Electronic Journal of Probability

Electron. J. Probab. **17** (2012), no. 9, 1–18.
ISSN: 1083-6489 DOI: 10.1214/EJP.v17-1765

Limit theorems for empirical processes based on dependent data

Patrizia Berti[*] Luca Pratelli[†] Pietro Rigo[‡]

Abstract

Let (X_n) be any sequence of random variables adapted to a filtration $(\mathcal{G}_n)$. Define $a_n(\cdot) = P\big(X_{n+1} \in \cdot \mid \mathcal{G}_n\big)$, $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$, and $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. Convergence in distribution of the empirical processes

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

is investigated under uniform distance. If (X_n) is conditionally identically distributed, convergence of B_n and C_n is studied according to Meyer-Zheng as well. Some CLTs, both uniform and non uniform, are proved. In addition, various examples and a characterization of conditionally identically distributed sequences are given.

Keywords: Conditional identity in distribution; Empirical process; Exchangeability; Predictive measure; Stable convergence.
AMS MSC 2010: 60B10; 60F05; 60G09; 60G57.
Submitted to EJP on November 17, 2011, final version accepted on January 21, 2012.

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. Special attention is paid to *conditionally identically distributed* (c.i.d.)

[*]University of Modena e Reggio-Emilia, Italy. E-mail: patrizia.berti@unimore.it
[†]Accademia Navale di Livorno, Italy. E-mail: pratel@mail.dm.unipi.it
[‡]University of Pavia, Italy. E-mail: pietro.rigo@unipv.it

sequences of random variables (see Section 3). This type of dependence, introduced in [4] and [16], includes exchangeability as a particular case and plays a role in Bayesian inference.

For convergence in distribution of empirical processes, the usual and (natural) distances are the uniform and the Skorohod ones. While such distances have various merits, they are often too strong to deal with non ergodic data. Thus, in case of c.i.d. data, empirical processes are also investigated under a weaker distance; see Meyer-Zheng's paper [19] and Subsection 5.2.

The paper is organized as follows. Sections 2 and 4 include preliminary material (with the only exception of Example 4.4). Results are in Sections 3 and 5. Section 3 includes a characterization of c.i.d. sequences and a couple of examples. Section 5 contains some uniform and non uniform CLTs. Suppose (X_n) is adapted to a filtration $(\mathcal{G}_n)$. Define the *predictive measure* $a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{G}_n)$, the empirical measure $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and the empirical processes

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

where $b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$. Our main results provide conditions for B_n and C_n to converge in distribution, under uniform distance as well as in Meyer-Zheng's sense; see Theorems 5.2-5.6.

2 Notation and basic definitions

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 $(\Omega, \mathcal{A}, P)$. The sequence (X_n) is adapted to the filtration $\mathcal{G} = (\mathcal{G}_n : n \geq 0)$. Apart from the final Section 5, (X_n) is assumed to be identically distributed.

Let S be a metric space. A random probability measure on S is a map γ on Ω such that: (i) $\gamma(\omega)$ is a Borel probability measure on S for each $\omega \in \Omega$; (ii) $\omega \mapsto \gamma(\omega)(B)$ is $\mathcal{A}$-measurable for each Borel set $B \subset S$. If $S = \mathbb{R}$, we call F a random distribution function if $F(t,\omega) = \gamma(\omega)(-\infty, t]$, $(t,\omega) \in \mathbb{R} \times \Omega$, for some random probability measure γ on $\mathbb{R}$.

A map $Y : \Omega \to S$ is *measurable*, or a *random variable*, if $Y^{-1}(B) \in \mathcal{A}$ 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 $V : \Omega \to \mathbb{R}$ is $E^*V = \inf EU$, where $\inf$ ranges over those bounded measurable $U : \Omega \to \mathbb{R}$ satisfying $U \geq V$.

Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a sequence of probability spaces and $Y_n : \Omega_n \to S$. The maps Y_n are not requested to be measurable. Denote by $C_b(S)$ the set of real bounded continuous functions on S. Given a Borel probability ν on S, say that Y_n *converges in distribution to* ν if

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

In such case, we write $Y_n \xrightarrow{d} Y$ for any S-valued random variable Y, defined on some probability space, with distribution ν. We refer to [11] and [21] for more on convergence in distribution of non measurable random elements.

Finally, we turn to stable convergence. Fix a random probability measure γ on S and suppose $(\Omega_n, \mathcal{A}_n, P_n) = (\Omega, \mathcal{A}, P)$ for all n. Say that Y_n *converges stably to* γ whenever

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

Here, $\gamma(f)$ denotes the real random variable $\omega \mapsto \int f(x)\,\gamma(\omega)(dx)$. Stable convergence clearly implies convergence in distribution. Indeed, Y_n 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 Y_n are measurable). We refer to [8], [15] and references therein for details.

