LIMIT THEOREMS FOR EMPIRICAL PROCESSES BASED ON DEPENDENT DATA

Berti, Patrizia

Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy [**berti.patrizia@unimore.it**]

Leisen, Fabrizio

Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy [**leisen.fabrizio@unimore.it**]

Pratelli, Luca

Accademia Navale, viale Italia 72, 57100 Livorno, Italy [**pratel@mail.dm.unipi.it**]

Rigo, Pietro

Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia, via S. Felice 5, 27100 Pavia, Italy [**prigo@eco.unipv.it**]

Abstract Empirical processes for non ergodic data are investigated under uniform distance. Some CLT's, both uniform and non uniform, are proved. In particular, conditions for the empirical process $B_n = \sqrt{n}(\mu_n - b_n)$ to converge in distribution are given, where μ_n is the empirical measure and b_n the arithmetic mean of the first n predictive measures. Such conditions apply under various assumptions on the data, including stationarity, conditional identity in distribution and pairwise exchangeability. A characterization of conditionally identically distributed probability laws is also given.

Keywords: Conditional identity in distribution, convergence in distribution, empirical process, exchangeability, non measurable random element, stable convergence.

1 Introduction

Almost all work on empirical processes, so far, concerned i.i.d. data; cf. [8] and [13]. To our knowledge, the non independent case is almost neglected and usually restricted to the stationary and ergodic one; see e.g. [1] and references therein.

This paper deals with convergence in distribution of empirical processes, based on non ergodic data, under uniform distance. We focus on *conditionally identically distributed* (c.i.d.) sequences of random variables (see Section 4). This type of dependence, introduced in [4] and [11], includes exchangeability as a particular case and plays a role in Bayesian predictive inference.

In addition to reviewing some known facts from [4] and [5], various new results are proved. We mention a characterization of c.i.d. probability distributions (Theorem 4.1) and some uniform or non uniform CLT's (Theorems 4.2, 6.1 and 6.2, Example 5.3). In particular, conditions for the empirical process $B_n = \sqrt{n}(\mu_n - b_n)$ to converge in distribution are given, where μ_n is the empirical measure and b_n the arithmetic mean of the first n predictive measures. Such conditions apply under various assumptions on the data, including stationarity, conditional identity in distribution and pairwise exchangeability.

2 Empirical processes

Throughout, "the data" are meant as an *identically distributed* sequence $(\xi_n)_{n\geq 1}$ of random variables, defined on the probability space $(\Omega, \mathcal{A}, P)$ and taking values in the Polish space $\mathcal{X}$. Let $\mathcal{B}$ denote the Borel σ-field on $\mathcal{X}$. A random probability measure on $\mathcal{X}$ is a map γ on Ω such that: (i) $\gamma(\omega)$ is a probability measure on $\mathcal{B}$ for all $\omega \in \Omega$; (ii) $\gamma(\cdot)(B)$ is $\mathcal{A}$-measurable for all $B \in \mathcal{B}$. One example is the n-th empirical measure, that is, the random probability measure on $\mathcal{X}$ given by

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i}.$$

For any set T, let $l^\infty(T)$ denote the space of real bounded functions on T equipped with the uniform norm

$$\|\phi\| = \sup_{x\in T}|\phi(x)|, \quad \phi \in l^\infty(T).$$

To introduce empirical processes, we fix an *uniformly bounded* class $\mathcal{F}$ of real Borel measurable functions on $\mathcal{X}$ and we let $\mu = P \circ \xi_1^{-1}$, i.e., we denote μ the probability distribution common to the ξ_n. Then, in the particular case where (ξ_n) is i.i.d., the n-th empirical process is

$$G_n = \sqrt{n}(\mu_n - \mu).$$

Apart from (ξ_n) is i.i.d. or not, G_n is a real process indexed by $\mathcal{F}$ with bounded paths, so that it can be seen as a map $G_n : \Omega \to l^\infty(\mathcal{F})$. In particular, if G_n converges in distribution (as a random element of $l^\infty(\mathcal{F})$), then

$$\|\mu_n - \mu\| = \frac{1}{\sqrt{n}}\|G_n\| \xrightarrow{P} 0.$$

If (ξ_n) is not i.i.d., G_n need not be the "right" empirical process to be dealt with. A first reason is that μ is only a part, usually not the most meaningful one, of the probability distribution of the sequence (ξ_n). Thus, in the dependent case, G_n is often not much interesting from the point of view of applications. A second and more stringent reason is that, unless (ξ_n) is stationary and ergodic, $\|\mu_n - \mu\|$ typically fails to converge to 0 in probability. In this case, G_n is definitively ruled out as far as convergence in distribution is concerned.

Hence, when $\|\mu_n - \mu\|$ fails to converge to 0 in probability, empirical processes are to be defined in some different way. One option is

$$\widetilde{G}_n = r_n(\mu_n - \gamma_n),$$

where the r_n are constants such that $r_n \to \infty$ and the γ_n random probability measures on $\mathcal{X}$ satisfying $\|\mu_n - \gamma_n\| \xrightarrow{P} 0$. In this paper, we focus on $\widetilde{G}_n$ in the particular case $r_n = \sqrt{n}$.

As an example, suppose (ξ_n) is exchangeable and $\mathcal{T}$ denotes the tail σ-field of (ξ_n). By de Finetti's theorem,

$$P\big((\xi_1, \xi_2, \dots) \in B\big) = \int \gamma(\omega)^\infty(B) P(d\omega), \quad B \in \mathcal{B}^\infty,$$

where γ is a (regular) version of $P\big(\xi_1 \in \cdot \mid \mathcal{T}\big)$ and $\gamma(\omega)^\infty = \gamma(\omega) \times \gamma(\omega) \times \dots$. In this case, it is tempting to let $\gamma_n = \gamma$ for all n. The corresponding empirical process

$$W_n = \sqrt{n}(\mu_n - \gamma)$$

has been examined in [4] and [5].

