Berti, Patrizia; Pratelli, Luca; Rigo, Pietro

Working Paper

Limit Theorems for Empirical Processes Based on Dependent Data

Quaderni di Dipartimento, No. 132

Provided in Cooperation with:
University of Pavia, Department of Economics and Quantitative Methods (EPMQ)

Suggested Citation: Berti, Patrizia; Pratelli, Luca; Rigo, Pietro (2010) : Limit Theorems for Empirical Processes Based on Dependent Data, Quaderni di Dipartimento, No. 132, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia

This Version is available at:
https://hdl.handle.net/10419/95307

Quaderni di Dipartimento

Limit Theorems for Empirical Processes Based on Dependent Data

Patrizia Berti
(Università di Modena e Reggio Emilia)

Luca Pratelli
(Accademia Navale di Livorno)

Pietro Rigo
(Università di Pavia)

132 (11-10)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Novembre 2010

LIMIT THEOREMS FOR EMPIRICAL PROCESSES BASED ON DEPENDENT DATA

PATRIZIA BERTI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. Empirical processes for non ergodic data are investigated under uniform distance. Some CLTs, both uniform and non uniform, are proved. In particular, conditions for $B_n = \sqrt{n}\,(\mu_n - b_n)$ and $C_n = \sqrt{n}\,(\mu_n - a_n)$ to converge in distribution are given, where μ_n is the empirical measure, a_n the predictive measure, and $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$. Such conditions can be applied to any adapted sequence of random variables. Various examples and a characterization of conditionally identically distributed sequences are given as well.

1. INTRODUCTION

Almost all work on empirical processes, so far, concerned ergodic sequences (X_n) of random variables. Slightly abusing terminology, here, (X_n) is called *ergodic* if the underlying probability measure P is 0-1 valued on the sub-σ-field

$$\sigma\big(\limsup_n \frac{1}{n}\sum_{i=1}^{n} I_B(X_i) : B \text{ a measurable set}\big).$$

In real problems, however, (X_n) is often non ergodic in the previous sense. Most stationary sequences, for instance, are non ergodic. Or else, an exchangeable sequence is ergodic if and only if it is i.i.d..

This paper deals with convergence in distribution of empirical processes, based on non ergodic data, under uniform distance. Special attention is paid to *conditionally identically distributed* (c.i.d.) sequences of random variables (see Section 4). This type of dependence, introduced in [4] and [15], includes exchangeability as a particular case and plays a role in Bayesian inference.

The paper is organized as follows. Sections 2, 3 and 5 include preliminary material (the only new result in these sections is Example 7). In particular, empirical processes for non ergodic data are discussed in Section 2 while the case of c.i.d. data is reviewed in Section 5. The core of the paper is in Sections 4 and 6. Section 4 includes a characterization of c.i.d. sequences and a couple of examples. Section 6 contains some uniform and non uniform CLTs. Suppose (X_n) is adapted to a filtration $(\mathcal{G}_n)$ and let $a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ denote the *predictive measure*. Our main results (Theorems 9, 10, 11) provide conditions for the empirical processes

$$B_n = \sqrt{n}\,(\mu_n - b_n) \quad \text{and} \quad C_n = \sqrt{n}\,(\mu_n - a_n)$$

to converge in distribution, where μ_n is the empirical measure and $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$. Such conditions can be applied to any adapted sequence of random variables.

2000 *Mathematics Subject Classification.* 60B10, 60F05, 60G09, 60G57.

Key words and phrases. Conditional identity in distribution, empirical process, exchangeability, predictive measure, stable convergence.

2. Empirical processes

Throughout, $(\Omega, \mathcal{A}, P)$ is a probability space, $\mathcal{X}$ a Polish space and $\mathcal{B}$ the Borel σ-field on $\mathcal{X}$. The "data" are meant as a sequence $(X_n : n \geq 1)$ of $\mathcal{X}$-valued random variables on $(\Omega, \mathcal{A}, P)$. Also, we fix $\mathcal{F} \subset \mathcal{B}$ and we let $l^\infty(\mathcal{F})$ denote the space of real bounded functions on $\mathcal{F}$ equipped with the sup-norm

$$\|\phi\| = \sup_{B \in \mathcal{F}} |\phi(B)|, \quad \phi \in l^\infty(\mathcal{F}).$$

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) $\omega \mapsto \gamma(\omega)(B)$ is $\mathcal{A}$-measurable for all $B \in \mathcal{B}$. One example is the empirical measure

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

In the particular case where (X_n) is i.i.d., the empirical process is

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

where $\mu = P \circ X_1^{-1}$ denotes the probability distribution common to the X_n. Anyway, apart from (X_n) is i.i.d. or not, G_n is a map $G_n : \Omega \to l^\infty(\mathcal{F})$. 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 (X_n) is not i.i.d., G_n needs not be the "right" empirical process to be dealt with. A first reason is that, even if (X_n) is identically distributed, μ is only a part of the probability distribution of the sequence (X_n) (and usually not the most meaningful part). 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, if (X_n) is non 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 (X_n) is non ergodic, empirical processes should 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 $\|\mu_n - \gamma_n\| \xrightarrow{P} 0$.

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

$$P\big((X_1, X_2, \ldots) \in B\big) = \int \gamma(\omega)^\infty(B)\, P(d\omega), \quad B \in \mathcal{B}^\infty,$$

where γ is a (regular) version of $P\big(X_1 \in \cdot \mid \mathcal{T}\big)$ and $\gamma(\omega)^\infty = \gamma(\omega) \times \gamma(\omega) \times \ldots$. In this case, it is tempting to let $r_n = \sqrt{n}$ and $\gamma_n = \gamma$. The corresponding empirical process

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

is examined in [4] and [5].

For another example, suppose (X_n) is adapted to a filtration $(\mathcal{G}_n : n \geq 0)$ and define the predictive measure

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

In Bayesian inference and discrete time filtering, evaluating a_n is a major goal. 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. For instance, μ_n is a sound estimate of a_n if (X_n) is exchangeable and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. 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, for $\|\mu_n - a_n\| = \frac{1}{r_n}\|r_n(\mu_n - a_n)\| \xrightarrow{P} 0$. Thus, in a Bayesian framework, it is quite reasonable to let $\gamma_n = a_n$. Letting also $r_n = \sqrt{n}$ leads to the empirical process

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

In case of c.i.d. data (see Section 4), C_n is investigated in [1], [4], [6].