3 Conditionally identically distributed random variables

3.1 Basics

Let $(X_n : n \geq 1)$ be adapted to the filtration $\mathcal{G} = (\mathcal{G}_n : n \geq 0)$. Then, (X_n) is *conditionally identically distributed* with respect to $\mathcal{G}$, or $\mathcal{G}$-c.i.d., if

$$P\big(X_k \in \cdot \mid \mathcal{G}_n\big) = P\big(X_{n+1} \in \cdot \mid \mathcal{G}_n\big) \quad \text{a.s. for all } k > n \geq 0. \tag{3.1}$$

Roughly speaking, (3.1) 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 (3.1) is equivalent to

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

(For any random variables U and V, we write $U \sim V$ to mean that U and V are identically distributed). 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

$$\big(X_1, \ldots, X_n, X_{n+2}\big) \sim \big(X_1, \ldots, X_n, X_{n+1}\big) \quad \text{for all } n \geq 0. \tag{3.2}$$

Exchangeable sequences are c.i.d., for they meet (3.2), while the converse is not true. In fact, by a result in [16], (X_n) is exchangeable if and only if it is stationary and c.i.d.. Forthcoming Examples 3.3, 3.4 and 4.4 exhibit non exchangeable c.i.d. sequences. We refer to [4] for more on c.i.d. sequences. Here, it suffices to mention the Strong Law of Large Numbers (SLLN) and some of its consequences.

Let (X_n) be $\mathcal{G}$-c.i.d. and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical measure. Then, there is a random probability measure γ on $\mathcal{X}$ satisfying

$$\mu_n(B) \xrightarrow{a.s.} \gamma(B) \quad \text{for every fixed } B \in \mathcal{B}.$$

As a consequence, given $n \geq 0$ and $B \in \mathcal{B}$, one obtains

$$E\big\{\gamma(B) \mid \mathcal{G}_n\big\} = \lim_k E\big\{\mu_k(B) \mid \mathcal{G}_n\big\} = \lim_k \frac{1}{k}\sum_{i=n+1}^{k} P\big(X_i \in B \mid \mathcal{G}_n\big) = P\big(X_{n+1} \in B \mid \mathcal{G}_n\big) \text{ a.s..}$$

Suppose next $\mathcal{X} = \mathbb{R}$. Up to enlarging the underlying probability space $(\Omega, \mathcal{A}, P)$, there is an i.i.d. sequence (U_n) with U_1 uniformly distributed on $(0,1)$ and (U_n) independent of (X_n). Define the random distribution function $F(t) = \gamma(-\infty, t]$, $t \in \mathbb{R}$, and

$$Z_n = \inf\{t \in \mathbb{R} : F(t) \geq U_n\}.$$

Then, (Z_n) is exchangeable and $\frac{1}{n}\sum_{i=1}^n I_{\{Z_i \in B\}} \xrightarrow{a.s.} \gamma(B)$ for each $B \in \mathcal{B}$. The exchangeable sequence (Z_n) plays a role in forthcoming Theorems 5.5 and 5.6.

3.2 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,

$$\pi_1 \sim \sigma(\emptyset) \text{ and } \{\sigma(q) : q \in \mathcal{X}^n\} \text{ is a version of the conditional distribution of } \pi_{n+1} \text{ given } (\pi_1,\ldots,\pi_n) \text{ for all } n \geq 1. \tag{3.3}$$

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 $B \in \mathcal{B}$.

If $q = (x_1,\ldots,x_n) \in \mathcal{X}^n$ and $x \in \mathcal{X}$, we write $(q,x) = (x_1,\ldots,x_n,x)$ and $(\emptyset,x) = x$. In this notation, the following result is available.

Theorem 3.1. *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 $q \in \mathcal{X}^n$, $n = 1,2,\ldots$.

Proof. Fix a strategy σ which induces ν. By (3.2) and (3.3), (X_n) is c.i.d. if and only if $X_2 \sim X_1$ and, under ν,

$$\{\sigma(q) : q \in \mathcal{X}^n\} \quad \text{is a version of the conditional distribution of } \pi_{n+2} \text{ given } (\pi_1,\ldots,\pi_n) \text{ for all } n \geq 1. \tag{3.4}$$

In view of (3.3), 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 (3.4) is equivalent to

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

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

Practically, Theorem 3.1 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 on MCMC, e.g. [18] and [20].

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 [14].

Theorem 3.2. *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 $q = \emptyset$).

3.3 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 3.3. *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 3.3 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 3.3. Note also that σ is of Ferguson-Dirichlet type if $z_n = 1$ for all n; see [13].

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

$$P\big((X_1, X_2, \ldots) \in B\big) = \int \nu(B)\, Q(d\nu), \quad B \in \mathcal{B}^\infty, \tag{3.5}$$

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 (3.5).

Example 3.4. *Let m be Lebesgue measure and $f : [0,1] \to [0,1]$ a Borel function satisfying*

$$\int_0^1 f(u)\,du = \frac{1}{2}, \quad \int_0^1 u\,f(u)\,du = \frac{1}{3}, \quad m\{u \in [0,1] : f(u) \neq u\} > 0. \tag{3.6}$$

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 (3.6)*, 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*

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

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 (3.5) *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 (3.6)*. Thus, the probability distribution of (X_n) cannot be written as in* (3.5)*.*

4 Empirical processes

This section includes preliminary material. Apart from Example 4.4, which is new, all other results are from [4]. First, the empirical processes B_n and C_n are introduced for an arbitrary $\mathcal{G}$-adapted sequence (X_n). Then, in case (X_n) is $\mathcal{G}$-c.i.d., some known facts on B_n and C_n are reviewed.

