the interval of functions or "brackets" between f and g, denoted by $[f, g]$, the set
$$\{h \in \mathcal{H} \,:\, f \leq h \leq g\}.$$
When $\mathcal{F}$ is endowed by a distance d, $d(f, g)$ is the diameter of $[f, g]$.
2. The class $\mathcal{F}$ is said totally bounded with brackets if for every $\delta > 0$, there exists a finite set $S(\delta)$ of brackets with diameter at most δ such that for all $f \in \mathcal{F}$, there exists $[h, g] \in S(\delta)$ such that $f \in [h, g]$.

3. We define the bracketing number of $\mathcal{F}$ as the lowest cardinal of sets $S(\delta)$. This bracketing number is denoted $\mathcal{N}_{[\,]}(\delta, \mathcal{F})$.
4. The entropy (with brackets) number of $\mathcal{F}$ is defined by

$$H_{[\,]}(\delta, \mathcal{F}, d) = \log\left(\max(\mathcal{N}_{[\,]}(\delta, \mathcal{F}), 2)\right).$$

Note that bracketing number and entropy with brackets both depend on the diameter δ and on the metric on $\mathcal{F}$.

For i.i.d. sequence with common distribution $\mathbb{P}$, when $\mathcal{F} \subset L^2(\mathbb{P})$, the condition

$$\int_0^1 \sqrt{H_{[\,]}(x, \mathcal{F}, \|.\|_2)}\mathrm{d}x < \infty$$

is sufficient so that the $\mathcal{F}-$indexed empirical process satisfies a functional central limit theorem. See [30]. However this condition can't be used as well in a dependent setting. For $(X_k)_{k\in\mathbb{Z}}$ stationary β-mixing sequence we introduce a weighted version of the Euclidean norm:

$$\|f\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u) U_f^2(u)\mathrm{d}(u)},$$

where β^{-1} is the inverse function of $r \in \mathbb{N} \mapsto \beta(r)$ for $(\beta(r))_{r\in\mathbb{N}}$ the sequence of β-mixing coefficients and U_f is the inverse function of $t \mapsto \mathbb{P}(|f(X_0)| > t)$. We denote by $L^{2,\beta}(\mathbb{P})$ the space of functions for which the norm $\|.\|_{2,\beta}$ is finite.

Theorem 5.5 ([17])**.** *Let $(X_k)_{k\in\mathbb{Z}}$ be stationary β-mixing sequence with common distribution $\mathbb{P}$ such that*

$$\sum_{r>0} \beta(r) < \infty.$$

If $\mathcal{F} \subset L^{2,\beta}(\mathbb{P})$ fulfills

$$\int_0^1 \sqrt{H_{[\,]}(x, \mathcal{F}, \|.\|_{2,\beta})}\mathrm{d}x < \infty, \tag{16}$$

then $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$.

The condition (16) is not that easy to be checked. Even the class of quadrant do not fulfill this condition. To avoid this issue, the idea is to use the maximal coupling of Goldstein [22] to built a suitable measure for which $\mathcal{F}$ fulfills the equicontinuity condition with the unweighted norms of L^2 or L^1.

Theorem 5.6. *Let $(X_k)_{k\in\mathbb{Z}}$ be a stationary β-mixing sequence with common distribution $\mathbb{P}$ such that*

$$\sum_{r>0} \beta(r) < \infty.$$

Let $\mathbb{Q}$ be a positive measure built by maximal coupling and $\mathcal{F} \subset L^1(\mathbb{Q})$ a class of functions taking their values in $[-1, 1]$. If $\mathcal{F}$ is totally bounded in $L^1(\mathbb{Q})$ and

$$\int_0^1 \sqrt{H_{[\,]}(x, \mathcal{F}, \|.\|_1)/x}\; \mathrm{d}x < \infty,$$

then $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$.

The measure $\mathbb{Q}$ exists and is positive when the coefficients $\beta(r)$ are summable and its construction is explicit. One can find this construction in the monography of Rio [33] p. 112.