As another example, suppose (ξ_n) is adapted to a filtration $(\mathcal{G}_n)_{n \geq 0}$ and define

$$a_n(\cdot) = P\big(\xi_{n+1} \in \cdot \mid \mathcal{G}_n\big).$$

In Bayesian predictive inference and discrete time filtering, one major goal is evaluating the *predictive measure* a_n. When a_n can not be calculated in closed form, one option is estimating it by data and a possible estimate is the empirical measure μ_n. Then, it is important to evaluate the limiting distribution of the error, that is, to investigate convergence in distribution of the process $r_n(\mu_n - a_n)$ for suitable constants $r_n \to \infty$. Among other things, if such a process converges in distribution then μ_n is a "consistent estimate" of a_n, since $\|\mu_n - a_n\| = \frac{1}{r_n}\|r_n(\mu_n - a_n)\| \overset{P}{\to} 0$. Thus, in Bayesian predictive problems, it is quite reasonable to let $\gamma_n = a_n$. Letting also $r_n = \sqrt{n}$ leads to the empirical process

$$C_n = \sqrt{n}(\mu_n - a_n).$$

One more interesting choice is $\gamma_n = b_n$ where $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$. In fact, there are problems where b_n plays a role, mainly in stochastic approximation, calibration and gambling; see [2] and references therein. In case of c.i.d. data (see Section 4), the corresponding empirical process

$$B_n = \sqrt{n}(\mu_n - b_n)$$

has been examined in [4].

In the sequel, the empirical processes B_n, C_n and W_n are investigated, by paying particular attention to B_n.

We finally note that, when (ξ_n) is i.i.d. and $\mathcal{G}_n = \sigma(\xi_1, \dots, \xi_n)$, B_n, C_n and W_n all reduce to G_n. Generally, however, the former are technically harder than the latter to work with. In fact, G_n is centered around the fixed measure μ, while B_n, C_n and W_n are centered around random measures (b_n, a_n and γ, respectively) possibly depending on n.

3 Convergence in distribution of non measurable random elements and stable convergence

The empirical processes B_n, C_n, G_n and W_n, regarded as maps from Ω into $l^\infty(\mathcal{F})$, can fail to be measurable if $l^\infty(\mathcal{F})$ is equipped with the Borel σ-field. To study their convergence in distribution, thus, we need a definition which works for non measurable random elements as well. One such definition is due to Hoffmann-Jørgensen. The resulting theory, developed in [8] and [13], is nice and usable in real problems. We recall here basic definitions.

Let S be a metric space and $(\Omega_0, \mathcal{A}_0, P_0)$ a probability space. A map $X : \Omega_0 \to S$ is called *measurable* if $X^{-1}(B) \in \mathcal{A}_0$ for all Borel sets $B \subset S$, and it is called *tight* if it is measurable and has a tight probability distribution. For each bounded function $Z : \Omega_0 \to \mathbb{R}$, the outer expectation of Z is defined as

$$E^* Z = \inf\{EU : U : \Omega_0 \to \mathbb{R} \text{ bounded and measurable, } U \geq Z\}.$$

Let ν be a probability measure on the Borel σ-field of S and $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ a net of probability spaces with associated maps $X_\alpha : \Omega_\alpha \to S$. For each bounded Borel function f on S, denote $\nu(f) = \int f d\nu$. Say that X_α *converges in distribution to* ν if

$$E^* f(X_\alpha) \to \nu(f) \quad \text{for all } f \in C_b(S).$$

In this case, we also write $X_\alpha \xrightarrow{d} X$ whenever X is a measurable S-valued map, defined on any probability space, with distribution ν.

The probabilistic meaning of the above definition is made more transparent by the following characterization, given in [3]. Indeed, X_α converges in distribution to ν if and only if

$$E_{Q_\alpha} f(X_\alpha) \to \nu(f) \quad \text{for all } f \in C_b(S), \text{ whenever each } Q_\alpha \text{ is a}$$
$$\text{finitely additive probability on the power set of } \Omega_\alpha \text{ extending } P_\alpha.$$

In other terms, convergence in distribution of X_α amounts to weak convergence of the probability laws $Q_\alpha \circ X_\alpha^{-1}$, in the usual sense, for *all finitely additive extensions* Q_α of P_α.

Finally, we turn to stable convergence. Let γ be a random probability measure on S and suppose that $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha) = (\Omega, \mathcal{A}, P)$ for all α. Say that X_α *converges stably to* γ in case X_α converges in distribution under $P(\cdot \mid H)$ to the probability measure $\nu_H(\cdot) = E(\gamma(\cdot) \mid H)$, for each $H \in \mathcal{A}$ with $P(H) > 0$. In other terms, stable convergence of X_α to γ means that

$$E^*\big(f(X_\alpha) \mid H\big) \to E(\gamma(f) \mid H) \quad \text{for all } f \in C_b(S) \text{ and } H \in \mathcal{A} \text{ with } P(H) > 0.$$

In particular, for $H = \Omega$, stable convergence implies convergence in distribution. Stable convergence has been introduced by Renyi in [12] and subsequently investigated by various authors (in case the X_α are measurable). We refer to [6] and [10] for more on stable convergence.

4 Conditionally identically distributed sequences of random variables

In the sequel, $\mathcal{G} = (\mathcal{G}_n)_{n \geq 0}$ is a filtration on $(\Omega, \mathcal{A}, P)$. Say that (ξ_n) is *conditionally identically distributed with respect to* $\mathcal{G}$, abbreviated $\mathcal{G}$-c.i.d., in case (ξ_n) is $\mathcal{G}$-adapted and

$$P\big(\xi_k \in \cdot \mid \mathcal{G}_n\big) = P\big(\xi_{n+1} \in \cdot \mid \mathcal{G}_n\big) \quad \text{a.s. for all } k > n \geq 0. \tag{1}$$

Roughly speaking, (1) means that, at each time $n \geq 0$, the future observations $(\xi_k)_{k>n}$ are identically distributed given the past $\mathcal{G}_n$. Condition (1) is also equivalent to

$$\xi_{T+1} \sim \xi_1 \quad \text{for each finite } \mathcal{G}\text{-stopping time } T.$$

When $\mathcal{G} = \mathcal{G}^\xi$, where $\mathcal{G}_0^\xi = \{\emptyset, \Omega\}$ and $\mathcal{G}_n^\xi = \sigma(\xi_1, \ldots, \xi_n)$, the filtration is not mentioned at all and (ξ_n) is just called c.i.d.. Clearly, if (ξ_n) is $\mathcal{G}$-c.i.d. then it is c.i.d. and identically distributed. Moreover, (ξ_n) is c.i.d. if and only if

$$\big(\xi_1, \ldots, \xi_n, \xi_{n+2}\big) \sim \big(\xi_1, \ldots, \xi_n, \xi_{n+1}\big) \quad \text{for all } n \geq 0. \tag{2}$$

Exchangeable sequences meet (2) so that are c.i.d., while the converse is not true. In fact, by a result in [11], (ξ_n) is exchangeable if and only if it is stationary and c.i.d.. Some significant examples of non exchangeable c.i.d. sequences are in [4].

