Asymptotics of Predictive Distributions

Patrizia Berti, Luca Pratelli and Pietro Rigo

Abstract Let (X_n) be a sequence of random variables, adapted to a filtration $(\mathcal{G}_n)$, and let $\mu_n = (1/n)\sum_{i=1}^n \delta_{X_i}$ and $a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ be the empirical and the predictive measures. We focus on $\|\mu_n - a_n\| = \sup_{B\in\mathcal{D}} |\mu_n(B) - a_n(B)|$, where $\mathcal{D}$ is a class of measurable sets. Conditions for $\|\mu_n - a_n\| \to 0$, almost surely or in probability, are given. Also, to determine the rate of convergence, the asymptotic behavior of $r_n\,\|\mu_n - a_n\|$ is investigated for suitable constants r_n. Special attention is paid to $r_n = \sqrt{n}$. The sequence (X_n) is exchangeable or, more generally, conditionally identically distributed.

1 Introduction

1.1 The Problem

Throughout, S is a Polish space and $X = (X_n : n \geq 1)$ a sequence of S-valued random variables on the probability space $(\Omega, \mathcal{A}, P)$. Further, $\mathcal{B}$ is the Borel σ-field on S and $\mathcal{G} = (\mathcal{G}_n : n \geq 0)$ a filtration on $(\Omega, \mathcal{A}, P)$. We fix a subclass $\mathcal{D} \subset \mathcal{B}$ and we let $\|\cdot\|$ denote the sup-norm over $\mathcal{D}$, namely, $\|\alpha - \beta\| = \sup_{B\in\mathcal{D}} |\alpha(B) - \beta(B)|$ whenever α and β are probabilities on $\mathcal{B}$.

Let

$$\mu_n = (1/n)\sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{G}_n).$$

P. Berti
Universita' di Modena e Reggio-Emilia, Modena, Italy
e-mail: patrizia.berti@unimore.it

L. Pratelli
Accademia Navale di Livorno, Livorno, Italy
e-mail: pratel@mail.dm.unipi.it

P. Rigo (✉)
Universita' di Pavia, Pavia, Italy
e-mail: pietro.rigo@unipv.it

M.B. Ferraro et al. (eds.), *Soft Methods for Data Science*, Advances in Intelligent Systems and Computing 456, DOI 10.1007/978-3-319-42972-4_7

Both μ_n and a_n are regarded as random probability measures on $\mathcal{B}$; μ_n is the empirical measure and (if X is $\mathcal{G}$-adapted) a_n is the predictive measure.

Under some conditions, $\mu_n(B) - a_n(B) \xrightarrow{a.s.} 0$ for fixed $B \in \mathcal{B}$. In that case, a (natural) question is whether $\mathcal{D}$ is such that $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$.

Such question is addressed in this paper. Conditions for $\|\mu_n - a_n\| \to 0$, almost surely or in probability, are given. Also, to determine the rate of convergence, the asymptotic behavior of $r_n \|\mu_n - a_n\|$ is investigated for suitable constants r_n. Special attention is paid to $r_n = \sqrt{n}$. The sequence X is assumed to be exchangeable or, more generally, conditionally identically distributed (see Sect. 2).

Our main concern is to connect and unify a few results from [1–4]. Thus, this paper is essentially a survey. However, in addition to report known facts, some new results and examples are given. This is actually the case of Theorem 1(d), Corollary 1 and Examples 1–3.

1.2 Heuristics

There are various (non-independent) reasons for investigating $\mu_n - a_n$. We now list a few of them under the assumption that $\mathcal{G} = \mathcal{G}^X$, where $\mathcal{G}_0^X = \{\emptyset, \Omega\}$ and $\mathcal{G}_n^X = \sigma(X_1, \ldots, X_n)$. Most remarks, however, apply to any filtration $\mathcal{G}$ which makes X adapted.