Dedecker and Louichi [15] established the counterpart of Theorem 5.6 in terms of ϕ-mixing. We recall that those mixing coefficients are defined for $r \in \mathbb{N}^*$ as

$$\phi_k(r) = \sup_{r<i_1\le\cdots\le i_k} \phi(\sigma(X_i,\ i \le 0), \sigma(X_{i_1}, \ldots, X_{i_k})).$$

Theorem 5.7 ([15])**.** *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence with common distribution $\mathbb{P}$ and $\mathcal{F} \subset L^2(\mathbb{P})$. Then $\mathbb{G}_n \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$ if one of the following assumptions is fulfilled.*

- *The sequence $(X_i)_{i\in\mathbb{Z}}$ is ϕ-mixing such that*

$$\sum_{k>0} k\phi_2(k) < \infty \quad \text{and} \quad \int_0^1 \sqrt{H_{[\,]}(x,\mathcal{F},\|.\|_4)}\mathrm{d}x < \infty.$$

- *The sequence $(X_i)_{i\in\mathbb{Z}}$ is ϕ-mixing such that $\phi_2(k) = O(k^{-b})$ for some $b \in (1,2)$ and*

$$\int_0^1 \sqrt{H_{[\,]}(x,\mathcal{F},\|.\|_{2b/(b-1)})}\mathrm{d}x < \infty.$$

5.2.2. Metric entropy numbers

Definition 5.2. Let $\mathcal{F}$ be a class of functions endowed by a seminorm ρ and a real $\varepsilon > 0$. The covering number $\mathcal{N}(\varepsilon,\mathcal{F},\rho)$ is a minimal number of balls of radius ε needed to cover $\mathcal{F}$. We so define the metric entropy number as

$$H(\varepsilon,\mathcal{F},\rho) = \log(\mathcal{N}(\varepsilon,\mathcal{F},\rho)).$$

In the i.i.d. case this approach is often preferred since the metric entropy number with respect to the uniform metric can be easily controlled for the classes of Vapnik-Chervonenkis [38]. But the consideration of those classes is not relevant in the present dependent case and many references restrict to classes of BV-functions.

A real function h is said to be a BV-function if there exists a finite signed measure dh such that, $h(x) = h(0) + dh([0,x))$ if $x > 0$ and $h(x) = h(0) - dh([x,0))$ if $x < 0$. Following the Hahn-Jordan decomposition, there exist a unique couple (dh_+, dh_-) of (positive) measures such that $dh = dh_+ - dh_-$. We so define the norm $\|dh\| = dh_+(\mathbb{R}) + dh_-(\mathbb{R})$. Moreover we said that h is BV_1 if h is BV and $\|dh\| \leq 1$. Note that the map $h \mapsto |h|_v = \|dh\|$ defines a seminorm over any class of BV functions.

In order to derive central limit theorem it is convenient to set ϕ-mixing coefficient in terms of BV_1 functions. For X a real valued random variable of distribution $\mathbb{P}_X$ and $\mathcal{M}$ a σ-algebra, Dedecker and Prieur [12] showed that

$$\phi(\mathcal{M},X) = \left\| \sup_{f\in BV_1} \left| \int f(x)\mathbb{P}_{X|\mathcal{M}}(\mathrm{d}x) - \int f(x)\mathbb{P}_X(\mathrm{d}x) \right| \right\|_1 .$$

Others mixing coefficients can be expressed in the same way, we refer to [12] for proofs and comparison relations between those coefficients. We further define, for a positive integer k, the coefficients

$$\phi(k) = \sup_{i\geq 0} \phi(\sigma(X_j, j \leq i), X_{k+i}).$$

Theorem 5.8. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence of real-valued random variables and $\mathcal{F}$ a class of BV functions. If we have*

$$\sum_{k=1}^{\infty} \phi(k) < \infty \text{ and } \int_0^1 \sqrt{H(\varepsilon,\mathcal{F},|.|_v)}\mathrm{d}\varepsilon < \infty,$$

then

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G} \quad \text{in } l^\infty(\mathcal{F}).$$

Note that a class of convex Lipschitz functions always satisfies the metric entropy condition.