4.1 The general case

Fix a subclass $\mathcal{F} \subset \mathcal{B}$. Also, for any set T, let $l^\infty(T)$ denote the space of real bounded functions on T equipped with the sup-norm

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

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

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

where $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure and $\mu = P \circ X_1^{-1}$ the probability distribution common to the X_n. From the point of view of convergence in distribution, G_n is regarded as a (non measurable) map $G_n : \Omega \to l^\infty(\mathcal{F})$.

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 μ only partially characterizes the probability distribution of the sequence (X_n) (and usually not in the most significant way). So, in the dependent case, G_n is often not much interesting for applications. A second reason is the following. If G_n converges in distribution, as a map $G_n : \Omega \to l^\infty(\mathcal{F})$, then

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

But $\|\mu_n - \mu\|$ typically fails to converge to 0 when (X_n) is non ergodic. 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, if there is a random probability measure γ on $\mathcal{X}$ satisfying

$$\mu_n(B) \xrightarrow{a.s.} \gamma(B) \quad \text{for each fixed } B \in \mathcal{B},$$

it is tempting to let $\gamma_n = \gamma$ for all n. Such γ is actually available when (X_n) is c.i.d.. Further, if (X_n) is exchangeable, (X_n) is conditionally i.i.d. given γ. In the latter case, it is rather natural to take $r_n = \sqrt{n}$. The corresponding empirical process

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

is examined in [4] and [5].

For another example, 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 cannot 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 $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. In such cases, it is important to evaluate the limiting distribution of the error, that is, to investigate convergence in distribution of $\widetilde{G}_n = r_n\,(\mu_n - a_n)$ for suitable constants $r_n \to \infty$. Among other things, μ_n is a "consistent estimate" of a_n if $\widetilde{G}_n$ converges in distribution. In this case, in fact, $\|\mu_n - a_n\| = \frac{1}{r_n}\|\widetilde{G}_n\| \xrightarrow{P} 0$. 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, C_n is investigated in [2], [4], [6].

A third example, considered in [4] for c.i.d. data, is

$$B_n = \sqrt{n}\,(\mu_n - b_n) \quad \text{where } b_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i.$$

The empirical process B_n plays a role in calibration, stochastic approximation and gambling; see [3] and references therein. Following [9] and focusing on calibration, we now give some motivations to B_n.

Example 4.1. *Let $\mathcal{X} = \mathbb{R}$ and T a real random variable. At each time $n \geq 0$, you are requested to predict the event $\{X_{n+1} \leq T\}$ basing on the available information $\mathcal{G}_n$. Your predictor is $P\big(X_{n+1} \leq T \mid \mathcal{G}_n\big)$ and prediction performances are assessed through*

$$V_n = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq T\}} - \frac{1}{n}\sum_{i=1}^{n} P\big(X_i \leq T \mid \mathcal{G}_{i-1}\big) = \mu_n(-\infty, T] - b_n(-\infty, T].$$

Loosely speaking, you are well calibrated *if V_n is small. Let $\mathcal{F} = \{(-\infty, t] : t \in \mathbb{R}\}$. Then, $|V_n| \leq \|\mu_n - b_n\|$. Furthermore, $\|\mu_n - b_n\| \xrightarrow{a.s.} 0$ provided μ_n converges uniformly on $\mathcal{F}$ a.s.; see [3]. Thus, the rate of convergence of $\|\mu_n - b_n\|$ should be investigated, and this leads to $r_n\,(\mu_n - b_n)$ for some choice of the constants r_n. The process B_n corresponds to $r_n = \sqrt{n}$.*

A last remark is that

$$B_n = C_n = W_n = G_n$$

when (X_n) is i.i.d., $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Generally, however, B_n, C_n and W_n are technically harder than G_n 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.

4.2 The case of c.i.d. data

In the sequel, we focus on

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

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 $l^\infty(\mathbb{R})$. Accordingly, B_n and C_n are regarded as maps from Ω into $l^\infty(\mathbb{R})$. Precisely, for each $t \in \mathbb{R}$, they can be written as

$$B_n(t) = \sqrt{n}\,\big\{F_n(t) - b_n(-\infty, t]\big\} \quad \text{and} \quad C_n(t) = \sqrt{n}\,\big\{F_n(t) - P\big(X_{n+1} \leq t \mid \mathcal{G}_n\big)\big\}$$
$$\text{where} \quad F_n(t) = \mu_n(-\infty, t] = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq t\}}.$$

Let $N_k(0, \Sigma)$ denote 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.. Then, a possible limit in distribution for B_n or C_n is a tight random variable $\mathbb{G} : \Omega_0 \to l^\infty(\mathbb{R})$, defined on some probability space $(\Omega_0, \mathcal{A}_0, P_0)$, such that

$$P_0\Big(\big(\mathbb{G}(t_1), \ldots, \mathbb{G}(t_k)\big) \in A\Big) = \int N_k\big(0, \Sigma(t_1, \ldots, t_k)\big)(A)\, dP \tag{4.1}$$

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 the following. Recall that there is a random distribution function F such that $F_n(t) \xrightarrow{a.s.} F(t)$ for each $t \in \mathbb{R}$. Define

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

where $\mathbb{B}$ and M are defined on $(\Omega_0, \mathcal{A}_0, P_0)$, $\mathbb{B}$ is a Brownian bridge, M a random distribution function independent of $\mathbb{B}$, and $M \sim F$. Then, equation (4.1) holds with $\mathbb{G} = \mathbb{G}^F$ and