- **Empirical processes for non-ergodic data**. Slightly abusing terminology, say that X is ergodic if P is 0–1 valued on the sub-σ-field $\sigma\big(\limsup_n \mu_n(B) : B \in \mathcal{B}\big)$. In real problems, X is often non-ergodic. Most stationary sequences, for instance, fail to be ergodic. Or else, an exchangeable sequence is ergodic if and only if is i.i.d. Now, if X is i.i.d., the empirical process is defined as $G_n = \sqrt{n}\,(\mu_n - \mu_0)$ where μ_0 is the probability distribution of X_1. But this definition has various drawbacks when X is not ergodic; see [5]. In fact, unless X is i.i.d., the probability distribution of X is not determined by that of X_1. More importantly, if G_n converges in distribution in $l^\infty(\mathcal{D})$ (the metric space $l^\infty(\mathcal{D})$ is recalled before Corollary 1) then $\|\mu_n - \mu_0\| = n^{-1/2}\|G_n\| \xrightarrow{P} 0$. But $\|\mu_n - \mu_0\|$ typically fails to converge to 0 in probability when X is not ergodic. Thus, empirical processes for non-ergodic data should be defined in some different way. In this framework, a meaningful option is to replace μ_0 with a_n, namely, to let $G_n = \sqrt{n}\,(\mu_n - a_n)$.
- **Bayesian predictive inference**. In a number of problems, the main goal is to evaluate a_n but the latter can not be obtained in closed form. Thus, a_n is to be estimated by the available data. Under some assumptions, a reasonable estimate of a_n is just μ_n. In these situations, the asymptotic behavior of the error $\mu_n - a_n$ plays a role. For instance, μ_n is a consistent estimate of a_n provided $\|\mu_n - a_n\| \longrightarrow 0$ in some sense.

- **Predictive distributions of exchangeable sequences**. Let X be exchangeable. Just very little is known on the general form of a_n for given n, and a representation theorem for a_n would be actually a major breakthrough. Failing the latter, to fix the asymptotic behavior of $\mu_n - a_n$ contributes to fill the gap.
- **de Finetti**. Historically, one reason for introducing exchangeability (possibly, the main reason) was to justify observed frequencies as predictors of future events. See [8–10]. In this sense, to focus on $\mu_n - a_n$ is in line with de Finetti's ideas. Roughly speaking, μ_n should be a good substitute of a_n in the exchangeable case.

2 Conditionally Identically Distributed Sequences

The sequence X is *conditionally identically distributed* (c.i.d.) with respect to $\mathcal{G}$ if it is $\mathcal{G}$-adapted and $P\big(X_k \in \cdot \mid \mathcal{G}_n\big) = P\big(X_{n+1} \in \cdot \mid \mathcal{G}_n\big)$ a.s. for all $k > n \geq 0$. Roughly speaking, at each time $n \geq 0$, the future observations $(X_k : k > n)$ are identically distributed given the past $\mathcal{G}_n$. When $\mathcal{G} = \mathcal{G}^X$, the filtration $\mathcal{G}$ is not mentioned at all and X is just called c.i.d. Then, X is c.i.d. if and only if $\big(X_1, \dots, X_n, X_{n+2}\big) \sim \big(X_1, \dots, X_n, X_{n+1}\big)$ for all $n \geq 0$.

Exchangeable sequences are c.i.d. while the converse is not true. Indeed, X is exchangeable if and only if it is stationary and c.i.d. We refer to [3] for more on c.i.d. sequences. Here, it suffices to mention a last fact.

If X is c.i.d., there is a random probability measure μ on $\mathcal{B}$ such that $\mu_n(B) \xrightarrow{a.s.} \mu(B)$ for every $B \in \mathcal{B}$. As a consequence, if X is c.i.d. with respect to $\mathcal{G}$, for each $n \geq 0$ and $B \in \mathcal{B}$ one obtains

$$E\big\{\mu(B) \mid \mathcal{G}_n\big\} = \lim_m E\big\{\mu_m(B) \mid \mathcal{G}_n\big\} = \lim_m \frac{1}{m} \sum_{k=n+1}^{m} P\big(X_k \in B \mid \mathcal{G}_n\big)$$
$$= P\big(X_{n+1} \in B \mid \mathcal{G}_n\big) = a_n(B) \quad \text{a.s.}$$

In particular, $a_n(B) = E\big\{\mu(B) \mid \mathcal{G}_n\big\} \xrightarrow{a.s.} \mu(B)$ and $\mu_n(B) - a_n(B) \xrightarrow{a.s.} 0$.