One more possible 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], [9] and Subsection 6.1. The corresponding empirical process

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

is concerned in [4] for c.i.d. data.

In Section 6, we focus on B_n and C_n in case of any (adapted) sequence (X_n) of random variables.

We finally note that B_n, C_n and W_n reduce to G_n in the particular case where (X_n) is i.i.d., $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. 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. Modes of 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 investigate their convergence in distribution, thus, we need a definition which allows for non measurable random elements. One such definition is due to Hoffmann-Jørgensen. The resulting theory, developed in [11] and [16], is nice and usable in real problems. We recall here basic definitions.

Let S be a metric space. A map $X : \Omega \to S$ is *measurable* (or a *random variable*) in case $X^{-1}(B) \in \mathcal{A}$ for all Borel sets $B \subset S$, and it is tight provided it is measurable and has a tight probability distribution. The outer expectation of a bounded function $Z : \Omega \to \mathbb{R}$ is

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

Let ν be a probability law on the Borel σ-field of S and $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ a net of probability spaces with associated maps $Z_\alpha : \Omega_\alpha \to S$. Denote $\nu(f) = \int f d\nu$ for all bounded Borel functions $f : S \to \mathbb{R}$. Say that Z_α *converges in distribution to* ν if

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

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

To make the previous definition more transparent, we recall that Z_α converges in distribution to ν if and only if

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

see [3]. Actually, convergence in distribution of Z_α amounts to weak convergence of the probability laws $Q_\alpha \circ Z_\alpha^{-1}$, 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 Z_α *converges stably to* γ in case

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

Stable convergence clearly implies convergence in distribution. Indeed, Z_α converges in distribution to the probability measure $E(\gamma(\cdot) \mid H)$, under $P(\cdot \mid H)$, for each $H \in \mathcal{A}$ with $P(H) > 0$. Stable convergence has been introduced by Renyi and subsequently investigated by various authors (in case the Z_α are measurable). We refer to [8], [14] and references therein for details.

4. Conditionally identically distributed random variables

4.1. **Basics.** In the sequel, $\mathcal{G} = (\mathcal{G}_n : n \geq 0)$ is a filtration on $(\Omega, \mathcal{A}, P)$. The sequence $(X_n : n \geq 1)$ is *conditionally identically distributed with respect to* $\mathcal{G}$, abbreviated $\mathcal{G}$-c.i.d., in case (X_n) is $\mathcal{G}$-adapted and

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

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

$$X_{T+1} \sim X_1 \quad \text{for each finite } \mathcal{G}\text{-stopping time } T.$$

When $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$, the filtration is not mentioned at all and (X_n) is just called c.i.d.. Clearly, if (X_n) is $\mathcal{G}$-c.i.d. then it is c.i.d. and identically distributed. Moreover, (X_n) is c.i.d. if and only if

$$\big(X_1, \ldots, X_n, X_{n+2}\big) \sim \big(X_1, \ldots, X_n, X_{n+1}\big) \quad \text{for all } n \geq 0. \tag{2}$$

Exchangeable sequences are c.i.d., for they meet (2), while the converse is not true. In fact, by a result in [15], (X_n) is exchangeable if and only if it is stationary and c.i.d.. Forthcoming Examples 3, 4 and 7 exhibit non exchangeable c.i.d. sequences. We refer to [4] for more on c.i.d. sequences. Here, we just mention the SLLN. Suppose $\mathcal{X} = \mathbb{R}$, (X_n) is c.i.d. and $E|X_1| < \infty$. Then, $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} X$ for a suitable random variable X.

4.2. **Characterizations.** Following [10], let us call *strategy* any collection

$$\sigma = \{\sigma(q) : q = \emptyset \text{ or } q \in \mathcal{X}^n \text{ for some } n = 1, 2, \ldots\}$$

where each $\sigma(q)$ is a probability on $\mathcal{B}$ and $(x_1, \ldots, x_n) \mapsto \sigma(x_1, \ldots, x_n)(B)$ is Borel measurable for all $n \geq 1$ and $B \in \mathcal{B}$. Here, $\emptyset$ denotes "the empty sequence". Let π_n be the n-th coordinate projection on $\mathcal{X}^\infty$, i.e.,

$$\pi_n(x_1, \ldots, x_n, \ldots) = x_n \quad \text{for all } n \geq 1 \text{ and } (x_1, \ldots, x_n, \ldots) \in \mathcal{X}^\infty.$$

By Ionescu Tulcea theorem, each strategy σ induces a unique probability ν on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$. By "$\sigma$ *induces* ν" we mean that, under ν,

$$\pi_1 \sim \sigma(\emptyset) \text{ and } \{\sigma(q) : q \in \mathcal{X}^n\} \text{ is a version of the conditional distribution of } \pi_{n+1} \text{ given } (\pi_1, \dots, \pi_n) \text{ for all } n \geq 1. \tag{3}$$

Conversely, since $\mathcal{X}$ is Polish, each probability ν on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ is induced by an (essentially unique) strategy σ.

Let α_0 and $\{\alpha(x) : x \in \mathcal{X}\}$ be probabilities on $\mathcal{B}$ such that the map $x \mapsto \alpha(x)(B)$ is Borel measurable for $B \in \mathcal{B}$. Say that $\{\alpha(x) : x \in \mathcal{X}\}$ is a *(Markov) kernel with stationary distribution* α_0 in case $\alpha_0(B) = \int \alpha(x)(B)\, \alpha_0(dx)$ for $B \in \mathcal{B}$.

In this notation, the following result is available.