5.2.3. The Sobolev balls

In this example, we consider a probability space $(\Omega,\mathcal{A},\mathbb{P})$ and $T : \Omega \to \Omega$ bimeasurable which preserve $\mathbb{P}$. Consider the sequence $(X_k)_{k\geq 0}$ defined as $X_k = X_0 \circ T^k$ and the filtration $\mathcal{M}_k = T^k(\mathcal{M}_0)$ with $\mathcal{M}_0$ a sub-σ-algebra of $\mathcal{A}$. Moreover, we denote by $\mathcal{I}$ the σ-algebra of all T-invariant sets. In this framework, Merlevède and Dedecker [10] considered the special case of the class of functions indexed by the general Sobolev balls. They see $F_n - F$, where F_n is the empirical cumulative distribution function of the $(X_i)_{1\leq i\leq n}$ with common distribution F, as a random element of $L^p(\mu)$ for a certain probability measure μ and $1 \leq p < \infty$. In this framework, the considered class of functions is

$$W_{p',1}(\mu) = \left\{ f : f(t) = f(0) + \int_{[0,t)} f'(x)\mu(dx)\mathbb{1}_{\{t>0\}} - \int_{(t,0]} \mu(dx)\mathbb{1}_{\{t\leq 0\}}, \|f'\|_{q,\mu} \right\}$$

with p' the conjugate exponent of p namely $1/p + 1/p' = 1$. With this class one can express the τ-mixing coefficient as

$$\tau_{\mu,p,q}(\mathcal{M}, X) = \Big\| \sup_{f \in W_{q,1}(\mu)} | \int f dF_{X|\mathcal{M}} - \int f dF_X | \Big\|_q .$$

We naturally define for an integer k, $\tau_{\mu,p,q}(k)$ as $\tau_{\mu,p,q}(\mathcal{M}_0, X_k)$. Since central limit theorems involve weak convergence in l^∞(class of functions), we need an isometry between $L^p(\mu)$ and $l^\infty(W_p, 1(\mu))$. The natural candidate for a such map is the application $h : L^p(\mu) \to l^\infty(W_p, 1(\mu))$ defined as $h(g) = \{\mu(f'g), f \in W_{p',1}(\mu)\}$. This need come from the fact that the present setup involves an L^p space instead of l^∞

Theorem 5.9 ([10]). *Define the function F_μ for $x \in \mathbb{R}$ by $F_\mu(x) = \mu([0, x))$ if $x \geq 0$ and $F_\mu(x) = -\mu([x, 0))$ if $x \leq 0$. Assume that $\| \, |F_\mu(X_0)|^{1/p} \|_2$ is finite. The empirical process $\{\mathbb{G}_n(f), f \in W_{1,p'}\}$ weakly converges in $h(W_{p',1}(\mu))$ to a tight process which is Gaussian centered conditionally to $\mathcal{I}$ if one of the following conditions is fulfilled:*

1. *$p \in [2, \infty)$ and $\sum_{k>0} \tau_{\mu,p,2}(k) < \infty$.*
2. *$p = 2$, $\mu(\mathbb{R}) < \infty$ and $\sum_{k>0} \tau_{\mu,2,1}(k) < \infty$.*
3. *$p = 2$, $F_{X_0|\mathcal{M}_{-\infty}} = F$ and $\sum_{k>0} \| \| \|F_{X_k|\mathcal{M}_0} - F_{X_k|\mathcal{M}_{-1}} \|_{2,\mu} \|_2 < \infty$.*

The knowledge of the result for τ-mixing sequence is enough to deal with a large amount of mixing sequences since many of this dependence coefficient can control the coefficient τ.