One property of c.i.d. sequences, to be mentioned for later purposes, is the SLLN: If (ξ_n) is c.i.d., $\mathcal{X} = \mathbb{R}$ and $E|\xi_1| < \infty$, then $\overline{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$ converges a.s. and in L_1.

We refer to [4] for more on $\mathcal{G}$-c.i.d. sequences. Here, we prove a couple of new results.

4.1 A characterization of c.i.d. probability distributions

Let $\mathcal{X}^*$ be the set of finite sequences of elements of $\mathcal{X}$, including the empty sequence. Following [7], let us call *strategy* any collection

$$\sigma = \{\sigma(q) : q \in \mathcal{X}^*\}$$

of probabilities on $\mathcal{B}$, indexed by $\mathcal{X}^*$, such that $(x_1, \ldots, x_n) \mapsto \sigma(x_1, \ldots, x_n)(B)$ is Borel measurable for each $n \geq 1$ and $B \in \mathcal{B}$. Denote σ_0 the element of σ corresponding to the empty sequence, i.e., $\sigma_0 = \sigma(\text{empty sequence})$. Also, let π_n be the n-th coordinate projection on $\mathcal{X}^\infty$. By Ionescu Tulcea theorem, to each strategy σ there corresponds a unique probability ν on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ such that, under ν,

(IT) σ_0 is the marginal of π_1 and, for each $n \geq 1$, $\{\sigma(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in \mathcal{X}^n\}$ is a version of the conditional distribution of π_{n+1} given $(\pi_1, \ldots, \pi_n)$.

Conversely, for each probability ν on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ there is an (essentially unique) strategy σ satisfying (IT) under ν.

Let α_0 and $\alpha(x)$ be probabilities on $\mathcal{B}$, where $x \in \mathcal{X}$. As usual, say that $\{\alpha(x) : x \in \mathcal{X}\}$ is *a Markov Kernel with stationary distribution* α_0 in case, for each $B \in \mathcal{B}$, the map $x \mapsto \alpha(x)(B)$ is Borel measurable and $\alpha_0(B) = \int \alpha(x)(B)\alpha_0(dx)$. In this notation, the following result is available.

Theorem 4.1. *Let ν be a probability on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$. Then, (π_n) is c.i.d. under ν if and only if there is a strategy σ, satisfying (IT) under ν, such that*

$$\{\sigma(q, x) : x \in \mathcal{X}\} \quad \textit{is a Markov kernel with stationary distribution } \sigma(q)$$

for each fixed $q \in \mathcal{X}^*$.

Proof. Consider the probability space $(\mathcal{X}^\infty, \mathcal{B}^\infty, \nu)$ and fix a strategy σ satisfying (IT) under ν. In view of (IT), (π_n) is c.i.d. if and only if $\pi_2 \sim \pi_1$ and, for each $n \geq 1$,

$$\begin{gathered}\{\sigma(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in \mathcal{X}^n\} \quad \text{is a version of} \\ \text{the conditional distribution of } \pi_{n+2} \text{ given } (\pi_1, \ldots, \pi_n).\end{gathered} \tag{3}$$

By (IT), the condition $\pi_2 \sim \pi_1$ amounts to

$$\int \sigma(x)(\cdot)\sigma_0(dx) = \nu(\pi_2 \in \cdot) = \nu(\pi_1 \in \cdot) = \sigma_0(\cdot)$$

which just means that $\{\sigma(x) : x \in \mathcal{X}\}$ is a Markov kernel with stationary distribution σ_0. Likewise, given $n \geq 1$, condition (3) is equivalent to

$$\int \sigma(x_1, \ldots, x_n, x)(\cdot)\sigma(x_1, \ldots, x_n)(dx) = \sigma(x_1, \ldots, x_n)(\cdot)$$

for almost all $(x_1, \ldots, x_n) \in \mathcal{X}^n$, where "almost all" is meant with respect to the probability distribution of $(\pi_1, \ldots, \pi_n)$. This proves the "if" part of the Theorem, while for the "only if" part it suffices to modify the strategy σ on a ν-null set. $\square$

Practically, Theorem 4.1 suggests how to assess c.i.d. sequences stepwise. First, select a law σ_0 on $\mathcal{B}$, the marginal distribution of ξ_1. Then, choose a Markov kernel $\{\sigma(x) : x \in \mathcal{X}\}$ with stationary distribution σ_0, where $\sigma(x)$ should be viewed as the conditional distribution of ξ_2 given $\xi_1 = x$. Next, for each $x \in \mathcal{X}$, select a Markov kernel $\{\sigma(x, y) : y \in \mathcal{X}\}$ with stationary distribution $\sigma(x)$, where $\sigma(x, y)$ should be viewed as the conditional distribution of ξ_3 given $\xi_1 = x$ and $\xi_2 = y$. And so on. In other terms, for getting a c.i.d. sequence, it is enough assigning at each step a Markov kernel with a given stationary distribution. Indeed, a plenty of methods for doing this are available: see the statistical literature concerning MCMC.

4.2 A CLT for c.i.d. sequences

For $a \in \mathbb{R}^k$ and Σ a symmetric nonnegative definite matrix of order k, let $N(a, \Sigma)$ denote the Gaussian law on $\mathbb{R}^k$ with mean a and covariance matrix Σ. If Σ is the null matrix, we let $N(a, 0) = \delta_a$. Note that, for $k = 1$, $N(0, L)$ is a random probability measure on $\mathbb{R}$ whenever L is a real nonnegative random variable.