Theorem 1. *Let ν be the probability distribution of the sequence (X_n). Then, (X_n) is c.i.d. if and only if ν is induced by a strategy σ satisfying*

(a) *the kernel $\{\sigma(q, x) : x \in \mathcal{X}\}$ has stationary distribution $\sigma(q)$ for $q = \emptyset$ and for almost all $q \in \mathcal{X}^n$, $n = 1, 2, \dots$.*

Proof. Fix a strategy σ which induces ν. By (2) and (3), (X_n) is c.i.d. if and only if $X_2 \sim X_1$ and, under ν,

$$\{\sigma(q) : q \in \mathcal{X}^n\} \text{ is a version of the conditional distribution of } \pi_{n+2} \text{ given } (\pi_1, \dots, \pi_n) \text{ for all } n \geq 1. \tag{4}$$

In view of (3), the condition $X_2 \sim X_1$ amounts to

$$\int \sigma(x)(B)\, \sigma(\emptyset)(dx) = P(X_2 \in B) = P(X_1 \in B) = \sigma(\emptyset)(B), \quad B \in \mathcal{B},$$

which just means that the kernel $\{\sigma(x) : x \in \mathcal{X}\}$ has stationary distribution $\sigma(\emptyset)$. Likewise, condition (4) is equivalent to

$$\text{for all } n \geq 1, \text{ there is } H_n \in \mathcal{B}^n \text{ such that } P\big((X_1, \dots, X_n) \in H_n\big) = 1$$
$$\text{and} \quad \int \sigma(q, x)(B)\, \sigma(q)(dx) = \sigma(q)(B) \quad \text{for all } q \in H_n \text{ and } B \in \mathcal{B}.$$

Therefore, (X_n) is c.i.d. if and only if σ can be taken to meet condition (a). □

Practically, Theorem 1 suggests how to assess a c.i.d. sequence (X_n) stepwise. First, select a law $\sigma(\emptyset)$ on $\mathcal{B}$, the marginal distribution of X_1. Then, choose a kernel $\{\sigma(x) : x \in \mathcal{X}\}$ with stationary distribution $\sigma(\emptyset)$, where $\sigma(x)$ should be viewed as the conditional distribution of X_2 given $X_1 = x$. Next, for each $x \in \mathcal{X}$, select a 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 X_3 given $X_1 = x$ and $X_2 = y$. And so on. In other terms, for getting a c.i.d. sequence, it is enough to assign at each step a kernel with a given stationary distribution. Indeed, a plenty of methods for doing this are available: see the statistical literature concerning MCMC.

Finally, we recall that exchangeable sequences admit an analogous characterization. Say that $\{\alpha(x) : x \in \mathcal{X}\}$ is a *reversible* kernel with respect to α_0 in case

$$\int_A \alpha(x)(B)\, \alpha_0(dx) = \int_B \alpha(x)(A)\, \alpha_0(dx) \quad \text{for all } A, B \in \mathcal{B}.$$

If a kernel is reversible with respect to a probability law, it admits such a law as a stationary distribution. The following result, firstly proved by de Finetti for $\mathcal{X} = \{0, 1\}$, is in [13].

Theorem 2. *The sequence (X_n) is exchangeable if and only if its probability distribution is induced by a strategy σ such that*

(b) $\{\sigma(q,x) : x \in \mathcal{X}\}$ *is a reversible kernel with respect to* $\sigma(q)$,

(c) $\sigma(\widetilde{q}) = \sigma(q)$ *whenever* $\widetilde{q}$ *is a permutation of* q,

for $q = \emptyset$ and for almost all $q \in \mathcal{X}^n$, $n = 1, 2, \ldots$. (with $\widetilde{q} = q$ if $q = \emptyset$).

4.3. **Examples.** It is not hard to see that condition (b) reduces to (a) whenever $\mathcal{X} = \{0,1\}$. Thus, for a sequence (X_n) of indicators, (X_n) is exchangeable if and only if it is c.i.d. and its conditional distributions $\sigma(q)$ are invariant under permutations of q. It is tempting to conjecture that (b) can be weakened into (a) in general, even if the X_n are not indicators. As shown by the next example, however, this is not true. It may be that (X_n) fails to be exchangeable, and yet it is c.i.d. and its conditional distributions meet condition (c).

Example 3. Let $\mathcal{X} = \mathcal{Y} \times (0,\infty)$, where $\mathcal{Y}$ is a Polish space. Fix a constant $r > 0$ and Borel probabilities μ_1 on $\mathcal{Y}$ and μ_2 on $(0,\infty)$. Define $\sigma(\emptyset) = \mu_1 \times \mu_2$ and

$$\begin{aligned}\sigma(x_1,\ldots,x_n)(A \times B) &= \sigma[(y_1,z_1),\ldots,(y_n,z_n)](A \times B) \\ &= \frac{r\,\mu_1(A) + \sum_{i=1}^n z_i\, I_A(y_i)}{r + \sum_{i=1}^n z_i}\ \mu_2(B)\end{aligned}$$

where $n \geq 1$, $x_i = (y_i, z_i) \in \mathcal{Y} \times (0,\infty)$ for all i and $A \subset \mathcal{Y}$, $B \subset (0,\infty)$ are Borel sets. By construction, σ satisfies condition (c). By Lemma 6 of [6], (π_n) is c.i.d. under ν, where ν is the probability on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ induced by σ. However, (π_1, π_2) is not distributed as (π_2, π_1) for various choices of μ_1, μ_2 (take for instance $\mathcal{Y} = \{0,1\}$, $\mu_1 = (\delta_0 + \delta_1)/2$ and $\mu_2 = (\delta_1 + \delta_2)/2$). Hence, (π_n) may fail to be exchangeable under ν.

The strategy σ of Example 3 makes sense in some real problems. In general, the z_n should be viewed as weights while the y_n describe the phenomenon of interest. As an example, consider an urn containing white and black balls. At each time $n \geq 1$, a ball is drawn and then replaced together with z_n more balls of the same color. Let y_n be the indicator of the event $\{\text{white ball at time } n\}$ and suppose z_n is chosen according to a fixed distribution on the integers, independently of $(y_1, z_1, \ldots, y_{n-1}, z_{n-1}, y_n)$. This situation is modelled by σ in Example 3. Note also that σ is of Ferguson-Dirichlet type in case $z_n = 1$ for all n.