From now on, X is c.i.d. with respect to $\mathcal{G}$. In particular, X is identically distributed and μ_0 denotes the probability distribution of X_1. We also let

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

where μ is the random probability measure on $\mathcal{B}$ introduced above. Note that, if X is i.i.d., then $\mu = \mu_0$ a.s. and W_n reduces to the usual empirical process.

3 Results

Let $\mathcal{D} \subset \mathcal{B}$. To avoid measurability problems, $\mathcal{D}$ is assumed to be *countably determined*. This means that there is a countable subclass $\mathcal{D}_0 \subset \mathcal{D}$ such that $\|\alpha - \beta\| = \sup_{B \in \mathcal{D}_0} |\alpha(B) - \beta(B)|$ for all probabilities α, β on $\mathcal{B}$. For instance, $\mathcal{D} = \mathcal{B}$ is countably determined (for $\mathcal{B}$ is countably generated). Or else, if $S = \mathbb{R}^k$, then $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}^k\}$, $\mathcal{D} = \{\text{closed balls}\}$ and $\mathcal{D} = \{\text{closed convex sets}\}$ are countably determined.

3.1 A General Criterion

Since $a_n(B) = E\{\mu(B) \mid \mathcal{G}_n\}$ a.s. for each $B \in \mathcal{B}$ and $\mathcal{D}$ is countably determined, one obtains

$$\|\mu_n - a_n\| = \sup_{B \in \mathcal{D}_0} |E\{\mu_n(B) - \mu(B) \mid \mathcal{G}_n\}| \leq E\{\|\mu_n - \mu\| \mid \mathcal{G}_n\} \quad \text{a.s.}$$

This simple inequality has some nice consequences. Recall that $\mathcal{D}$ is a *universal Glivenko-Cantelli class* if $\|\mu_n - \mu_0\| \xrightarrow{a.s.} 0$ whenever X is i.i.d.

Theorem 1 *Suppose $\mathcal{D}$ is countably determined and X is c.i.d. with respect to $\mathcal{G}$. Then,*

(a) $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$ *if* $\|\mu_n - \mu\| \xrightarrow{a.s.} 0$ *and* $\|\mu_n - a_n\| \xrightarrow{P} 0$ *if* $\|\mu_n - \mu\| \xrightarrow{P} 0$.
(b) $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$ *provided X is exchangeable, $\mathcal{G} = \mathcal{G}^X$ and $\mathcal{D}$ is a universal Glivenko-Cantelli class.*
(c) $r_n \|\mu_n - a_n\| \xrightarrow{P} 0$ *whenever the constants r_n satisfy $r_n/\sqrt{n} \to 0$ and* $\sup_n E\{\|W_n\|^b\} < \infty$ *for some $b \geq 1$.*
(d) $n^u \|\mu_n - a_n\| \xrightarrow{a.s.} 0$ *whenever $u < 1/2$ and* $\sup_n E\{\|W_n\|^b\} < \infty$ *for each $b \geq 1$.*

Proof Since $\|\mu_n - \mu\| \leq 1$, point (a) follows from the martingale convergence theorem in the version of [7]. (If $\|\mu_n - \mu\| \xrightarrow{P} 0$, it suffices to apply an obvious argument based on subsequences). Next, suppose X, $\mathcal{G}$ and $\mathcal{D}$ are as in (b). By de Finetti's theorem, conditionally on μ, the sequence X is i.i.d. with common distribution μ. Since $\mathcal{D}$ is a universal Glivenko-Cantelli class, it follows that $P(\|\mu_n - \mu\| \to 0) = \int P\{\|\mu_n - \mu\| \to 0 \mid \mu\}\, dP = \int 1 dP = 1$. Hence, (b) is a consequence of (a). As to (c), just note that

$$E\Big\{\big(r_n \|\mu_n - a_n\|\big)^b\Big\} \leq r_n^b\, E\{\|\mu_n - \mu\|^b\} = (r_n/\sqrt{n})^b\, E\{\|W_n\|^b\}.$$