Next result unifies and slightly improves Theorem 3.3 of [4]. In case $\mathcal{X} = \mathbb{R}$ and $E|\xi_1| < \infty$, define $\overline{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$ and

$$U_n = \sqrt{n}\big(\overline{\xi}_n - E(\xi_{n+1} \mid \mathcal{G}_n)\big).$$

Theorem 4.2. *Suppose $\mathcal{X} = \mathbb{R}$, (ξ_n) is $\mathcal{G}$-c.i.d., $E\xi_1^2 < \infty$ and $\sup_n EU_n^2 < \infty$. Let*

$$D_n = \big(\xi_n - E(\xi_n \mid \mathcal{G}_{n-1})\big) - n\big(E(\xi_{n+1} \mid \mathcal{G}_n) - E(\xi_n \mid \mathcal{G}_{n-1})\big).$$

In order to $U_n \to N(0, L)$ stably, where L is a real nonnegative random variable, it is enough that

$$M_n = \frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{P} L, \quad \frac{1}{n}\max_{i \leq n} D_i^2 \xrightarrow{P} 0. \tag{4}$$

In particular,

(i) $U_n \to N(0, L)$ stably if $M_n \to L$ a.s.;

(ii) $U_n \to N(0, L)$ stably, where $L = \lim_n \frac{1}{n}\sum_{i=1}^n \big(\xi_i - \overline{\xi}_n\big)^2$ a.s., if

$$\frac{1}{n}\sum_{i=1}^n i^2\big(E(\xi_{i+1} \mid \mathcal{G}_i) - E(\xi_i \mid \mathcal{G}_{i-1})\big)^2 \xrightarrow{P} 0. \tag{5}$$

Proof. For $n \geq 1$ and $j = 1, \ldots, n$, define $Y_{n,j} = \frac{D_j}{\sqrt{n}}$, $\mathcal{F}_{n,0} = \mathcal{G}_0$ and $\mathcal{F}_{n,j} = \mathcal{G}_j$. Then, $\mathcal{F}_{n+1,j} = \mathcal{F}_{n,j}$ and $Y_{n,j}$ is $\mathcal{F}_{n,j}$-measurable. Since (ξ_n) is $\mathcal{G}$-c.i.d.,

$$E(Y_{n,j} \mid \mathcal{F}_{n,j-1}) = \frac{j}{\sqrt{n}}\big(E(\xi_{j+1} \mid \mathcal{G}_{j-1}) - E(\xi_j \mid \mathcal{G}_{j-1})\big) = 0 \quad \text{a.s.}.$$

Letting

$$V_j = E(\xi_{j+1} \mid \mathcal{G}_j),$$

one also obtains

$$\sum_{j=1}^{n} Y_{n,j} = \frac{1}{\sqrt{n}}\Big(n\overline{\xi}_n - \sum_{j=1}^{n}(jV_j - (j-1)V_{j-1})\Big) = \frac{1}{\sqrt{n}}\Big(n\overline{\xi}_n - nV_n\Big) = U_n.$$

Therefore,

$$\sup_n E\big(\max_{j\le n} Y_{n,j}^2\big) \le \sup_n E\Big(\sum_{j=1}^{n} Y_{n,j}^2\Big) = \sup_n EU_n^2 < \infty.$$

Moreover, condition (4) implies

$$\sum_{j=1}^{n} Y_{n,j}^2 = M_n \xrightarrow{P} L, \quad \max_{j\le n} Y_{n,j}^2 = \frac{1}{n}\max_{j\le n} D_j^2 \xrightarrow{P} 0.$$

Thus, the martingale CLT yields $U_n = \sum_{j=1}^{n} Y_{n,j} \to N(0,L)$ stably; see Theorem 3.2, p. 58, of [10]. This concludes the proof of the first part of the Theorem.

As to (i), if $M_n \to L$ a.s. then $Y_{n,n}^2 = M_n - \frac{n-1}{n}M_{n-1} \to 0$ a.s., and thus

$$\max_{j\le n} Y_{n,j}^2 = \max_{j\le n} \frac{j}{n} Y_{j,j}^2 \to 0 \quad \text{a.s.}.$$

We now prove (ii). Suppose (5) holds. By the SLLN, $\overline{\xi}_n \to V$ a.s. and $\frac{1}{n}\sum_{i=1}^{n}\xi_i^2 \to Z$ a.s. for some real random variables V and Z. Since (V_n) is a $\mathcal{G}$-martingale and $EV_n^2 \le E\xi_1^2$ for all n, one also obtains $V_n \to V$ a.s.. Recall that, if (r_n) and (s_n) are real sequences, then $\frac{1}{n}\sum_{i=1}^{n} r_i s_i \to rs$ provided $s_n \to s$, $\frac{1}{n}\sum_{i=1}^{n} r_i \to r$ and $r_n \ge 0$ for all n. Hence,

$$\frac{1}{n}\sum_{i=1}^{n} V_{i-1}\xi_i = \frac{1}{n}\sum_{i=1}^{n} V_{i-1}\xi_i^+ - \frac{1}{n}\sum_{i=1}^{n} V_{i-1}\xi_i^- \to V\Big(\lim_n \frac{1}{n}\sum_{i=1}^{n}\xi_i^+\Big) - V\Big(\lim_n \frac{1}{n}\sum_{i=1}^{n}\xi_i^-\Big) = V^2 \quad \text{a.s.}.$$

It follows that

$$A_{1,n} := \frac{1}{n}\sum_{i=1}^{n}(\xi_i - V_{i-1})^2 \to Z + V^2 - 2V^2 = \lim_n \frac{1}{n}\sum_{i=1}^{n}\big(\xi_i - \overline{\xi}_n\big)^2 \text{ a.s.}.$$

Define further

$$A_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - V_{i-1})(V_i - V_{i-1})i, \quad A_{3,n} = \frac{1}{n}\sum_{i=1}^{n}\Big((V_i - V_{i-1})i\Big)^2.$$

By (5), $A_{3,n} \xrightarrow{P} 0$. Thus,

$$A_{2,n}^2 \le \Big(\frac{1}{n}\sum_{i=1}^{n}|\xi_i - V_{i-1}||V_i - V_{i-1}|i\Big)^2 \le A_{1,n}A_{3,n} \xrightarrow{P} 0.$$

Therefore,

$$M_n = A_{1,n} - 2A_{2,n} + A_{3,n} \xrightarrow{P} \lim_n \frac{1}{n}\sum_{i=1}^{n}\big(\xi_i - \overline{\xi}_n\big)^2.$$

It remains to prove that $\frac{1}{n}\max_{i\le n} D_i^2 \xrightarrow{P} 0$. Let

$$B_{1,n}=\frac{1}{n}\max_{i\le n}(\xi_i-V_{i-1})^2,\quad B_{2,n}=\frac{1}{n}\max_{i\le n}\Big((V_i-V_{i-1})i\Big)^2.$$

Since

$$\frac{1}{n}\max_{i\le n} D_i^2\le B_{1,n}+B_{2,n}+2\sqrt{B_{1,n}B_{2,n}}\quad\text{ and }\quad B_{2,n}\le A_{3,n}\xrightarrow{P}0,$$

it is enough proving that $B_{1,n}\xrightarrow{P}0$. To this end, note that $\frac{1}{n}\max_{i\le n}V_{i-1}^2\to 0$ a.s. and

$$E\big(\frac{1}{n}\max_{i\le n}\xi_i^2\big)\le\frac{c^2}{n}+\frac{1}{n}\sum_{i=1}^n E\big(\xi_i^2 I_{\{|\xi_i|>c\}}\big)=\frac{c^2}{n}+E\big(\xi_1^2 I_{\{|\xi_1|>c\}}\big)\quad\text{for all } c>0.$$

Thus,

$$B_{1,n}\le\frac{1}{n}\max_{i\le n}\xi_i^2+\frac{1}{n}\max_{i\le n}V_{i-1}^2+2\sqrt{\big(\frac{1}{n}\max_{i\le n}\xi_i^2\big)\big(\frac{1}{n}\max_{i\le n}V_{i-1}^2\big)}\xrightarrow{P}0,$$

and this concludes the proof. □

5 Back to empirical processes

In this section, some empirical process theory for c.i.d. data is summarized. With the only exception of Example 5.3, which is new, all other results are from [4] and [5].