Finally, suppose (X_n) is 2-exchangeable, that is,

$$(X_i, X_j) \sim (X_1, X_2) \quad \text{for all } i \neq j.$$

Suggested by de Finetti's representation theorem, another conjecture is that the probability distribution of (X_n) is a mixture of 2-independent identically distributed laws. More precisely, this means that

$$P\big((X_1, X_2, \ldots) \in B\big) = \int \nu(B)\, Q(d\nu), \quad B \in \mathcal{B}^\infty, \tag{5}$$

where Q is some mixing measure supported by those probability laws ν on $(\mathcal{X}^\infty, \mathcal{B}^\infty)$ such that (π_n) is 2-independent and identically distributed under ν. Once again, the conjecture turns out to be false. As shown by the following example, it may be that (X_n) is c.i.d. and 2-exchangeable and yet its probability distribution does not admit representation (5).

Example 4. Let m be Lebesgue measure and $f : [0,1] \to [0,1]$ a Borel function satisfying

$$\int_0^1 f(u)\,du = \frac{1}{2}, \quad \int_0^1 u\,f(u)\,du = \frac{1}{3}, \quad m\{u \in [0,1] : f(u) \neq u\} > 0. \tag{6}$$

Let $(U_n : n \geq 0)$ be i.i.d. with U_0 uniformly distributed on $[0,1]$. Define $\mathcal{X} = \{0,1\}$ and $X_n = I_{H_n}$, where

$$H_1 = \{U_1 \leq f(U_0)\}, \quad H_n = \{U_n \leq U_0\} \quad \text{for } n > 1.$$

Conditionally on U_0, the sequence (X_n) is independent with

$$P(X_1 = 1 \mid U_0) = f(U_0) \quad \text{and} \quad P(X_n = 1 \mid U_0) = U_0 \quad \text{a.s. for all } n > 1.$$

Basing on this fact and (6), it is straightforward to check that (X_n) is c.i.d. and 2-exchangeable. Moreover, $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} U_0$. By Etemadi's SLLN, if (π_n) is 2-independent and identically distributed under ν, then

$$\frac{1}{n}\sum_{i=1}^n \pi_i \xrightarrow{\nu-a.s.} E_\nu(\pi_1).$$

Letting $\pi_* = \limsup_n \frac{1}{n}\sum_{i=1}^n \pi_i$, it follows that

$$\nu\big(\pi_* \in I,\, \pi_1 = 1\big) = \nu\big(\pi_* \in I,\, \pi_2 = 1\big) \quad \text{for all Borel sets } I \subset [0,1].$$

Hence, if representation (5) holds, one obtains

$$\begin{aligned}\int_I f(u)\,du &= \int_{\{U_0 \in I\}} P(X_1 = 1 \mid U_0)\,dP = P(U_0 \in I,\, X_1 = 1)\\ &= \int \nu\big(\pi_* \in I,\, \pi_1 = 1\big)\,Q(d\nu) = \int \nu\big(\pi_* \in I,\, \pi_2 = 1\big)\,Q(d\nu)\\ &= P(U_0 \in I,\, X_2 = 1) = \int_I u\,du \quad \text{for all Borel sets } I \subset [0,1].\end{aligned}$$

This implies $f(u) = u$, for m-almost all u, contrary to (6). Thus, the probability distribution of (X_n) can not be written as in (5).

5. Back to empirical processes

In this section, the empirical process theory for c.i.d. data is summarized. With the only exception of Example 7, which is new, all other results are from [4].

In the sequel, we focus on

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

Accordingly, for each $\phi \in l^\infty(\mathcal{F})$, we write $\phi(t)$ instead of $\phi\big((-\infty,t]\big)$ and we regard ϕ as a member of $l^\infty(\mathbb{R})$. Moreover, $N_k(0,\Sigma)$ denotes the Gaussian law on the Borel sets of $\mathbb{R}^k$ with mean 0 and covariance matrix Σ (possibly singular). We let $N_k(0,0) = \delta_0$ and, for $k = 1$ and $u \geq 0$, we write $N(0,u)$ instead of $N_1(0,u)$.

Suppose (X_n) is $\mathcal{G}$-c.i.d. and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure. By the SLLN, there is a random probability measure γ on $\mathbb{R}$ such that

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

Let $F_\gamma(t) = \gamma(-\infty, t]$, $t \in \mathbb{R}$. As in Section 2, define also $a_n(\cdot) = P\big(X_{n+1} \in \cdot \mid \mathcal{G}_n\big)$, $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$ and

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

As usual, B_n and C_n are regarded as maps from Ω into $l^\infty(\mathbb{R})$.

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

$$P_0\big((\mathbb{G}(t_1), \ldots, \mathbb{G}(t_k)) \in A\big) = \int N_k\big(0, \Sigma(t_1, \ldots, t_k)\big)(A)\, dP \tag{7}$$

where $t_1, \ldots, t_k \in \mathbb{R}$, $A \subset \mathbb{R}^k$ is a Borel set and $\Sigma(t_1, \ldots, t_k)$ a random covariance matrix on $(\Omega, \mathcal{A}, P)$. One significant case is

$$\mathbb{G}^F(t) = \mathbb{B}(F(t)), \quad t \in \mathbb{R},$$

where $\mathbb{B}$ and F are defined on $(\Omega_0, \mathcal{A}_0, P_0)$, $\mathbb{B}$ is a Brownian bridge, F a random distribution function independent of $\mathbb{B}$, and $F \sim F_\gamma$. Then, (7) holds with $\mathbb{G} = \mathbb{G}^F$ and

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

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; see Example 11 of [3]. However, $\mathbb{G}^F$ is measurable and tight whenever every F-path is continuous on D^c for some *fixed* countable set $D \subset \mathbb{R}$.

As a trivial example, suppose (X_n) i.i.d., $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Then $F = H$ a.s., where H is the distribution function common to the X_n, and D can be taken as $D = \{t : H(t) > H(t-)\}$. 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 $(\Omega, \mathcal{A}, P)$. 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 $\mathbb{R}$ by right-open intervals such that

$$\limsup_n P\big(\max_j \sup_{s,t \in I_j} |Z_n(s) - Z_n(t)| > \epsilon\big) < \eta. \tag{8}$$

We are now able to state a couple of results.