Proposition 5.10. *Let X be a real valued random variable and $\mathcal{M}$ a sub-σ-algebra of $\mathcal{A}$. Then*

1. *for any $p, q \in [1, \infty]$ and for any measure μ: $\tau_{\mu,p,q}(\mathcal{M}, X) \leq \mu(\mathbb{R})^{1/p} \phi(\mathcal{M}, X)$,*
2. *for any $p, q \in [1, \infty]$ and for any measure μ: $\tau_{\mu,p,q}(\mathcal{M}, X) \leq \mu(\mathbb{R})^{1/p} \beta(\mathcal{M}, X)^{1/q}$,*
3. *if $t \mapsto \mu((-\infty, t])$ is K-Lipschitz, then for any $p, q \in [1, \infty]$: $\tau_{\mu,p,q}(\mathcal{M}, X) \leq (K\phi(\mathcal{M}, X))^{1/p}$,*
4. *if $t \mapsto \mu((-\infty, t])$ is K-Lipschitz, then for any $p \in [1, \infty]$ and $q < p$: $\tau_{\mu,p,q}(\mathcal{M}, X) \leq (K\tau(\mathcal{M}, X))^{1/p}$.*

5.2.4. Classes of functions with bounded variations

In this section we denote by $\mathbb{Z}_n(x)$, $x \in \mathbb{R}$, the empirical process indexed by indicators of half lines in $\mathbb{R}$ (this is the case treated in section 5.1). The aim of this section is to deduce asymptotic gaussianity for $\mathbb{G}_n$ from the asymptotic gaussianity of $\mathbb{Z}_n$ together with regularity conditions. Consider first, for a function $g : \mathbb{R} \to \mathbb{R}$, the total variation norm defined by

$$\|g\|_{TV} = \sup_{\Pi} \sum_{x_i, x_{i+1} \in \Pi} |g(x_{i+1}) - g(x_i)|,$$

where Π denote the set of all countable partitions of $\mathbb{R}$. Consider the set, for $T > 0$,

$$BV_T = \{g : \mathbb{R} \to \mathbb{R} \text{ such that } \|g\|_{TV} \leq T, \|g\|_\infty \leq T\}.$$

Caution, do not confuse BV_T with the set BV introduced in section 5.2.2. Moreover we denote by BV_T' the subset of BV_T of right continuous functions.

Theorem 5.11 ([32]). *Assume that there exists a distribution function F_0 such at the process $\mathbb{Z}_n$ converges weakly (as $n \to \infty$) to a tight Gaussian process with uniformly continuous sample paths with respect to the distance $d(s, t) = |F_0(s) - F_0(t)|$. Then for any $T > 0$ and any class of functions $\mathcal{G} \subset BV_T'$, the $\mathcal{G}$-indexed empirical process converges weakly on $l^\infty(\mathcal{G})$ to a Gaussian process. Moreover, the Gaussian limit has uniformly $L_1(F_0)$-continuous sample paths.*

Note that, Theorem 5.11 remains true if we replace $\mathbb{Z}_n$ by any processes satisfying the following conditions:

- $\lim_{|t| \to \infty} \mathbb{Z}_n(t) = 0$,
- the sample paths of $\mathbb{Z}_n$ are right continuous and of bounded variations,

and if we interpret $\mathbb{G}_n$ as a process indexed by $\mathcal{G}$, namely

$$\Big\{ \int g(x) d\mathbb{Z}_n(x), g \in \mathcal{G} \Big\}.$$

Acknowledgements

The authors would like to thank the editor and an anonymous referee for their helpful comments. This work was funded by CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013, and by the Labex MME-DII (https://labex-mme-dii.u-cergy.fr/).