Suppose (ξ_n) is $\mathcal{G}$-c.i.d.. Letting $\mu_n=\frac{1}{n}\sum_{i=1}^n\delta_{\xi_i}$ denote the n-th empirical measure, by the SLLN, there is a random probability measure γ on $\mathcal{X}$ such that

$$\mu_n(\omega)\to\gamma(\omega)\quad\text{weakly, for almost all }\omega.$$

According to Section 2, we let $a_n(\cdot)=P\big(\xi_{n+1}\in\cdot\mid\mathcal{G}_n\big)$, $b_n=\frac{1}{n}\sum_{i=0}^{n-1}a_i$ and

$$B_n=\sqrt{n}(\mu_n-b_n),\quad C_n=\sqrt{n}(\mu_n-a_n),\quad W_n=\sqrt{n}(\mu_n-\gamma).$$

¿From now on, we focus on the particular case

$$\mathcal{X}=\mathbb{R}\quad\text{ and }\quad\mathcal{F}=\{I_{(-\infty,t]}:t\in\mathbb{R}\}.$$

Accordingly, we write $B_n(t)$ instead of $B_n(I_{(-\infty,t]})$ and we regard B_n as a map $B_n:\Omega\to l^\infty(\mathbb{R})$. The same for C_n and W_n. Moreover, F_γ denotes the random distribution function corresponding to γ, i.e.

$$F_\gamma(t)=\gamma(-\infty,t],\quad t\in\mathbb{R}.$$

A possible limit in distribution for B_n, C_n or W_n is a tight random element $\mathbb{G}:\Omega_0\to l^\infty(\mathbb{R})$, defined on some probability space $(\Omega_0,\mathcal{A}_0,P_0)$, with distribution

$$P_0\big((\mathbb{G}(t_1),\dots,\mathbb{G}(t_k))\in A\big)=\int N\big(0,\Sigma(t_1,\dots,t_k)\big)(A)dP \tag{6}$$

where $t_1,\dots,t_k\in\mathbb{R}$, $A\in\mathcal{B}^k$ and $\Sigma(t_1,\dots,t_k)$ is a random covariance matrix on $(\Omega,\mathcal{A},P)$. One significant particular case is $\mathbb{G}(t)=\mathbb{B}(F(t))$, $t\in\mathbb{R}$, where $\mathbb{B}$ is a standard Brownian bridge on $[0,1]$ and F an independent copy of F_γ. Precisely, F is a random distribution function, independent of $\mathbb{B}$, such that $F \sim F_\gamma$ (recall that $\mathbb{B}$ and F are defined on $(\Omega_0, \mathcal{A}_0, P_0)$). In this case, (6) holds with

$$\Sigma(t_1, \ldots, t_k) = \big(F_\gamma(t_i \wedge t_j)(1 - F_\gamma(t_i \vee t_j)) : 1 \leq i, j \leq k\big).$$

Let us denote $\mathbb{G}^F$ such a process, i.e.

$$\mathbb{G}^F(t) = \mathbb{B}(F(t)) \quad \text{for all } t \in \mathbb{R}.$$

Generally, $\mathbb{G}^F : \Omega_0 \to l^\infty(\mathbb{R})$ can fail to be measurable if $l^\infty(\mathbb{R})$ is equipped with the Borel σ-field. However, $\mathbb{G}^F$ is measurable and tight whenever every F-path is continuous on A^c for some *fixed* countable set $A \subset \mathbb{R}$.

As a trivial example, suppose $\mathcal{G} = \mathcal{G}^\xi$ and (ξ_n) is i.i.d.. Then $F = H$ a.s., where H is the distribution function common to the ξ_n, so that $A = \{t : H(t) > H(t-)\}$. Thus, $\mathbb{G}^F = \mathbb{G}^H$ is measurable and tight and $G_n \xrightarrow{d} \mathbb{G}^H$ (recall that $B_n = C_n = W_n = G_n$ in this particular case).

Let (Z_n) be any sequence of real random processes indexed by $\mathbb{R}$, with bounded cadlag paths, defined on $(\Omega, \mathcal{A}, P)$. Then, a necessary condition for Z_n to converge in distribution to a tight limit is: For all $\epsilon, \eta > 0$, there is a finite partition $I_1, \ldots, I_m$ of $\mathbb{R}$ such that

$$\limsup_n P\big(\max_j \sup_{s,t \in I_j} |Z_n(s) - Z_n(t)| > \epsilon\big) < \eta. \tag{7}$$

We are now able to state a couple of results.

Theorem 5.1. *If (ξ_n) is $\mathcal{G}$-c.i.d. and B_n meets (7) (i.e., (7) holds with $Z_n = B_n$), then $B_n \xrightarrow{d} \mathbb{G}^F$ and $\mathbb{G}^F$ is tight.*

Theorem 5.2. *Suppose (ξ_n) is $\mathcal{G}$-c.i.d., C_n meets (7), and $\sup_n E\big(C_n(t)^2\big) < \infty$ for all $t \in \mathbb{R}$. If*

$$\frac{1}{n}\sum_{i=1}^n q_i(s) q_i(t) \to \sigma(s,t) \quad \textit{a.s. for all } s,t \in \mathbb{R}$$
$$\textit{where} \quad q_i(t) = I_{\{\xi_i \leq t\}} - iP\big(\xi_{i+1} \leq t \mid \mathcal{G}_i\big) + (i-1)P\big(\xi_i \leq t \mid \mathcal{G}_{i-1}\big),$$

then $C_n \xrightarrow{d} \mathbb{G}$, where $\mathbb{G}$ is a tight process with distribution (6) and $\Sigma(t_1, \ldots, t_k) = \big(\sigma(t_i, t_j) : 1 \leq i, j \leq k\big)$. Moreover, $C_n \xrightarrow{d} \mathbb{G}^F$ and $\mathbb{G}^F$ is tight whenever

$$\frac{1}{n}\sum_{i=1}^n i^2 \big[P\big(\xi_{i+1} \leq t \mid \mathcal{G}_i\big) - P\big(\xi_i \leq t \mid \mathcal{G}_{i-1}\big)\big]^2 \xrightarrow{P} 0 \quad \textit{for all } t \in \mathbb{R}.$$

Both Theorems 5.1 and 5.2 require condition (7), that is, asymptotic tightness. Thus, it would be useful to have some criterion for testing it. In the exchangeable case, one such criterion is tightness of the process G^F. Indeed, checking tightness of G^F is often not very hard. Unfortunately, as it is shown in the next example, this criterion can fail in the general $\mathcal{G}$-c.i.d. case.