Theorem 5. *If (X_n) is $\mathcal{G}$-c.i.d. and B_n meets (8) (i.e., (8) holds with $Z_n = B_n$), then $B_n \xrightarrow{d} \mathbb{G}^F$ and $\mathbb{G}^F$ is tight.*

Theorem 6. *Suppose (X_n) is $\mathcal{G}$-c.i.d., C_n meets (8), and $\sup_n E\{C_n(t)^2\} < \infty$ for all $t \in \mathbb{R}$. If*

$$\frac{1}{n}\sum_{i=1}^n q_i(s)\, q_i(t) \xrightarrow{a.s.} \sigma(s,t) \quad \text{for all } s, t \in \mathbb{R}$$

$$\text{where} \quad q_i(t) = I_{\{X_i \le t\}} - i\, P\big(X_{i+1} \le t \mid \mathcal{G}_i\big) + (i-1)\, P\big(X_i \le t \mid \mathcal{G}_{i-1}\big),$$

then $C_n \xrightarrow{d} \mathbb{G}$, where $\mathbb{G}$ is a tight process with distribution (7) and

$$\Sigma(t_1, \ldots, t_k) = \Big(\sigma(t_i, t_j) : 1 \le i, j \le k\Big).$$

Both Theorems 5 and 6 require condition (8). Thus, it would be useful to have a criterion for testing it. In the exchangeable case, one such criterion is tightness of the process $\mathbb{G}^F$. Suppose in fact (X_n) is exchangeable. Then, condition (8) holds for B_n and C_n provided $\mathbb{G}^F$ is tight. In particular, $B_n \xrightarrow{d} \mathbb{G}^F$ if $\mathbb{G}^F$ is tight and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$ for all n. Furthermore, $\mathbb{G}^F$ is tight whenever $P(X_1 = X_2) = 0$ or X_1 has a discrete distribution. Unfortunately, this useful criterion can fail in the

$\mathcal{G}$-c.i.d. case. We close this section with an example where (X_n) is $\mathcal{G}$-c.i.d., $\mathbb{G}^F$ is tight, and yet condition (8) fails for C_n.

Example 7. 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

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

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

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

Hence, $\gamma = \delta_X$ and $\mathbb{G}^F = 0$, so that $\mathbb{G}^F$ is tight.

The finite dimensional distributions of C_n converge weakly to 0. In fact,

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

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

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

Since both $X_n \xrightarrow{a.s.} X$ and $S_n \xrightarrow{a.s.} X$, it is not hard to see that $C_n(t) \xrightarrow{a.s.} 0$ for every fixed t.

Toward a contradiction, suppose now that C_n meets (8) and define

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

Then $C_n \xrightarrow{d} 0$, so that $|I_n| \le 2\,\|C_n\| \xrightarrow{P} 0$ (recall that $\|\cdot\|$ denotes the sup-norm). On the other hand,

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

Let

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

Then $I_n - J_n \xrightarrow{a.s.} 0$, due to $S_n - X_n \xrightarrow{a.s.} 0$, 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,

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

Therefore, condition (8) fails for C_n.

Incidentally, neither $W_n = \sqrt{n}\,(\mu_n - \gamma)$ satisfies condition (8). In fact,

$$W_n(t) = \sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le t\}} - I_{\{X \le t\}}\Big) \xrightarrow{a.s.} 0 \quad \text{for fixed } t.$$

If W_n meets (8), thus, $W_n \xrightarrow{d} 0$ so that $\sup_t |W_n(t) - W_n(t-)| \le 2\,\|W_n\| \xrightarrow{P} 0$. But this is again a contradiction, for $P(X_i \neq X \text{ for all } i) = 1$ and

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

6. Uniform CLTs for the empirical processes B_n and C_n

If (X_n) is $\mathcal{G}$-c.i.d., conditions for B_n and C_n to converge in distribution are given by Theorems 5 and 6. In this section, the latter results are extended to any $\mathcal{G}$-adapted sequence (X_n). We again let $\mathcal{X} = \mathbb{R}$.

6.1. **Heuristics.** Motivations for dealing with C_n have been given in Section 2; see also [6] and [7]. Following [9], we now give analogous motivations for B_n. We assume that (X_n) is $\mathcal{G}$-adapted and $E|X_n| < \infty$ for all n.

Suppose that, at each time $n \ge 0$, you are requested to predict the next observation X_{n+1} basing on the available information $\mathcal{G}_n$. Your predictor is $E\big(X_{n+1} \mid \mathcal{G}_n\big)$ and prediction performances are assessed through

$$V_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i \mid \mathcal{G}_{i-1}) = \frac{\sum_{i=1}^{n}\big\{X_i - E(X_i \mid \mathcal{G}_{i-1})\big\}}{n}.$$

Loosely speaking, you are *well calibrated* when V_n is small.

By standard martingale arguments, $V_n \xrightarrow{a.s.} 0$ under quite general conditions, for instance when $\sup_n EX_n^2 < \infty$. In this case, it is useful to know the rate of convergence, and this leads to investigate the asymptotic behavior of $\sqrt{n}\,V_n$.

Suppose next you are interested in the events $\{X_n \le t\}$ and you aim to be well calibrated at a random value T of t. Define

$$F_n(t) = \mu_n(-\infty, t] = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le t\}} \quad \text{and}$$

$$F_n^*(t) = b_n(-\infty, t] = \frac{1}{n}\sum_{i=1}^{n} P\big(X_i \le t \mid \mathcal{G}_{i-1}\big), \quad t \in \mathbb{R}.$$

Then, you want $F_n(T) - F_n^*(T)$ small for some random variable T. On the other hand, $|F_n(T) - F_n^*(T)| \le \|F_n - F_n^*\|$ and

$$\|F_n - F_n^*\| \xrightarrow{a.s.} 0$$

whenever the empirical distribution function F_n converges uniformly on a set of probability 1; see [2]. Again, the rate of convergence of $\|F_n - F_n^*\|$ should be investigated, and this leads to the process

$$B_n(t) = \sqrt{n}\,\big\{F_n(t) - F_n^*(t)\big\}, \quad t \in \mathbb{R}.$$

One reason for dealing with B_n, thus, is calibration. Other reasons can be found in gambling and stochastic approximation; see [2] and references therein.

