ASYMPTOTICS OF CERTAIN CONDITIONALLY IDENTICALLY DISTRIBUTED SEQUENCES

PATRIZIA BERTI, EMANUELA DREASSI, LUCA PRATELLI, AND PIETRO RIGO

Abstract. Let S be a Borel subset of a Polish space and $(X_n : n \geq 1)$ a sequence of S-valued random variables. Fix a Borel probability measure σ_0 on S, a constant $q_0 \in [0,1]$ and a measurable function $q_i : S^i \to [0,1]$ for each $i \geq 1$. Suppose $X_1 \sim \sigma_0$ and

$$P\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big) = \sigma_0 \prod_{i=1}^{n} Q_i + \delta_{X_n}\,(1 - Q_n) + \sum_{i=1}^{n-1} \delta_{X_i}\,(1 - Q_i) \prod_{j=i+1}^{n} Q_j$$

where $Q_i = q_{i-1}(X_1, \ldots, X_{i-1})$. Sequences of this type, introduced in [10], are conditionally identically distributed and play a role in Bayesian predictive inference. This paper deals with the asymptotics of (X_n). As expected, (X_n) exhibits different behaviors depending on the Q_i. For instance, (X_n) converges a.s. if $\alpha \leq Q_i \leq \beta$ a.s. for all i, where $0 < \alpha \leq \beta < 1$ are constants, while (X_n) does not converge even in probability if σ_0 is nondegenerate, $Q_i > 0$ for all i and $\sum_i (1 - Q_i) < \infty$ a.s. A stable CLT for (X_n) is proved as well.

1. Introduction

Throughout, S is a Borel subset of a Polish space, $\mathcal{B}$ the Borel σ-field on S, and $\mathcal{P}$ the collection of all probability measures on $\mathcal{B}$. Moreover, X_n is the n-th coordinate projection on S^∞, i.e.

$$X_n(s_1, \ldots, s_n, \ldots) = s_n$$

for each $n \geq 1$ and each $(s_1, \ldots, s_n, \ldots) \in S^\infty$.

Following Dubins and Savage [14], a *strategy* is a sequence $\sigma = (\sigma_0, \sigma_1, \ldots)$ such that

- $\sigma_0 \in \mathcal{P}$ and $\sigma_n = \{\sigma_n(x) : x \in S^n\}$ is a collection of elements of $\mathcal{P}$;
- The map $x \mapsto \sigma_n(x)(B)$ is $\mathcal{B}^n$-measurable for fixed $n \geq 1$ and $B \in \mathcal{B}$.

Here, σ_0 should be regarded as the marginal distribution of X_1 and $\sigma_n(x)$ as the conditional distribution of X_{n+1} given that $(X_1, \ldots, X_n) = x$. The probabilities σ_0 and $\sigma_n(x)$ are also called the *predictive distributions* of the sequence (X_n).

According to the Ionescu-Tulcea theorem, for any strategy σ, there is a unique probability measure P on $(S^\infty, \mathcal{B}^\infty)$ satisfying

$$P(X_1 \in \cdot) = \sigma_0 \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid (X_1, \ldots, X_n) = x\big) = \sigma_n(x)$$
$$\text{for all } n \geq 1 \text{ and } P\text{-almost all } x \in S^n.$$

Such a P is denoted P_σ in the sequel.

2010 *Mathematics Subject Classification.* 62F15, 62M20, 60G25, 60G09.

Key words and phrases. Bayesian nonparametrics, Conditional identity in distribution, Exchangeability, Limit theorem, Predictive distribution, Predictive inference.

The Ionescu-Tulcea theorem plays a role in Bayesian predictive inference. In fact, in a Bayesian framework, to make predictions on the sequence (X_n) the inferrer needs to select a strategy σ. At each time $n \geq 1$, having observed $(X_1, \ldots, X_n) = x$, the next observation X_{n+1} is predicted through the predictive distribution $\sigma_n(x)$. This procedure makes sense, for *any* strategy σ, because of the Ionescu-Tulcea theorem.

1.1. **Standard and non-standard approach for exchangeable data.** Usually, (X_n) is requested to be exchangeable. Under this assumption, the standard approach to obtain σ is quite involved. Indeed, to get σ, the inferrer should:

(i) Select a prior π, namely, a probability measure on $\mathcal{P}$;

(ii) Calculate the posterior of π given that $(X_1, \ldots, X_n) = x$, say $\pi_n(x)$;

(iii) Evaluate σ as

$$\sigma_0(B) = \int_{\mathcal{P}} p(B)\, \pi(dp) \quad \text{and} \quad \sigma_n(x)(B) = \int_{\mathcal{P}} p(B)\, \pi_n(x)(dp) \quad \text{for all } B \in \mathcal{B}.$$

Steps (i)-(ii) are troublesome. To assess a prior π is clearly hard. But even when π is selected, to evaluate the posterior π_n may be not straightforward. Frequently, π_n can not be written in closed form but only approximated numerically.