Example 5.3. Let (α_n) and (β_n) be independent sequences of independent real random variables, with $\alpha_n \sim N(0, c_n - c_{n-1})$ and $\beta_n \sim N(0, 1 - c_n)$ where $c_n = 1 - (\frac{1}{n+1})^{\frac{1}{5}}$. Define

$$\xi_n = \sum_{i=1}^{n} \alpha_i + \beta_n, \quad \mathcal{G}_0 = \{\emptyset, \Omega\}, \quad \mathcal{G}_n = \sigma(\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n).$$

In Example 1.2 of [4], it is shown that (ξ_n) is $\mathcal{G}$-c.i.d. and $\xi_n \to \xi$ a.s. for some random variable ξ. Since $\xi_n \to \xi$ a.s.,

$$\mu_n(\omega) \to \delta_{\xi(\omega)} \quad \text{weakly, for almost all } \omega.$$

Hence, $\gamma = \delta_\xi$ and $G^F = 0$, so that G^F is tight.

The finite dimensional distributions of the processes C_n and W_n converge weakly (to 0). To see this, first note that

$$P\big(\xi_{n+1} \leq t \mid \mathcal{G}_n\big) = \Phi\Big(\frac{t - S_n}{\sqrt{1 - c_n}}\Big)$$

where $S_n = \sum_{i=1}^{n} \alpha_i$ and Φ is the standard normal distribution function. Hence,

$$C_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} I_{\{\xi_i \leq t\}} - I_{\{S_n \leq t\}}\Big) + \sqrt{n}\Big(I_{\{S_n \leq t\}} - \Phi\big(\frac{t - S_n}{\sqrt{1 - c_n}}\big)\Big),$$

$$W_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} I_{\{\xi_i \leq t\}} - I_{\{\xi \leq t\}}\Big).$$

Given t, since both $\xi_n \to \xi$ a.s. and $S_n \to \xi$ a.s., it can be checked that $C_n(t) \xrightarrow{P} 0$ and $W_n(t) \to 0$ a.s..

Finally, condition (7) fails for C_n and W_n. As to W_n, under (7), one would obtain $W_n \xrightarrow{d} 0$ so that $\sup_t |W_n(t) - W_n(t-)| \xrightarrow{P} 0$. But, $P(\xi_i \neq \xi \text{ for all } i) = 1$ and

$$\sup_t \Big|W_n(t) - W_n(t-)\Big| \geq \Big|W_n(\xi) - W_n(\xi-)\Big| = \sqrt{n} \quad \text{on the set } \{\xi_i \neq \xi \text{ for all } i\}.$$

Toward a contradiction, suppose now that C_n meets (7) and define

$$I_n = \int_{S_n - 1}^{S_n + 1} C_n(t)\,dt.$$

Then $C_n \xrightarrow{d} 0$, so that $|I_n| \leq 2 \sup_t |C_n(t)| \xrightarrow{P} 0$. On the other hand,

$$\begin{aligned} I_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(S_n + 1 - \xi_i \vee (S_n - 1)\Big)^+ - \sqrt{n} \int_{S_n - 1}^{S_n + 1} \Phi\big(\frac{t - S_n}{\sqrt{1 - c_n}}\big)dt \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(S_n + 1 - \xi_i \vee (S_n - 1)\Big)^+ - \sqrt{n}. \end{aligned}$$

Let

$$J_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(S_n + 1 - \xi_i\Big) - \sqrt{n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (S_n - \xi_i).$$

Then $I_n - J_n \to 0$ a.s., due to $S_n - \xi_n \to 0$ a.s., and thus $J_n \xrightarrow{P} 0$. But this is a contradiction, since $J_n \sim N(0, \sigma_n^2)$ with $\sigma_n^2 \to \infty$. Precisely,

$$\sigma_n^2 = -\frac{n}{(n+1)^{\frac{1}{5}}} + \frac{2}{n}\sum_{i=1}^{n} \frac{i}{(i+1)^{\frac{1}{5}}} \quad \text{so that} \quad \frac{\sigma_n^2}{n^{\frac{4}{5}}} \to \frac{1}{9}.$$

In the rest of this section, (ξ_n) is assumed *exchangeable.* Then, a nicer asymptotic theory is available.

Theorem 5.4. *If* (ξ_n) *is exchangeable,* $\mathcal{G} = \mathcal{G}^\xi$*, and* $\mathbb{G}^F$ *is tight, then*

$$B_n \xrightarrow{d} \mathbb{G}^F \quad \textit{and} \quad W_n \xrightarrow{d} \mathbb{G}^F.$$

Moreover, C_n *is asymptotically tight and relatively sequentially compact.*

Unlike Theorems 5.1 and 5.2, Theorem 5.4 includes a criterion for asymptotic tightness (i.e., tightness of $\mathbb{G}^F$). This makes Theorem 5.4 usable in various real problems. For instance, $\mathbb{G}^F$ is tight if ξ_1 has a discrete distribution or if $P(\xi_1 = \xi_2) = 0$.

When $\mathbb{G}^F$ is not tight, W_n can fail to converge in distribution due to non existence of any measurable limit; see Example 11 of [3]. However, something can be said even if $\mathbb{G}^F$ is not tight. Precisely, if $(\Omega, \mathcal{A}, P)$ is a perfect probability space, there is a random element W of $l^\infty(\mathbb{R})$ satisfying

$$E^* f(W_n) \to E^* f(W)$$

for all bounded uniformly continuous functions f on $l^\infty(\mathbb{R})$. Such a W is defined on $(\Omega, \mathcal{A}, P)$, it is not necessarily measurable if $l^\infty(\mathbb{R})$ is equipped with the Borel σ-field, while it is measurable if $l^\infty(\mathbb{R})$ is equipped with the smaller ball σ-field. Moreover, on the ball σ-field, the distributions of W and $\mathbb{G}^F$ coincide.