6.2. **Results.** Our main tools are the following two (non uniform) CLTs. The first is already known (see Theorem 2 of [7]) while the second is new.

Theorem 8. *Suppose (X_n) is $\mathcal{G}$-adapted and (X_n^2) uniformly integrable. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $Z_n = E(X_{n+1} \mid \mathcal{G}_n)$. Then,*

$$\sqrt{n}\,\big\{\overline{X}_n - Z_n\big\} \longrightarrow N(0, L) \quad \textit{stably}$$

provided

$$n^3\, E\Big\{\big(E(Z_{n+1} \mid \mathcal{G}_n) - Z_n\big)^2\Big\} \longrightarrow 0,$$

$$\frac{1}{\sqrt{n}}\, E\Big\{\max_{1\le i\le n} i\,|Z_{i-1} - Z_i|\Big\} \longrightarrow 0,$$

$$\frac{1}{n}\sum_{i=1}^n \big\{X_i - Z_{i-1} + i\,(Z_{i-1} - Z_i)\big\}^2 \xrightarrow{P} L.$$

Theorem 9. *Suppose (X_n) is $\mathcal{G}$-adapted and (X_n^2) uniformly integrable. Then,*

$$\sqrt{n}\,V_n = \frac{\sum_{i=1}^n \big\{X_i - E(X_i \mid \mathcal{G}_{i-1})\big\}}{\sqrt{n}} \longrightarrow N(0, L) \quad \textit{stably}$$

whenever

$$\frac{1}{n}\sum_{i=1}^n \big\{X_i - E(X_i \mid \mathcal{G}_{i-1})\big\}^2 \xrightarrow{P} L. \tag{9}$$

Moreover, condition (9) *applies if*

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} Y \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^n E(X_i \mid \mathcal{G}_{i-1})^2 \xrightarrow{P} Y - L$$

for some random variable Y, or if

$$E(X_n^2 \mid \mathcal{G}_{n-1}) - E(X_n \mid \mathcal{G}_{n-1})^2 \xrightarrow{P} L.$$

Proof. For $n \ge 1$ and $i = 1, \ldots, n$, define $\mathcal{F}_{n,0} = \mathcal{G}_0$, $\mathcal{F}_{n,i} = \mathcal{G}_i$ and

$$Y_{n,i} = n^{-1/2}\big\{X_i - E(X_i \mid \mathcal{G}_{i-1})\big\}.$$

Then, $\sqrt{n}\,V_n = \sum_{i=1}^n Y_{n,i}$. Further, $Y_{n,i}$ is $\mathcal{F}_{n,i}$-measurable, $\mathcal{F}_{n+1,i} = \mathcal{F}_{n,i}$, and

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

So, by the martingale CLT (see Theorem 3.2, p. 58, of [14]), it suffices proving that

$$\sum_{i=1}^n Y_{n,i}^2 \xrightarrow{P} L, \quad \max_{1\le i\le n} |Y_{n,i}| \xrightarrow{P} 0, \quad \sup_n E\big(\max_{1\le i\le n} Y_{n,i}^2\big) < \infty.$$

Let $D_i = X_i - E(X_i \mid \mathcal{G}_{i-1})$. By (9), $\sum_{i=1}^n Y_{n,i}^2 = \frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{P} L$. Since (X_n^2) is uniformly integrable, (D_n^2) is uniformly integrable as well. Given $\epsilon > 0$, take $a > 0$ such that $E\big(D_i^2 I_{\{|D_i|>a\}}\big) < \epsilon$ for all i. Then,

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

Therefore, $\lim_n E\big(\max_{1\le i\le n} Y_{n,i}^2\big) = 0$, and this implies that $\max_{1\le i\le n}|Y_{n,i}| \xrightarrow{P} 0$ and $\sup_n E\big(\max_{1\le i\le n} Y_{n,i}^2\big) < \infty$.

This concludes the proof of the first part. We next prove the sufficient conditions for (9). Define $\Delta_i = E(X_i^2 \mid \mathcal{G}_{i-1}) - E(X_i \mid \mathcal{G}_{i-1})^2$ and note that

$$E\Big|\sum_{i=1}^n (D_i^2 - \Delta_i)\Big| \le E\Big|\sum_{i=1}^n \big(X_i^2 - E(X_i^2 \mid \mathcal{G}_{i-1})\big)\Big| + 2\,E\Big|\sum_{i=1}^n D_i\,E(X_i \mid \mathcal{G}_{i-1})\Big|.$$

Since (X_n^2) is uniformly integrable, given $\epsilon > 0$, there is $a > 0$ such that

$$\sup_i E\big\{X_i^2\,(1 - I_{A_i})\big\} < \epsilon \quad \text{where } A_i = \{|X_i| \le a\}.$$

Further,

$$\Big\{E\Big|\sum_{i=1}^n \big(X_i^2 I_{A_i} - E(X_i^2 I_{A_i} \mid \mathcal{G}_{i-1})\big)\Big|\Big\}^2 \le E\Big\{\Big(\sum_{i=1}^n \big(X_i^2 I_{A_i} - E(X_i^2 I_{A_i} \mid \mathcal{G}_{i-1})\big)\Big)^2\Big\}$$
$$= \sum_{i=1}^n E\big\{\big(X_i^2 I_{A_i} - E(X_i^2 I_{A_i} \mid \mathcal{G}_{i-1})\big)^2\big\} \le n\,a^4.$$

Thus,

$$\frac{1}{n}\,E\Big|\sum_{i=1}^n \big(X_i^2 - E(X_i^2 \mid \mathcal{G}_{i-1})\big)\Big| \le \frac{a^2}{\sqrt{n}} + 2\sup_i E\big\{X_i^2\,(1 - I_{A_i})\big\} < \frac{a^2}{\sqrt{n}} + 2\,\epsilon.$$