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

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 [5] and references therein. 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 to be $D = \{t : H(t) > H(t-)\}$. 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 (Y_n) be any sequence of real random processes indexed by $\mathbb{R}$, with bounded cadlag paths, defined on $(\Omega, \mathcal{A}, P)$. According to Theorem 1.5.6 of [21], a necessary condition for Y_n to converge in distribution to a tight limit is: For all ϵ, $\eta > 0$, there is a finite partition $I_1, \ldots, I_m$ of $\mathbb{R}$ by right-open intervals such that

$$\limsup_n P\big(\max_j \sup_{s,t \in I_j} |Y_n(s) - Y_n(t)| > \epsilon\big) < \eta. \tag{4.2}$$

We are now able to state a couple of results from [4].

Theorem 4.2. *Suppose (X_n) is $\mathcal{G}$-c.i.d. and B_n meets (4.2) (i.e., (4.2) holds with $Y_n = B_n$). Then $B_n \xrightarrow{d} \mathbb{G}^F$, under uniform distance on $l^\infty(\mathbb{R})$, and $\mathbb{G}^F$ is tight.*

Theorem 4.3. *Suppose (X_n) is $\mathcal{G}$-c.i.d., C_n meets (4.2), and $\sup_n E\{C_n(t)^2\} < \infty$ for all $t \in \mathbb{R}$. Suppose also that*

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

Then $C_n \xrightarrow{d} \mathbb{G}$, under uniform distance on $l^\infty(\mathbb{R})$, where $\mathbb{G}$ is a tight process with distribution (4.1) and $\Sigma(t_1, \ldots, t_k) = \Big(\sigma(t_i, t_j) : 1 \leq i, j \leq k\Big)$.

Both Theorems 4.2 and 4.3 require condition (4.2) and 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) exchangeable and $\mathbb{G}^F$ tight. Then,

$$W_n = \sqrt{n}\,\{F_n - F\} \xrightarrow{d} \mathbb{G}^F.$$

Hence, W_n meets (4.2) and, as a consequence, B_n and C_n satisfy (4.2) as well; see Remark 4.4 of [4]. Furthermore, $\mathbb{G}^F$ is tight whenever $P(X_1 = X_2) = 0$ or X_1 has a discrete distribution. Unfortunately, this useful criterion fails in the $\mathcal{G}$-c.i.d. case. As we now prove, it may be that (X_n) is $\mathcal{G}$-c.i.d., $\mathbb{G}^F$ tight, and yet condition (4.2) fails for C_n.

Example 4.4. *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*. Given* $t \in \mathbb{R}$*, since* $X_n \xrightarrow{a.s.} X$ *and* $P(X = t) = 0$*, then* $F_n(t) \xrightarrow{a.s.} I_{\{X \le t\}}$*. Therefore,* $F = I_{[X,\infty)}$ *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,*

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

For fixed t*, since* $P(X = t) = 0$*,* $X_n \xrightarrow{a.s.} X$ *and* $S_n \xrightarrow{a.s.} X$*, it is not hard to see that* $C_n(t) \xrightarrow{a.s.} 0$*.*

Toward a contradiction, suppose now that C_n *meets* (4.2) *and define*

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

Then $C_n \xrightarrow{d} 0$*, so that* $|I_n| \le 2\,\|C_n\| \xrightarrow{P} 0$*. 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 (4.2) *fails for* C_n*.*

Incidentally, neither $W_n = \sqrt{n}\,(F_n - F)$ *meets* (4.2)*. In fact,* $W_n(t) \xrightarrow{a.s.} 0$ *for fixed* t*. Thus, if* W_n *meets* (4.2)*, then* $\sup_t |W_n(t) - W_n(t-)| \le 2\,\|W_n\| \xrightarrow{P} 0$*. But this is again a contradiction, for* $P(X_i \ne X \text{ for all } i) = 1$ *and*

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

5 Uniform CLTs for the empirical processes B_n and C_n

Theorems 4.2 and 4.3 apply to $\mathcal{G}$-c.i.d. sequences and refer to uniform distance. In this section, two types of results are obtained. First, Theorems 4.2 and 4.3 are extended to any $\mathcal{G}$-adapted sequence. Second, conditions for B_n and C_n to converge in distribution, under a certain distance weaker than the uniform one, are given for $\mathcal{G}$-c.i.d. sequences. We again let $\mathcal{X} = \mathbb{R}$ and $\mathcal{F} = \{(-\infty, t] : t \in \mathbb{R}\}$.

5.1 Convergence in distribution under uniform distance

Our main tools are the following two (non uniform) CLTs. The first is already known (see Theorem 1 of [7]) while the second is new.