References

[1] Ango Nze, P., Bühlmann, P., and Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for non parametric regression. *Ann. Statist.*, 30(2):397–430.
[2] Ango-Nze, P. and Doukhan, P. (2002). Weak dependence: models and applications. *Dehling W. et al. (ed.), Empirical Processes Techniques for Dependent Data*, pages 117–137.
[3] Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.*, 7(1):47–71.
[4] Bardet, J. M., Doukhan, P., Lang, G., and Ragache, N. (2008). Dependent Lindeberg central limit theorem and some applications. *ESAIM: Probab. Statist.*, 12:154–172.
[5] Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney.
[6] Blum, J. R., Hanson, D. L., and Koopmans, L. H. (1963). On the strong law of large numbers for a class of stochastic processes. *Probab. Theory Related Fields*, 2:1–11.
[7] Chen, X., Shao, Q.-M., Wu, W. B., and Xu, L. (2016). Self-normalized Cramér-type moderate deviations under dependence. *Ann. Statist.*, 44(4):1593–1617.
[8] Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Process. Appl.*, 106(1):63–80.
[9] Dedecker, J., Doukhan, P., Lang, G., Leon, J. R., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications.* Springer Verlag.
[10] Dedecker, J. and Merlevede, F. (2007). The empirical distribution function for dependent variables: asymptotic and nonasymptotic results in L^p. *ESAIM: Probab. Statist.*, 11:102–114.
[11] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132(2):203–236.
[12] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *S-tochastic Process. Appl.*, 117(1):121–142.
[13] Donoho, D. and Johnstone, I. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3):879–921.
[14] Doukhan, P. (1994). *Mixing: Examples and Properties*, volume 85. Springer.
[15] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84(2):313–342.
[16] Doukhan, P., Massart, P., and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 30(1):63–82.
[17] Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(2):393–427.
[18] Doukhan, P. and Neumann, M. H. (2008). The notion of ψ-weak dependence and its applications to bootstrapping time series. *Probab. Surv.*, 5:146–168.
[19] Doukhan, P. and Wintenberger, O. (2007). An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist.*, 27(1):45–73.
[20] Fan, X. (2019). Cramér type moderate deviations for stationary sequences of bounded random variables. *C. R. Acad. Sci. Paris, Ser. I*, 357(5):463–477.
[21] Fan, X. (2020). Cramér type moderate deviations for self-normalized ψ-mixing sequences. *J. Math. Anal. Appl.*, 486(2):123902.
[22] Goldstein, S. (1979). Maximal coupling. *Probab. Theory Related Fields*, 46:193–204.
[23] Hoffmann-Jørgensen, J. (1991). *Stochastic processes on Polish spaces*, volume 39 of *Various Publications Series (Aarhus)*. Aarhus Universitet, Matematisk Institut, Aarhus.
[24] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.*, 7:361–392.
[25] Kolmogorov, A. N. and Rozanov, J. A. (1960). On a strong mixing condition for stationary Gaussian processes. *Theory Probab. Appl.*, 5:222–227.
[26] Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statist. Sinica*, 23:1257–1272.
[27] Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields*, 151:435–474.

[28] Neumann, M. H. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probab. Statist.*, 17:120–134.

[29] Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models : Central limit theory and bootstrap approximations. *Bernoulli*, 14:14–46.

[30] Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.*, 15(3):897–919.

[31] Pollard, D. (1990). *Empirical processes: theory and applications*, volume 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA.

[32] Radulovic, D. and Wegkamp, M. (2018). Weak convergence of stationary empirical processes. *J. Statist. Plann. Inference*, 194:75–84.

[33] Rio, E. (2017). *Asymptotic theory of weakly dependent random processes*, volume 80 of *Probability Theory and Stochastic Modelling*. Springer, Berlin. Translated from the 2000 French edition [MR2117923].

[34] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42:43–47.

[35] Rosenblatt, M. (1991). *Stochastic Curve Estimation*, volume 3 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA.

[36] Shao, Q.-M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.*, 24(4):2098–2127.

[37] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

[38] Vapnik, V. N. and Chervonenkis, A. Y. (1982). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.*, 26:532–553.

[39] Volkonskiĭ, V. A. and Rozanov, Y. A. (1959). Some limit theorems for random functions. I. *Theory Probab. Appl.*, 4:178–197.

[40] Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154.

[41] Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.*, 41(2):425–436.

[42] Wu, W. B. and Zhao, Z. (2008). Moderate deviations for stationary processes. *Statist. Sinica*, 18(2):769–782.