6 A uniform CLT for the empirical process B_n

In this section, (ξ_n) is a $\mathcal{G}$-adapted and identically distributed sequence of real random variables. In order to investigate the asymptotic behaviour of B_n, which is our main goal, we first prove a CLT. To this end, a set of reasonable assumptions is

$$E\xi_1^2 < \infty, \quad \frac{1}{n}\sum_{i=1}^{n} E(\xi_i \mid \mathcal{G}_{i-1})^2 \xrightarrow{P} Y, \quad \frac{1}{n}\sum_{i=1}^{n} \xi_i^2 \xrightarrow{P} Z \tag{8}$$

for some real random variables Y and Z.

Theorem 6.1. *If* (ξ_n) *is* $\mathcal{G}$*-adapted, identically distributed and meets condition* (8)*, then*

$$\sqrt{n}(\overline{\xi}_n - m_n) \to N(0, Z - Y) \textit{ stably, where } m_n = \frac{1}{n}\sum_{i=1}^{n} E(\xi_i \mid \mathcal{G}_{i-1}).$$

Proof. For $n \geq 1$ and $j = 1, \ldots, n$, define $Y_{n,j} = \frac{\xi_j - E(\xi_j \mid \mathcal{G}_{j-1})}{\sqrt{n}}$, $\mathcal{F}_{n,0} = \mathcal{G}_0$ and $\mathcal{F}_{n,j} = \mathcal{G}_j$. Then, $Y_{n,j}$ is $\mathcal{F}_{n,j}$-measurable, $\mathcal{F}_{n+1,j} = \mathcal{F}_{n,j}$, and

$$E(Y_{n,j} \mid \mathcal{F}_{n,j-1}) = 0 \text{ a.s.}.$$

So, by the martingale CLT (see Theorem 3.2, p. 58, of [10]), it suffices proving that

$$\sum_{j=1}^{n} Y_{n,j}^2 \overset{P}{\to} Z - Y, \quad \max_{j\leq n}|Y_{n,j}| \overset{P}{\to} 0, \quad \sup_n E\big(\max_{j\leq n} Y_{n,j}^2\big) < \infty.$$

Let $Z_n = \xi_n - E(\xi_n \mid \mathcal{G}_{n-1})$. Since $E\xi_1^2 < \infty$ and (ξ_n) is identically distributed, (ξ_n^2) is uniformly integrable, so that (Z_n^2) is uniformly integrable, too. Given $\epsilon > 0$, take $a > 0$ such that $E\big(Z_j^2 I_{\{|Z_j|>a\}}\big) < \epsilon$ for all j. Then,

$$E\big(\max_{j\leq n} Y_{n,j}^2\big) \leq \frac{a^2}{n} + \frac{1}{n}\sum_{j=1}^{n} E\big(Z_j^2 I_{\{|Z_j|>a\}}\big) < \frac{a^2}{n} + \epsilon.$$

Hence, $\lim_n E\big(\max_{j\leq n} Y_{n,j}^2\big) = 0$, and this implies that $\max_{j\leq n}|Y_{n,j}| \overset{P}{\to} 0$ and $\sup_n E\big(\max_{j\leq n} Y_{n,j}^2\big) < \infty$.

Since (ξ_n^2) is uniformly integrable, given $\epsilon > 0$, there is $a > 0$ such that $E(I_{A_i}\xi_i^2) < \epsilon$ for all i, where $A_i = \{|E(\xi_i \mid \mathcal{G}_{i-1})| > a\}$. Hence,

$$E\Big|\frac{1}{n}\sum_{i=1}^{n} I_{A_i} E(\xi_i \mid \mathcal{G}_{i-1}) Z_i\Big| \leq \frac{2}{n}\sum_{i=1}^{n} E\big(I_{A_i} E(\xi_i^2 \mid \mathcal{G}_{i-1})\big) = \frac{2}{n}\sum_{i=1}^{n} E\big(I_{A_i}\xi_i^2\big) < 2\epsilon.$$

Next, define the $\mathcal{G}$-martingale

$$L_n = \sum_{i=1}^{n} \frac{(1 - I_{A_i}) E(\xi_i \mid \mathcal{G}_{i-1}) Z_i}{i}.$$

On noting that $EZ_i^2 \leq 4E\xi_1^2$ for all i,

$$\sum_n E\big((L_n - L_{n-1})^2\big) \leq \sum_n \frac{4a^2 E\xi_1^2}{n^2} < \infty.$$

Thus, L_n converges a.s., and Kronecker lemma yields

$$\frac{1}{n}\sum_{i=1}^{n}(1 - I_{A_i}) E(\xi_i \mid \mathcal{G}_{i-1}) Z_i \to 0 \text{ q.c..}$$

It follows that

$$\frac{1}{n}\sum_{i=1}^{n} E(\xi_i \mid \mathcal{G}_{i-1})(\xi_i - E(\xi_i \mid \mathcal{G}_{i-1})) = \frac{1}{n}\sum_{i=1}^{n} E(\xi_i \mid \mathcal{G}_{i-1}) Z_i \overset{P}{\to} 0.$$

Finally, by (8), the latter fact implies $\sum_{j=1}^{n} Y_{n,j}^2 \overset{P}{\to} Z - Y$. □

Theorem 6.1 applies to a number of real situations. We mention, among others, three meaningful examples. Suppose $E\xi_1^2 < \infty$ and (ξ_n) is $\mathcal{G}$-c.i.d. or stationary or 2-exchangeable (that is, $(\xi_i, \xi_j) \sim (\xi_1, \xi_2)$ for all $i \neq j$). In all these cases, $\frac{1}{n}\sum_{i=1}^{n} \xi_i^2$ converges a.s.. (In the 2-exchangeable one, just note that (ξ_n^2) is still 2-exchangeable and apply the SLLN of Etemadi and Kaminski in [9]). If (ξ_n) is stationary or 2-exchangeable, convergence in probability of $\frac{1}{n}\sum_{i=1}^{n} E\big(\xi_i \mid \mathcal{G}_{i-1}\big)^2$ is not granted, though it holds in various problems. Instead, $\frac{1}{n}\sum_{i=1}^{n} E\big(\xi_i \mid \mathcal{G}_{i-1}\big)^2$ converges a.s. whenever (ξ_n) is $\mathcal{G}$-c.i.d..