Theorem 5.1. *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 $U_n = E(X_{n+1} \mid \mathcal{G}_n)$. Then,*

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

provided

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

$$\frac{1}{\sqrt{n}}\,E\{\max_{1\le i\le n} i\,|U_{i-1} - U_i|\} \longrightarrow 0,$$

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

Theorem 5.2. *Suppose (X_n) is $\mathcal{G}$-adapted and (X_n^2) uniformly integrable. Then,*

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

whenever

$$\frac{1}{n}\sum_{i=1}^n \{X_i - E(X_i \mid \mathcal{G}_{i-1})\}^2 \xrightarrow{P} L. \tag{5.1}$$

Moreover, condition (5.1) *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 $i = 1, \dots, n$, define $\mathcal{F}_{n,0} = \mathcal{G}_0$, $\mathcal{F}_{n,i} = \mathcal{G}_i$ and

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

Then, $M_n = \sum_{i=1}^n Y_{n,i}$. 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 [15]), 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 (5.1), $\sum_{i=1}^n Y_{n,i}^2 = \frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{P} L$. 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 (5.1). 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 $d = \sqrt{\sup_i E D_i^2}$, one obtains

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

It follows that

$$E\Big|\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{i=1}^n E(X_i^2 \mid \mathcal{G}_{i-1})\Big| \longrightarrow 0, \tag{5.2}$$

$$E\Big|\frac{1}{n}\sum_{i=1}^n D_i^2 - \frac{1}{n}\sum_{i=1}^n \Delta_i\Big| \longrightarrow 0. \tag{5.3}$$

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 (5.2), so that $\frac{1}{n}\sum_{i=1}^n \Delta_i \xrightarrow{P} L$. Thus, (5.3) implies $\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{P} L$, i.e., condition (5.1) 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, (5.1) follows from (5.3). □

In order to apply Theorem 5.2, 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 example possibly exists.

We now turn to uniform CLTs.

Theorem 5.3. *Suppose* (X_n) *is* $\mathcal{G}$*-adapted,* B_n *meets condition* (4.2)*, 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 all $s,\,t\in\mathbb{R}$*. Then* $B_n \xrightarrow{d} \mathbb{G}$*, under uniform distance on* $l^\infty(\mathbb{R})$*, where* $\mathbb{G}$ *is a tight process with distribution* (4.1) *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 (4.2), it suffices to prove convergence of finite dimensional distributions; see e.g. Theorem 1.5.4 of [21]. 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 5.2 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 next result, as in Theorem 4.3, 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 5.4. *Suppose* (X_n) *is* $\mathcal{G}$*-adapted,* C_n *meets condition* (4.2)*, and*

$$n^3\, E\Big\{\big(P(X_{n+2}\le t\mid \mathcal{G}_n)-P(X_{n+1}\le t\mid \mathcal{G}_n)\big)^2\Big\}\longrightarrow 0,$$
$$\frac{1}{\sqrt{n}}\,E\big\{\max_{1\le i\le n}|q_i(t)|\big\}\longrightarrow 0,\quad \frac{1}{n}\sum_{i=1}^n q_i(s)\,q_i(t)\overset{P}{\longrightarrow}\sigma(s,t),$$

for all $s,\,t\in\mathbb{R}$. *Then* $C_n\overset{d}{\longrightarrow}\mathbb{G}$*, under uniform distance on* $l^\infty(\mathbb{R})$*, where* $\mathbb{G}$ *is a tight process with distribution* (4.1) *and* $\Sigma(t_1,\dots,t_k)=\Big(\sigma(t_i,t_j):1\le i,j\le k\Big)$.

Proof. We just give a sketch of the proof, for it is quite analogous to that of Theorem 5.3. By (4.2), it is enough to prove finite dimensional convergence. Fix $t_1,\dots,t_k,u_1,\dots,u_k\in\mathbb{R}$ and define $L=\sum_{r=1}^k\sum_{j=1}^k u_r\,u_j\,\sigma(t_r,t_j)$ and $\nu(\cdot)=\int N(0,L)(\cdot)\,dP$. Since $\sum_{r=1}^k u_r\,\mathbb{G}(t_r)$ has probability distribution ν, 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 5.1 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

$$n^3\,E\big\{\big(E(U_{n+1}\mid\mathcal{G}_n)-U_n\big)^2\big\}\le n^3\,k^2\max_{1\le r\le k}u_r^2\,E\big\{\big(P(X_{n+2}\le t_r\mid\mathcal{G}_n)-P(X_{n+1}\le t_r\mid\mathcal{G}_n)\big)^2\big\}\longrightarrow 0,$$
$$\frac{1}{\sqrt{n}}\,E\big\{\max_{1\le i\le n} i\,|U_{i-1}-U_i|\big\}\le\sum_{r=1}^k|u_r|\,\frac{E\big\{\max_{1\le i\le n}|q_i(t_r)|\big\}+1}{\sqrt{n}}\longrightarrow 0,$$
$$\frac{1}{n}\sum_{i=1}^n\big\{f(X_i)-U_{i-1}+i\,(U_{i-1}-U_i)\big\}^2=\sum_{r=1}^k\sum_{j=1}^k u_r\,u_j\,\frac{1}{n}\sum_{i=1}^n q_i(t_r)\,q_i(t_j)\overset{P}{\longrightarrow}L.$$

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. □

5.2 Convergence in distribution according to Meyer and Zheng

Checking condition (4.2) is hard in real problems. Indeed, unless (X_n) is exchangeable, uniform distance is often too strong for tightness of empirical processes. Conditions for convergence in distribution under some weaker distance, thus, are useful.