Finally, as to (d), fix $u < 1/2$ and take b such that $b(1/2 - u) > 1$. Then,

$$\sum_n P\big(n^u \|\mu_n - a_n\| > \epsilon\big) \leq \sum_n \frac{E\big\{\|\mu_n - a_n\|^b\big\}}{\epsilon^b\, n^{-ub}} \leq \sum_n \frac{E\big\{\|\mu_n - \mu\|^b\big\}}{\epsilon^b\, n^{-ub}}$$
$$= \sum_n \frac{E\big\{\|W_n\|^b\big\}}{\epsilon^b\, n^{(1/2-u)b}} \leq \sum_n \frac{\text{const}}{n^{(1/2-u)b}} < \infty \quad \text{for each } \epsilon > 0.$$

Therefore, $n^u \|\mu_n - a_n\| \xrightarrow{a.s.} 0$ because of the Borel-Cantelli lemma.

Some remarks are in order.

Theorem 1 is essentially known. Apart from (d), it is implicit in [2, 4].

If X is exchangeable, the second part of (a) is redundant. In fact, $\|\mu_n - \mu_0\|$ converges a.s. (not necessarily to 0) whenever X is i.i.d. Applying de Finetti's theorem as in the proof of Theorem 1(b), it follows that $\|\mu_n - \mu\|$ converges a.s. even if X is exchangeable. Thus, $\|\mu_n - \mu\| \xrightarrow{P} 0$ implies $\|\mu_n - \mu\| \xrightarrow{a.s.} 0$.

Sometimes, the condition in (a) is necessary as well, namely, $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$ if and only if $\|\mu_n - \mu\| \xrightarrow{a.s.} 0$. For instance, this happens when $\mathcal{G} = \mathcal{G}^X$ and $\mu \ll \lambda$ a.s., where λ is a (non-random) σ-finite measure on $\mathcal{B}$. In this case, in fact, $\|a_n - \mu\| \xrightarrow{a.s.} 0$ by [6, Theorem 1].

Several examples of universal Glivenko-Cantelli classes are available; see [11] and references therein. Similarly, for many choices of $\mathcal{D}$ and $b \geq 1$ there is a universal constant $c(b)$ such that $\sup_n E\big\{\|W_n\|^b\big\} \leq c(b)$ provided X is i.i.d.; see e.g. [11, Sects. 2.14.1 and 2.14.2]. In these cases, de Finetti's theorem yields $\sup_n E\big\{\|W_n\|^b\big\} \leq c(b)$ even if X is exchangeable. Thus, points (b)–(d) are especially useful when X is exchangeable.

In (c), convergence in probability can not be replaced by a.s. convergence. As a trivial example, take $\mathcal{D} = \mathcal{B}$, $\mathcal{G} = \mathcal{G}^X$, $r_n = \sqrt{\frac{n}{\log\log n}}$, and X an i.i.d. sequence of indicators. Letting $p = P(X_1 = 1)$, one obtains $E\big\{\|W_n\|^2\big\} = n\, E\big\{\big(\mu_n\{1\} - p\big)^2\big\} = p\,(1-p)$ for all n. However, the LIL yields

$$\limsup_n\, r_n\, \|\mu_n - a_n\| = \limsup_n \frac{|\sum_{i=1}^n (X_i - p)\,|}{\sqrt{n}\, \log\log n} = \sqrt{2\, p\,(1-p)} \quad \text{a.s.}$$

We finally give a couple of examples.

Example 1 Let $\mathcal{D} = \mathcal{B}$. If X is i.i.d., then $\|\mu_n - \mu_0\| \xrightarrow{a.s.} 0$ if and only if μ_0 is discrete. By de Finetti's theorem, it follows that $\|\mu_n - \mu\| \xrightarrow{a.s.} 0$ whenever X is exchangeable and μ is a.s. discrete. Thus, under such assumptions and $\mathcal{G} = \mathcal{G}^X$, Theorem 1(a) implies $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$. This result has possible practical interest. In fact, in Bayesian nonparametrics, most priors are such that μ is a.s. discrete.