In fact, $E(\xi_n \mid \mathcal{G}_{n-1})$ converges a.s. and $\lim_n E(\xi_n \mid \mathcal{G}_{n-1}) = \lim_n \overline{\xi}_n$ a.s.. Thus, in the $\mathcal{G}$-c.i.d. case one obtains

$$Z - Y = \lim_n \frac{1}{n}\sum_{i=1}^n (\xi_i - \overline{\xi}_n)^2 \quad \text{a.s.};$$

see also Section 3 of [4].

Basing on Theorem 6.1, we now prove a uniform CLT for the empirical process

$$B_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n I_{\{\xi_i \le t\}} - \frac{1}{n}\sum_{i=1}^n P\big(\xi_i \le t \mid \mathcal{G}_{i-1}\big)\Big), \quad t \in \mathbb{R}.$$

Our last result extends Theorem 5.1 (from $\mathcal{G}$-c.i.d.) to various other types of dependence for the sequence (ξ_n).

Theorem 6.2. *Suppose (ξ_n) is $\mathcal{G}$-adapted and identically distributed, B_n meets condition (7), and for all $s, t \in \mathbb{R}$:*

$$\frac{1}{n}\sum_{i=1}^n I_{\{\xi_i \le t\}} \xrightarrow{P} a(t), \quad \frac{1}{n}\sum_{i=1}^n P(\xi_i \le s \mid \mathcal{G}_{i-1})P(\xi_i \le t \mid \mathcal{G}_{i-1}) \xrightarrow{P} b(s,t).$$

Then $B_n \xrightarrow{d} \mathbb{G}$, where $\mathbb{G}$ is a tight process with distribution (6) and

$$\Sigma(t_1, \ldots, t_k) = \big(a(t_i \wedge t_j) - b(t_i, t_j) : 1 \le i, j \le k\big).$$

Proof. By (7), it is enough proving that the finite dimensional distributions of B_n converge weakly to those of $\mathbb{G}$; see e.g. Theorem 1.5.4 of [13]. Fix $t_1, \ldots, t_k, c_1, \ldots, c_k \in \mathbb{R}$, define

$$\sigma^2 = \sum_{r=1}^k \sum_{s=1}^k c_r c_s \big(a(t_r \wedge t_s) - b(t_r, t_s)\big),$$

and note that the real random variable $\sum_{r=1}^k c_r \mathbb{G}(t_r)$ has distribution

$$\nu(\cdot) = \int \mathcal{N}(0, \sigma^2)(\cdot)dP.$$

Next, define $f = \sum_{r=1}^k c_r I_{(-\infty, t_r]}$ and consider the sequence $(f(\xi_n))$. Then,

$$\frac{1}{n}\sum_{i=1}^n f(\xi_i)^2 = \sum_{r=1}^k \sum_{s=1}^k c_r c_s \frac{1}{n}\sum_{i=1}^n I_{\{\xi_i \le t_r \wedge t_s\}} \xrightarrow{P} \sum_{r=1}^k \sum_{s=1}^k c_r c_s a(t_r \wedge t_s).$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^n E\big(f(\xi_i) \mid \mathcal{G}_{i-1}\big)^2 = \frac{1}{n}\sum_{i=1}^n \Big(\sum_{r=1}^k c_r P\big(\xi_i \le t_r \mid \mathcal{G}_{i-1}\big)\Big)^2$$

$$= \sum_{r=1}^k \sum_{s=1}^k c_r c_s \frac{1}{n}\sum_{i=1}^n P\big(\xi_i \le t_r \mid \mathcal{G}_{i-1}\big)P\big(\xi_i \le t_s \mid \mathcal{G}_{i-1}\big) \xrightarrow{P} \sum_{r=1}^k \sum_{s=1}^k c_r c_s b(t_r, t_s).$$

Thus, Theorem 6.1 applies to $(f(\xi_n))$ with $Z - Y = \sigma^2$, so that

$$\sum_{r=1}^k c_r B_n(t_r) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n f(\xi_i) - \frac{1}{n}\sum_{i=1}^n E\big(f(\xi_i) \mid \mathcal{G}_{i-1}\big)\Big) \to N(0, \sigma^2) \text{ stably.}$$

In particular, $\sum_{r=1}^k c_r B_n(t_r)$ converges in distribution to ν, i.e., $\sum_{r=1}^k c_r B_n(t_r) \xrightarrow{d} \sum_{r=1}^k c_r \mathbb{G}(t_r)$. By letting $c_1, \ldots, c_k$ vary, it follows that $(B_n(t_1), \ldots, B_n(t_k)) \xrightarrow{d} (\mathbb{G}(t_1), \ldots, \mathbb{G}(t_k))$. □

References

[1] Ben Hariz S. (2005) Uniform CLT for empirical process. *Stoch. Proc. Appl.*, Vol. 115. 339-358.

[2] Berti P., Rigo P. (2002) A uniform limit theorem for predictive distributions. *Statist. Probab. Letters*, Vol. 56. 113–120.

[3] Berti P., Rigo P. (2004) Convergence in distribution of non measurable random elements. *Ann. Probab.*, Vol. 32. 365–379.

[4] Berti P., Pratelli L., Rigo P. (2004) Limit theorems for a class of identically distributed random variables. *Ann. Probab.*, Vol. 32. 2029–2052.

[5] Berti P., Pratelli L., Rigo P. (2005) Asymptotic behaviour of the empirical process for exchangeable data. *Stoch. Proc. Appl.*, to appear.

[6] Crimaldi I., Letta G., Pratelli L. (2005) A strong form of stable convergence. *Seminaire de Probabilites*, to appear.

[7] Dubins L.E., Savage L.J. (1965) *How to gamble if you must: Inequalities for stochastic processes.* McGraw Hill.

[8] Dudley R. (1999) *Uniform central limit theorems.* Cambridge University Press.

[9] Etemadi N., Kaminski M. (1996) Strong law of large numbers for 2-exchangeable random variables. *Statist. Probab. Letters*, Vol. 28. 245–250.

[10] Hall P., Heyde C.C. (1980) *Martingale limit theory and its applications.* Academic Press.

[11] Kallenberg O. (1988) Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.*, Vol. 16. 508–534.

[12] Renyi A. (1963) On stable sequences of events. *Sankhya* A, Vol. 25. 293–302.

[13] van der Vaart A., Wellner J.A. (1996) *Weak convergence and empirical processes.* Springer.