Similarly, letting $d = \sqrt{\sup_i ED_i^2}$, one obtains

$$\frac{2}{n}\,E\Big|\sum_{i=1}^n D_i\,E(X_i \mid \mathcal{G}_{i-1})\Big| \le \frac{2\,a\,d}{\sqrt{n}} + 2\sup_i E\big\{|D_i|\,E\big(|X_i|\,(1 - I_{A_i}) \mid \mathcal{G}_{i-1}\big)\big\}$$
$$\le \frac{2\,a\,d}{\sqrt{n}} + 2\sup_i \sqrt{ED_i^2\,E\big\{X_i^2\,(1 - I_{A_i})\big\}} < \frac{2\,a\,d}{\sqrt{n}} + 2\,d\,\sqrt{\epsilon}.$$

It follows that

$$E\Big|\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{i=1}^n E(X_i^2 \mid \mathcal{G}_{i-1})\Big| \longrightarrow 0, \tag{10}$$

$$E\Big|\frac{1}{n}\sum_{i=1}^n D_i^2 - \frac{1}{n}\sum_{i=1}^n \Delta_i\Big| \longrightarrow 0. \tag{11}$$

Suppose that $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} Y$ and $\frac{1}{n}\sum_{i=1}^n E(X_i \mid \mathcal{G}_{i-1})^2 \xrightarrow{P} Y - L$. Then, $\frac{1}{n}\sum_{i=1}^n E(X_i^2 \mid \mathcal{G}_{i-1}) \xrightarrow{P} Y$ by (10), so that $\frac{1}{n}\sum_{i=1}^n \Delta_i \xrightarrow{P} L$. Thus, (11) implies $\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{P} L$, i.e., condition (9) holds.

Finally, suppose $\Delta_n \xrightarrow{P} L$. Then $E|\Delta_n - L| \longrightarrow 0$, due to (Δ_n) is uniformly integrable, so that $E|\frac{1}{n}\sum_{i=1}^n \Delta_i - L| \longrightarrow 0$. Again, (9) follows from (11). □

In order to apply Theorem 9, note that $\frac{1}{n}\sum_{i=1}^n X_i^2$ converges a.s. under various assumptions. This happens, for instance, if $EX_1^2 < \infty$ and (X_n) is $\mathcal{G}$-c.i.d. or stationary or 2-exchangeable. (In the 2-exchangeable case, just apply the SLLN in [12]). In turn, $\frac{1}{n}\sum_{i=1}^n E\big(X_i \mid \mathcal{G}_{i-1}\big)^2$ converges a.s. provided $E\big(X_n \mid \mathcal{G}_{n-1}\big)$ converges a.s., which is true at least in the $\mathcal{G}$-c.i.d. case. We do not know of any example where (X_n) is stationary, $EX_1^2 < \infty$, and yet $\frac{1}{n}\sum_{i=1}^n E\big(X_i \mid \mathcal{G}_{i-1}\big)^2$ fails to converge in probability. But such an example possibly exists.

We finally turn to uniform CLTs.

Theorem 10. *Suppose (X_n) is $\mathcal{G}$-adapted, B_n meets condition (8), and*

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

for s, $t\in\mathbb{R}$. Then $B_n \xrightarrow{d} \mathbb{G}$, where $\mathbb{G}$ is a tight process with distribution (7) and

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

Proof. By (8), it suffices to prove convergence of finite dimensional distributions; see e.g. Theorem 1.5.4 of [16]. Fix $t_1,\dots,t_k,u_1,\dots,u_k\in\mathbb{R}$ and define

$$L=\sum_{r=1}^{k}\sum_{j=1}^{k} u_r\,u_j\,\big(a(t_r\wedge t_j)-b(t_r,t_j)\big).$$

Define also $f=\sum_{r=1}^{k} u_r\,I_{(-\infty,t_r]}$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i)^2=\sum_{r=1}^{k}\sum_{j=1}^{k} u_r\,u_j\,\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i\le t_r\wedge t_j\}} \xrightarrow{P} \sum_{r=1}^{k}\sum_{j=1}^{k} u_r\,u_j\,a(t_r\wedge t_j).$$

Moreover,

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

$$=\sum_{r=1}^{k}\sum_{j=1}^{k} u_r\,u_j\,\frac{1}{n}\sum_{i=1}^{n} P\big(X_i\le t_r\mid\mathcal{G}_{i-1}\big)\,P\big(X_i\le t_j\mid\mathcal{G}_{i-1}\big) \xrightarrow{P} \sum_{r=1}^{k}\sum_{j=1}^{k} u_r\,u_j\,b(t_r,t_j).$$

Thus, Theorem 9 applies to $(f(X_n))$, so that

$$\sum_{r=1}^{k} u_r B_n(t_r)=\sqrt{n}\Big\{\frac{1}{n}\sum_{i=1}^{n} f(X_i)-\frac{1}{n}\sum_{i=1}^{n} E\big(f(X_i)\mid\mathcal{G}_{i-1}\big)\Big\} \longrightarrow N(0,L) \ \text{ stably.}$$

In particular, $\sum_{r=1}^{k} u_r B_n(t_r)$ converges in distribution to the probability measure

$$\nu(B)=\int N(0,L)(B)\,dP \quad \text{for all Borel sets } B\subset\mathbb{R}.$$

On noting that $\sum_{r=1}^{k} u_r\,\mathbb{G}(t_r)\sim\nu$, one obtains $\sum_{r=1}^{k} u_r\,B_n(t_r)\xrightarrow{d}\sum_{r=1}^{k} u_r\,\mathbb{G}(t_r)$. By letting $u_1,\dots,u_k$ vary, it follows that

$$\big(B_n(t_1),\dots,B_n(t_k)\big)\xrightarrow{d}\big(\mathbb{G}(t_1),\dots,\mathbb{G}(t_k)\big).$$

□

For the last result, as in Theorem 6, we let

$$q_i(t)=I_{\{X_i\le t\}}-i\,P\big(X_{i+1}\le t\mid\mathcal{G}_i\big)+(i-1)\,P\big(X_i\le t\mid\mathcal{G}_{i-1}\big).$$