Let $L_0=L_0(m)$, where m is Lebesgue measure on the Borel σ-field on $\mathbb{R}$. Thus, L_0 is the space of (equivalence classes of) Borel functions $f:\mathbb{R}\to\mathbb{R}$. Define

$$\rho(f,g)=\int_{-\infty}^{\infty}1\wedge|f(t)-g(t)|\,e^{-|t|}\,dt\quad\text{for } f,\,g\in L_0.$$

Then, ρ is a distance on L_0 and (L_0,ρ) is a Polish space. Further, $\rho(f_n,f)\to 0$ means that $f_n\overset{m}{\longrightarrow}f$ on compacts. Convergence in distribution under ρ has been investigated by Meyer and Zheng in [19]. A clear exposition is in Section 4 of [17].

In the sequel, the expression "almost all $(t_1,\dots,t_k)\in\mathbb{R}^k$" is meant with respect to Lebesgue measure on $\mathbb{R}^k$. A real process U on $(\Omega,\mathcal{A},P)$ is called *jointly measurable* if the map $(t,\omega)\mapsto U(t,\omega)$ is measurable. Two basic facts are to be mentioned.

(i) Let U and V be real jointly measurable processes indexed by $\mathbb{R}$. Then, U and V are L_0-valued random variables and $U \sim V$ if and only if

$$\big(U(t_1),\dots,U(t_k)\big) \sim \big(V(t_1),\dots,V(t_k)\big)$$

for all $k \ge 1$ and almost all $(t_1,\dots,t_k) \in \mathbb{R}^k$.

(ii) Let U_n and U be real jointly measurable processes indexed by $\mathbb{R}$. For $U_n \xrightarrow{d} U$ under ρ, it suffices that the sequence (U_n) is tight under ρ and

$$\big(U_n(t_1),\dots,U_n(t_k)\big) \xrightarrow{d} \big(U(t_1),\dots,U(t_k)\big)$$

for all $k \ge 1$ and almost all $(t_1,\dots,t_k) \in \mathbb{R}^k$.

In view of (i), $B_n : \Omega \to L_0$ and $C_n : \Omega \to L_0$ are L_0-valued random variables. Another fact needs to be recalled from Subsection 3.1. Let (X_n) be $\mathcal{G}$-c.i.d.. Up to enlarging the underlying probability space $(\Omega,\mathcal{A},P)$, there are a random distribution function F and an exchangeable sequence (Z_n) such that

$$F_n(t) \xrightarrow{a.s.} F(t) \quad \text{and} \quad F_n^*(t) \xrightarrow{a.s.} F(t) \quad \text{for each } t \in \mathbb{R}, \tag{5.4}$$
$$\text{where } F_n(t) = \frac{1}{n}\sum_{i=1}^n I_{\{X_i \le t\}} \quad \text{and} \quad F_n^*(t) = \frac{1}{n}\sum_{i=1}^n I_{\{Z_i \le t\}}.$$

Our last results deal with the asymptotic behavior of B_n and C_n under ρ. In such results, (X_n) is $\mathcal{G}$-c.i.d., (Z_n) is exchangeable, F is a random distribution function, and F_n, F_n^* and F satisfy condition (5.4).

Theorem 5.5. *If*

$$\sqrt{n}\,\big\{F_n(t) - F_n^*(t)\big\} \xrightarrow{a.s.} 0 \quad \textit{and} \quad \sqrt{n}\,\big\{F(t) - E(F(t) \mid \mathcal{G}_n)\big\} \xrightarrow{a.s.} 0 \tag{5.5}$$

for almost all $t \in \mathbb{R}$*, then* $B_n \xrightarrow{d} \mathbb{G}^F$ *and* $C_n \xrightarrow{d} \mathbb{G}^F$ *under* ρ*.*

Proof. Let $R_n = \sqrt{n}\,\big\{F_n^* - F\big\}$. If (Z_n) is i.i.d., $R_n \xrightarrow{d} \mathbb{G}^F$ under uniform distance. Hence, $R_n \xrightarrow{d} \mathbb{G}^F$ under ρ, for uniform distance is stronger than ρ. Next, suppose (Z_n) exchangeable (and not necessarily i.i.d.). Since (L_0,ρ) is Polish, de Finetti's representation theorem again implies $R_n \xrightarrow{d} \mathbb{G}^F$ under ρ. Thus, it suffices to prove that $\rho(B_n,R_n) \xrightarrow{P} 0$ and $\rho(C_n,R_n) \xrightarrow{P} 0$.

Since $P(X_i \le t \mid \mathcal{G}_{i-1}) = E(F(t) \mid \mathcal{G}_{i-1})$ a.s., then

$$B_n(t) - R_n(t) = \sqrt{n}\,\big\{F_n(t) - F_n^*(t)\big\} + \sum_{i=1}^n (n\,i)^{-1/2}\Big(\sqrt{i}\,\big\{F(t) - E(F(t) \mid \mathcal{G}_{i-1})\big\}\Big) \text{ a.s..}$$

Because of (5.5) and $\sum_{i=1}^n (n\,i)^{-1/2} \to 2$, one obtains $B_n(t) - R_n(t) \xrightarrow{a.s.} 0$ for almost all $t \in \mathbb{R}$. Letting $A = \{(t,\omega) : B_n(t,\omega) - R_n(t,\omega) \text{ does not converge to } 0\}$, it follows that

$$\int_\Omega m\{t : (t,\omega) \in A\}\, P(d\omega) = \int_{-\infty}^{\infty} P\{\omega : (t,\omega) \in A\}\, dt = 0.$$