Example 2 Let $S = \mathbb{R}^k$ and $\mathcal{D} = \{\text{closed convex sets}\}$. Given any probability α on $\mathcal{B}$, denote by $\alpha^{(c)} = \alpha - \sum_x \alpha\{x\}\delta_x$ the continuous part of α. If X is i.i.d. and $\mu_0^{(c)} \ll m$, where m is Lebesgue measure, then $\|\mu_n - \mu_0\| \xrightarrow{a.s.} 0$. Applying Theorem 1(a) again, one obtains $\|\mu_n - a_n\| \xrightarrow{a.s.} 0$ provided X is exchangeable, $\mathcal{G} = \mathcal{G}^X$ and $\mu^{(c)} \ll m$ a.s. While "morally true", this argument does not work for $\mathcal{D} = \{\text{Borel convex sets}\}$ since the latter choice of $\mathcal{D}$ is not countably determined.

3.2 The Dominated Case

In this Subsection, $\mathcal{G} = \mathcal{G}^X$, $\mathcal{A} = \sigma\big(\cup_n \mathcal{G}_n^X\big)$, Q is a probability on $(\Omega, \mathcal{A})$ and $b_n(\cdot) = Q(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ is the predictive measure under Q. Also, we say that Q is a Ferguson-Dirichlet law if

$$b_n(\cdot) = \frac{c\, Q(X_1 \in \cdot) + n\, \mu_n(\cdot)}{c+n}, \quad Q\text{-a.s. for some constant } c > 0.$$

If $P \ll Q$, the asymptotic behavior of $\mu_n - a_n$ under P should be affected by that of $\mu_n - b_n$ under Q. This (rough) idea is realized by the next result.

Theorem 2 (Theorems 1 and 2 of [4]) *Suppose $\mathcal{D}$ is countably determined, X is c.i.d., and $P \ll Q$. Then, $\sqrt{n}\,\|\mu_n - a_n\| \xrightarrow{P} 0$ provided $\sqrt{n}\,\|\mu_n - b_n\| \xrightarrow{Q} 0$ and the sequence (W_n) is uniformly integrable under both P and Q. In addition, $n\,\|\mu_n - a_n\|$ converges a.s. to a finite limit whenever Q is a Ferguson-Dirichlet law, $\sup_n E_Q\big\{\|W_n\|^2\big\} < \infty$, and*

$$\sup_n n\, \Big\{E_Q\big\{(dP/dQ)^2\big\} - E_Q\big\{E_Q(dP/dQ \mid \mathcal{G}_n)^2\big\}\Big\} < \infty.$$

To make Theorem 2 effective, the condition $P \ll Q$ should be given a simple characterization. This happens in at least one case.

Let S be finite, say $S = \{x_1, \dots, x_k, x_{k+1}\}$, X exchangeable and $\mu_0\{x\} > 0$ for all $x \in S$. Then $P \ll Q$, with Q a Ferguson-Dirichlet law, if and only if the distribution of $\big(\mu\{x_1\}, \dots, \mu\{x_k\}\big)$ is absolutely continuous (with respect to Lebesgue measure). This fact is behind the next result.

Theorem 3 (Corollaries 4 and 5 of [4]) *Suppose $S = \{0, 1\}$ and X is exchangeable. Then, $\sqrt{n}\,\big(\mu_n\{1\} - a_n\{1\}\big) \xrightarrow{P} 0$ whenever the distribution of $\mu\{1\}$ is absolutely continuous. Moreover, $n\,\big(\mu_n\{1\} - a_n\{1\}\big)$ converges a.s. (to a finite limit) provided the distribution of $\mu\{1\}$ is absolutely continuous with an almost Lipschitz density.*

In Theorem 3, a real function f on $(0, 1)$ is said to be *almost Lipschitz* in case $x \mapsto f(x)x^u(1-x)^v$ is Lipschitz on $(0, 1)$ for some reals $u,\ v < 1$.