Theorem 11. *Suppose (X_n) is $\mathcal{G}$-adapted, C_n meets condition (8), and*

$$n^3\,E\Big\{\big(P(X_{n+2}\le t\mid\mathcal{G}_n)-P(X_{n+1}\le t\mid\mathcal{G}_n)\big)^2\Big\}\longrightarrow 0,$$

$$\frac{1}{\sqrt{n}}\,E\Big\{\max_{1\le i\le n}|q_i(t)|\Big\}\longrightarrow 0, \quad \frac{1}{n}\sum_{i=1}^{n} q_i(s)\,q_i(t)\xrightarrow{P}\sigma(s,t),$$

for s, $t \in \mathbb{R}$. Then $C_n \xrightarrow{d} \mathbb{G}$, where $\mathbb{G}$ is a tight process with distribution (7) *and*

$$\Sigma(t_1, \ldots, t_k) = \Big(\sigma(t_i, t_j) : 1 \le i, j \le k\Big).$$

Proof. We just give a sketch of the proof, for it is quite analogous to that of Theorem 10. By (8), it is enough to prove finite dimensional convergence. Fix $t_1, \ldots, t_k, u_1, \ldots, u_k \in \mathbb{R}$ and define $L = \sum_{r=1}^k \sum_{j=1}^k u_r\, u_j\, \sigma(t_r, t_j)$ and $\nu(\cdot) = \int N(0, L)(\cdot)\, dP$. Since $\sum_{r=1}^k u_r\, \mathbb{G}(t_r) \sim \nu$, it suffices to show that

$$\sum_{r=1}^k u_r\, C_n(t_r) \longrightarrow N(0, L) \quad \text{stably.}$$

To this end, we let $f = \sum_{r=1}^k u_r\, I_{(-\infty, t_r]}$ and we apply Theorem 8 to $(f(X_n))$. Define $U_n = E\big(f(X_{n+1}) \mid \mathcal{G}_n\big) = \sum_{r=1}^k u_r\, P\big(X_{n+1} \le t_r \mid \mathcal{G}_n\big)$. On noting that

$$E\big(U_{n+1} \mid \mathcal{G}_n\big) = E\big(f(X_{n+2}) \mid \mathcal{G}_n\big) = \sum_{r=1}^k u_r\, P\big(X_{n+2} \le t_r \mid \mathcal{G}_n\big),$$

one obtains

$$n^3\, E\Big\{\big(E(U_{n+1} \mid \mathcal{G}_n) - U_n\big)^2\Big\} \le n^3\, k^2 \max_{1 \le r \le k} u_r^2\, E\Big\{\big(P(X_{n+2} \le t_r \mid \mathcal{G}_n) - P(X_{n+1} \le t_r \mid \mathcal{G}_n)\big)^2\Big\} \longrightarrow 0,$$

$$\frac{1}{\sqrt{n}}\, E\Big\{\max_{1 \le i \le n} i\, |U_{i-1} - U_i|\Big\} \le \sum_{r=1}^k |u_r|\, \frac{E\big\{\max_{1 \le i \le n} |q_i(t_r)|\big\} + 1}{\sqrt{n}} \longrightarrow 0,$$

$$\frac{1}{n} \sum_{i=1}^n \big\{f(X_i) - U_{i-1} + i\,(U_{i-1} - U_i)\big\}^2 = \sum_{r=1}^k \sum_{j=1}^k u_r\, u_j\, \frac{1}{n} \sum_{i=1}^n q_i(t_r)\, q_i(t_j) \xrightarrow{P} L.$$

Hence, $\sum_{r=1}^k u_r\, C_n(t_r) = \sqrt{n}\, \big\{\frac{1}{n} \sum_{i=1}^n f(X_i) - U_n\big\} \longrightarrow N(0, L)$ stably. □

References

[1] Bassetti F., Crimaldi I., Leisen F. (2010) Conditionally identically distributed species sampling sequences, *Adv. in Appl. Probab.*, 42, 433-459.

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

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

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

[5] Berti P., Pratelli L., Rigo P. (2006) Asymptotic behaviour of the empirical process for exchangeable data, *Stoch. Proc. Appl.*, 116, 337-344.

[6] Berti P., Crimaldi I., Pratelli L., Rigo P. (2009) Rate of convergence of predictive distributions for dependent data, *Bernoulli*, 15, 1351-1367.

[7] Berti P., Crimaldi I., Pratelli L., Rigo P. (2010) A central limit theorem and its applications to multicolor randomly reinforced urns, *Submitted*, currently available at: `http://economia.unipv.it/pagp/pagine_personali/prigo/urn.pdf`

[8] Crimaldi I., Letta G., Pratelli L. (2007) A strong form of stable convergence, *Seminaire de Probabilites* XL, Lect. Notes in Math., 1899, 203-225.

[9] Dawid P. (1982) The well-calibrated Bayesian (with discussion), *J. Amer. Statist. Assoc.*, 77, 605-613.
[10] Dubins L.E., Savage L.J. (1965) *How to gamble if you must: Inequalities for stochastic processes*, McGraw Hill.
[11] Dudley R.M. (1999) *Uniform central limit theorems*, Cambridge University Press.
[12] Etemadi N., Kaminski M. (1996) Strong law of large numbers for 2-exchangeable random variables, *Statist. Probab. Letters*, 28, 245-250.
[13] Fortini S., Ladelli L., Regazzini E. (2000) Exchangeability, predictive distributions and parametric models, *Sankhya* A, 62, 86-109.
[14] Hall P., Heyde C.C. (1980) *Martingale limit theory and its applications*, Academic Press.
[15] Kallenberg O. (1988) Spreading and predictable sampling in exchangeable sequences and processes, *Ann. Probab.*, 16, 508-534.
[16] van der Vaart A., Wellner J.A. (1996) *Weak convergence and empirical processes*, Springer.

Patrizia Berti, Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy
E-mail address: patrizia.berti@unimore.it

Luca Pratelli, Accademia Navale, viale Italia 72, 57100 Livorno, Italy
E-mail address: pratel@mail.dm.unipi.it

Pietro Rigo (corresponding author), Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia, via S. Felice 5, 27100 Pavia, Italy
E-mail address: prigo@eco.unipv.it