Hence $B_n(\cdot,\omega) - R_n(\cdot,\omega) \longrightarrow 0$, m-a.s., for each ω in a set of P-probability 1. This implies $\rho(B_n,R_n) \xrightarrow{a.s.} 0$. Finally, $\rho(C_n,R_n) \xrightarrow{a.s.} 0$ follows from exactly the same argument. □

Theorem 5.6. *Suppose X_1 discrete or*

$$\inf_{\epsilon>0}\liminf_n P\big(|X_n - X_{n+1}| < \epsilon\big) = 0.$$

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

$$\sqrt{n}\,E\Big|F_n(t) - F_n^*(t)\Big| \longrightarrow 0 \quad \text{for almost all } t \in \mathbb{R}, \tag{5.6}$$

then $B_n \xrightarrow{d} \mathbb{G}^F$ under ρ. If, in addition,

$$\frac{1}{\sqrt{n}}\,E\big\{\max_{1\le i\le n}|q_i(t)|\big\} \longrightarrow 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n q_i(s)\,q_i(t) \xrightarrow{P} \sigma(s,t) \quad \text{for almost all } (s,t) \in \mathbb{R}^2,$$

then $C_n \xrightarrow{d} \mathbb{G}$ under ρ. Here, $\mathbb{G}$ is a process satisfying equation (4.1) with $\Sigma(t_1,\ldots,t_k) = \Big(\sigma(t_i,t_j) : 1 \le i,j \le k\Big)$ for all $k \ge 1$ and almost all $(t_1,\ldots,t_k) \in \mathbb{R}^k$.

Proof. Since (X_n) is $\mathcal{G}$-c.i.d., the finite dimensional distributions of B_n converge weakly to those of $\mathbb{G}^F$; see Theorem 4.2 of [4] and its proof. Similarly, arguing as in the proof of Theorem 5.4 and using the conditions on $q_i(t)$, one obtains

$$\big(C_n(t_1),\ldots,C_n(t_k)\big) \xrightarrow{d} \big(\mathbb{G}(t_1),\ldots,\mathbb{G}(t_k)\big)$$

for all $k \ge 1$ and almost all $(t_1,\ldots,t_k) \in \mathbb{R}^k$. Therefore, it is enough to prove tightness of (B_n) and (C_n) under ρ.

By Theorem 2.5 of [4], $(X_n, X_{n+1}) \xrightarrow{d} (Z_1, Z_2)$. Thus, Z_1 is discrete if X_1 is discrete. Otherwise, if X_1 is not discrete,

$$P(Z_1 = Z_2) = \inf_{\epsilon>0} P\big(|Z_1 - Z_2| < \epsilon\big) \le \inf_{\epsilon>0}\liminf_n P\big(|X_n - X_{n+1}| < \epsilon\big) = 0.$$

Let $R_n = \sqrt{n}\,\big\{F_n^* - F\big\}$. Since (Z_n) is exchangeable and Z_1 is discrete or $P(Z_1 = Z_2) = 0$, then $R_n \xrightarrow{d} \mathbb{G}^F$ under uniform distance; see Subsection 4.2. Given $t \in \mathbb{R}$, recall $E\big\{F(t) \mid \mathcal{G}_n\big\} = P\big(X_{n+1} \le t \mid \mathcal{G}_n\big)$ a.s. and define

$$Y_n(t) = E\big\{R_n(t) \mid \mathcal{G}_n\big\} = \sqrt{n}\,\Big\{E\big\{F_n^*(t) \mid \mathcal{G}_n\big\} - P\big(X_{n+1} \le t \mid \mathcal{G}_n\big)\Big\}.$$

Since R_n satisfies condition (4.2), Y_n meets (4.2) as well; see Remark 4.4 of [4]. Hence, (Y_n) is tight under ρ (for uniform distance is stronger than ρ). Further,

$$E\big\{\rho(C_n, Y_n)\big\} = \int_{-\infty}^{\infty} E\Big\{1 \wedge \Big|\sqrt{n}\,E\big\{F_n(t) - F_n^*(t) \mid \mathcal{G}_n\big\}\Big|\Big\}\,e^{-|t|}\,dt$$

$$\le \int_{-\infty}^{\infty} 1 \wedge \Big\{\sqrt{n}\,E\Big|F_n(t) - F_n^*(t)\Big|\Big\}\,e^{-|t|}\,dt \longrightarrow 0 \quad \text{by condition (5.6).}$$

Thus, (C_n) is tight under ρ, because (Y_n) is tight under ρ and $\rho(C_n, Y_n) \xrightarrow{P} 0$.

We finally prove (B_n) tight under ρ. Let

$$D_n(t) = R_n(t) + \frac{E(F(t) \mid \mathcal{G}_n) - E(F(t) \mid \mathcal{G}_0)}{\sqrt{n}} + \sum_{i=1}^n (n\,i)^{-1/2}\,\big\{Y_i(t) - R_i(t)\big\}.$$

Since (D_n) is tight under ρ, it suffices to show that $E\big\{\rho(B_n, D_n)\big\} \longrightarrow 0$. Write B_n as

$$B_n(t) = W_n(t) + \frac{E(F(t) \mid \mathcal{G}_n) - E(F(t) \mid \mathcal{G}_0)}{\sqrt{n}} + \sum_{i=1}^n (n\,i)^{-1/2}\,\big\{C_i(t) - W_i(t)\big\}$$

where $W_n = \sqrt{n}\,\{F_n - F\}$. Then, condition (5.6) implies $E\Big|B_n(t) - D_n(t)\Big| \longrightarrow 0$ for almost all $t \in \mathbb{R}$, which in turn implies $E\{\rho(B_n, D_n)\} \longrightarrow 0$.