A consequence of Theorem 3 is to be stressed. For each $B \in \mathcal{B}$, define

$$T_n(B) = \sqrt{n}\,\Big\{a_n(B) - P\big\{X_{n+1} \in B \mid \mathcal{G}_n^B\big\}\Big\}$$

where $\mathcal{G}_n^B = \sigma\big(I_B(X_1), \ldots, I_B(X_n)\big)$. Also, let $l^\infty(\mathcal{D})$ be the set of real bounded functions on $\mathcal{D}$, equipped with uniform distance. In the next result, W_n is regarded as a random element of $l^\infty(\mathcal{D})$ and convergence in distribution is meant in Hoffmann-Jørgensen's sense; see [11].

Corollary 1 *Let $\mathcal{D}$ be countably determined and X exchangeable. Suppose*

(i) $\mu(B)$ has an absolutely continuous distribution for each $B \in \mathcal{D}$ such that $0 < P(X_1 \in B) < 1$;
(ii) the sequence $(\|W_n\|)$ is uniformly integrable;
(iii) W_n converges in distribution to a tight limit in $l^\infty(\mathcal{D})$.

Then, $\sqrt{n}\,\|\mu_n - a_n\| \xrightarrow{P} 0$ if and only if $T_n(B) \xrightarrow{P} 0$ for each $B \in \mathcal{D}$.

Proof Let $U_n(B) = \sqrt{n}\,\Big\{\mu_n(B) - P\big\{X_{n+1} \in B \mid \mathcal{G}_n^B\big\}\Big\}$. Then, $U_n(B) \xrightarrow{P} 0$ for each $B \in \mathcal{D}$. In fact, $U_n(B) = 0$ a.s. if $P(X_1 \in B) \in \{0, 1\}$. Otherwise, $U_n(B) \xrightarrow{P} 0$ follows from Theorem 3, since $(I_B(X_n))$ is an exchangeable sequence of indicators and $\mu(B)$ has an absolutely continuous distribution. Next, suppose $T_n(B) \xrightarrow{P} 0$ for each $B \in \mathcal{D}$. Letting $C_n = \sqrt{n}\,(\mu_n - a_n)$, we have to prove that $\|C_n\| \xrightarrow{P} 0$. Equivalently, regarding C_n as a random element of $l^\infty(\mathcal{D})$, we have to prove that $C_n(B) \xrightarrow{P} 0$ for fixed $B \in \mathcal{D}$ and the sequence (C_n) is asymptotically tight; see e.g. [11, Sect. 1.5]. Given $B \in \mathcal{D}$, since both $U_n(B)$ and $T_n(B)$ converge to 0 in probability, then $C_n(B) = U_n(B) - T_n(B) \xrightarrow{P} 0$. Moreover, since $C_n(B) = E\big\{W_n(B) \mid \mathcal{G}_n\big\}$ a.s., the asymptotic tightness of (C_n) follows from (ii) and (iii); see [3, Remark 4.4]. Hence, $\|C_n\| \xrightarrow{P} 0$. Conversely, if $\|C_n\| \xrightarrow{P} 0$, one trivially obtains

$$|T_n(B)| = |U_n(B) - C_n(B)| \le |U_n(B)| + \|C_n\| \xrightarrow{P} 0 \quad \text{for each } B \in \mathcal{D}.$$

If X is exchangeable, it frequently happens that $\sup_n E\big\{\|W_n\|^2\big\} < \infty$, which in turn implies condition (ii). Similarly, (iii) is not unusual. As an example, conditions (ii) and (iii) hold if $S = \mathbb{R}$, $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$ and μ_0 is discrete or $P(X_1 = X_2) = 0$; see [3, Theorem 4.5].

Unfortunately, as shown by the next example, $T_n(B)$ may fail to converge to 0 even if $\mu(B)$ has an absolutely continuous distribution. This suggests the following general question. In the exchangeable case, in addition to $\mu_n(B)$, which further information is required to evaluate $a_n(B)$? Or at least, are there reasonable conditions for $T_n(B) \xrightarrow{P} 0$? Even if intriguing, to our knowledge, such a question does not have a satisfactory answer.