□

The idea underlying Theorems 5.5-5.6 is straightforward: B_n and C_n behave nicely under ρ provided (X_n) is close enough to a suitable exchangeable sequence (Z_n). In the spirit of [1], we note that (X_{n_j}) is actually close to (Z_{n_j}) for some subsequence (n_j). Finally, we give a (technical) remark and an example.

Remark 5.7. *Under the assumptions of Theorem 5.6, one obtains*

$$B_n = D_n + D_n^* \quad \text{and} \quad C_n = Y_n + Y_n^*$$

with D_n and Y_n satisfying (4.2) *and $D_n^* \xrightarrow{P} 0$, $Y_n^* \xrightarrow{P} 0$ under ρ. Moreover, if the conditions on $q_i(t)$ hold for* every $(s,t) \in \mathbb{R}^2$ *and*

$$\sqrt{n}\, E\|F_n - F_n^*\| \longrightarrow 0,$$

then $B_n \xrightarrow{d} \mathbb{G}^F$ and $C_n \xrightarrow{d} \mathbb{G}$ under uniform distance.

Example 5.8 (Example 4.4 continued)**.** *Let the notation of Example 4.4 prevail. Since $F = I_{[X,\infty)}$, the exchangeable sequence (Z_n) can be taken to be $Z_n = X$ for all n. In particular, $\mathbb{G}^F = 0$ and $F_n^* = F$ for all n. Since $X_n \xrightarrow{a.s.} X$, $S_n \xrightarrow{a.s.} X$, and $P(X = t) = 0$ for fixed t, condition* (5.5) *is trivially true. Thus, Theorem 5.5 yields $B_n \xrightarrow{d} 0$ and $C_n \xrightarrow{d} 0$ under ρ.*

References

[1] Aldous, D.J.: Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. *Prob. Theo. Relat. Fields* **40**, (1977) , 59–82. MR-0455090

[2] Bassetti, F., Crimaldi, I. and Leisen, F.: Conditionally identically distributed species sampling sequence. *Adv. in Appl. Probab.* **42**, (2010), 433–459. MR-2675111

[3] Berti, P. and Rigo, P.: A uniform limit theorem for predictive distributions. *Statist. Probab. Letters* **56**, (2002), 113–120. MR-1881164

[4] Berti, P., Pratelli, L. and Rigo, P.: Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **32**, (2004), 2029–2052. MR-2073184

[5] Berti, P., Pratelli, L. and Rigo, P.: Asymptotic behaviour of the empirical process for exchangeable data. *Stoch. Proc. Appl.* **116**, (2006), 337–344. MR-2197981

[6] Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P.: Rate of convergence of predictive distributions for dependent data. *Bernoulli* **15**, (2009), 1351–1367. MR-2597596

[7] Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P.: A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.* **48**, (2011), 527–546. MR-2840314

[8] Crimaldi, I., Letta, G. and Pratelli, L.: A strong form of stable convergence. *Seminaire de Probabilites* XL, Lect. Notes in Math., 1899, (2007), 203–225. MR-2409006

[9] Dawid, P.: The well-calibrated Bayesian (with discussion). *J. Amer. Statist. Assoc.* **77**, (1982), 605–613. MR-0675887

[10] Dubins, L.E. and Savage, L.J.: *How to gamble if you must: Inequalities for stochastic processes.* McGraw Hill, (1965). MR-0236983

[11] Dudley, R.M.: *Uniform central limit theorems.* Cambridge University Press, (1999). MR-1720712

[12] Etemadi, N. and Kaminski, M.: Strong law of large numbers for 2-exchangeable random variables. *Statist. Probab. Letters* **28**, (1996), 245–250. MR-1406997

[13] Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, (1973), 209–230. MR-0350949

[14] Fortini, S., Ladelli, L. and Regazzini, E.: Exchangeability, predictive distributions and parametric models. *Sankhya A* **62**, (2000), 86–109. MR-1769738

[15] Hall, P. and Heyde, C.C.: *Martingale limit theory and its applications.* Academic Press, (1980). MR-0624435

[16] Kallenberg, O.: Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.* **16**, (1988), 508–534. MR-0929061

[17] Kurtz, T.G.: Random time changes and convergence in distribution under the Meyer-Zheng conditions. *Ann. Probab.* **19**, (1991), 1010–1034. MR-1112405

[18] Liu, J.S.: *Monte Carlo strategies in scientific computing.* Springer, (2001). MR-1842342

[19] Meyer, P.A. and Zheng, W.A.: Tightness criteria for laws of semimartingales. *Ann. Inst. H. Poincare Probab. Statist.* **20**, (1984), 353–372. MR-0771895

[20] Tierney, L.: Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, (1994), 1701–1762. MR-1329166

[21] van der Vaart, A. and Wellner, J.A.: *Weak convergence and empirical processes.* Springer, (1996). MR-1385671