Example 3 Let $S = \mathbb{R}$ and $X_n = Y_n\, Z^{-1}$, where Y_n and Z are independent real random variables, $Y_n \sim N(0, 1)$ for all n, and Z has an absolutely continuous distribution supported by $[1, \infty)$. Conditionally on Z, the sequence $X = (X_1, X_2, \ldots)$ is i.i.d. with common distribution $N(0, Z^{-2})$. Thus, X is exchangeable and $\mu(B) = P(X_1 \in B \mid Z) = f_B(Z)$ a.s., where

$$f_B(z) = (2\pi)^{-1/2} z \int_B \exp\big(-(xz)^2/2\big)\, dx \quad \text{for } B \in \mathcal{B} \text{ and } z \geq 1.$$

Fix $B \in \mathcal{B}$, with $B \subset [1, \infty)$ and $P(X_1 \in B) > 0$, and define $C = \{-x : x \in B\}$. Since $f_B = f_C$, then $\mu(B) = \mu(C)$ a.s. Further, $\mu(B)$ has an absolutely continuous distribution, for f_B is differentiable and $f_B' \neq 0$. Nevertheless, one between $T_n(B)$ and $T_n(C)$ does not converge to 0 in probability. Define in fact $g = I_B - I_C$ and $R_n = n^{-1/2} \sum_{i=1}^n g(X_i)$. Since $\mu(g) = \mu(B) - \mu(C) = 0$ a.s., then R_n converges stably to the kernel $N(0,\, 2\mu(B))$; see [3, Theorem 3.1]. On the other hand, since $E\big\{g(X_{n+1}) \mid \mathcal{G}_n\big\} = E\big\{\mu(g) \mid \mathcal{G}_n\big\} = 0$ a.s., one obtains

$$R_n = \sqrt{n}\,\big\{\mu_n(B) - \mu_n(C)\big\} = T_n(C) - T_n(B) +$$
$$+\sqrt{n}\,\Big\{\mu_n(B) - P\big\{X_{n+1} \in B \mid \mathcal{G}_n^B\big\}\Big\} - \sqrt{n}\,\Big\{\mu_n(C) - P\big\{X_{n+1} \in C \mid \mathcal{G}_n^C\big\}\Big\}.$$

Hence, if $T_n(B) \xrightarrow{P} 0$ and $T_n(C) \xrightarrow{P} 0$, Corollary 1 (applied with $\mathcal{D} = \{B, C\}$) implies the contradiction $R_n \xrightarrow{P} 0$.

References

1. Berti P, Rigo P (1997) A Glivenko-Cantelli theorem for exchangeable random variables. Stat Probab Lett 32:385–391
2. Berti P, Mattei A, Rigo P (2002) Uniform convergence of empirical and predictive measures. Atti Sem Mat Fis Univ Modena 50:465–477
3. Berti P, Pratelli L, Rigo P (2004) Limit theorems for a class of identically distributed random variables. Ann Probab 32:2029–2052
4. Berti P, Crimaldi I, Pratelli L, Rigo P (2009) Rate of convergence of predictive distributions for dependent data. Bernoulli 15:1351–1367
5. Berti P, Pratelli L, Rigo P (2012) Limit theorems for empirical processes based on dependent data. Electron J Probab 17:1–18
6. Berti P, Pratelli L, Rigo P (2013) Exchangeable sequences driven by an absolutely continuous random measure. Ann Probab 41:2090–2102
7. Blackwell D, Dubins LE (1962) Merging of opinions with increasing information. Ann Math Stat 33:882–886
8. Cifarelli DM, Regazzini E (1996) De Finetti's contribution to probability and statistics. Stat Sci 11:253–282
9. Cifarelli DM, Dolera E, Regazzini E (2016) Frequentistic approximations to Bayesian prevision of exchangeable random elements. arXiv:1602.01269v1
10. Fortini S, Ladelli L, Regazzini E (2000) Exchangeability, predictive distributions and parametric models. Sankhya A 62:86–109
11. van der Vaart A, Wellner JA (1996) Weak convergence and empirical processes. Springer