A non-standard approach (henceforth, NSA) is to assign σ_n directly, without passing through π and π_n. Merely, instead of choosing π and then evaluating π_n and σ_n, the inferrer just selects his/her predictive distribution σ_n. As noted above, this procedure makes sense because of the Ionescu-Tulcea theorem. See [3], [6], [10], [12], [13], [14], [15], [18], [19]; see also [16], [21], [22], [23] and references therein.

NSA is in line with de Finetti, Dubins and Savage, among others. Recently, NSA has been adopted in [18] to obtain a fast online Bayesian prediction via copulas. In addition, NSA is quite implicit in most of the machine learning literature. From our point of view, NSA has essentially two merits. Firstly, it requires to place probabilities on *observable facts* only. The value of the next observation X_{n+1} is actually observable, while π and π_n (being probabilities on $\mathcal{P}$) do not deal with observable facts. Secondly, NSA is much more direct than the standard approach. In fact, if the main goal is to predict future observations, why to select the prior π explicitly ? Rather than wondering about π, it looks reasonable to reflect on how the next observation X_{n+1} is affected by $(X_1, \ldots, X_n)$.

However, if (X_n) is requested to be exchangeable, NSA has a gap. Given an arbitrary strategy σ, the Ionescu-Tulcea theorem does not grant exchangeability of (X_n) under P_σ. Therefore, for NSA to apply, one should first characterize those strategies σ which make (X_n) exchangeable under P_σ. A nice characterization is [15, Theorem 3.1]. However, the conditions on σ for making (X_n) exchangeable are quite hard to be checked in real problems. This is the main reason for NSA has not developed so far.

1.2. **Predictive inference with conditionally identically distributed data.** To bypass the gap mentioned in the above paragraph, the exchangeability assumption could be weakened. One option is to request (X_n) to be *conditionally identically* *distributed* (c.i.d.), namely

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

where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}_0$ is the trivial σ-field.

Roughly speaking, condition (1) means that, at each time $n \geq 0$, the future observations $(X_k : k > n)$ are identically distributed given the past $\mathcal{F}_n$. Condition (1) is actually weaker than exchangeability. Indeed, (X_n) is exchangeable if and only if is stationary and c.i.d.

We refer to Subsection 2.1 for the essentials of c.i.d. sequences. Here, we just mention three reasons for taking c.i.d. data into account.

- (j) It is not hard to characterize the strategies σ which make (X_n) c.i.d. under P_σ; see Theorem 1. Therefore, unlike the exchangeable case, NSA can be easily implemented.
- (jj) C.i.d. sequences behave asymptotically much in the same way as exchangeable ones; see Subsection 2.1.
- (jjj) A number of meaningful strategies can not be used if (X_n) is requested to be exchangeable, but are available if (X_n) is only asked to be c.i.d. A trivial example is the strategy (3) reported below. Various other examples are in [1], [2] and [10].

Motivated by (j)-(jjj), in [10], a few strategies σ which makes (X_n) c.i.d. are introduced. One of such strategies is the following.

Fix $\sigma_0 \in \mathcal{P}$, a constant $q_0 \in [0,1]$ and the measurable functions $q_n : S^n \to [0,1]$. For all $n \geq 1$ and $x = (x_1, \ldots, x_n) \in S^n$, define

$$\sigma_n(x) = \sigma_0 \prod_{i=0}^{n-1} q_i + \delta_{x_n}(1 - q_{n-1}) + \sum_{i=1}^{n-1} \delta_{x_i}(1 - q_{i-1}) \prod_{j=i}^{n-1} q_j \tag{2}$$

where δ_{x_i} is the unit mass at x_i and q_i is a shorthand notation to denote

$$q_i = q_i(x_1, \ldots, x_i).$$

Then, (X_n) is c.i.d. under P_σ. Further, σ satisfies the recursive equation

$$\sigma_{n+1}(x, y) = q_n(x)\, \sigma_n(x) + \big\{1 - q_n(x)\big\}\, \delta_y$$

for all $n \geq 0$, $x \in S^n$ and $y \in S$. Thus, when a new observation y becomes available, $\sigma_{n+1}(x, y)$ can be obtained by a simple recursive update of $\sigma_n(x)$.

The strategy (2) is connected to Beta-GOS processes, as meant in [1], and is analogous to formula (10) of [18]. Note also that, if σ_0 vanishes on singletons, the q_i have the following interpretation. Let $x = (x_1, \ldots, x_n)$. Since $\sigma_0\big(\{x_1, \ldots, x_n\}\big) = 0$ and $\delta_{x_i}\big(\{x_1, \ldots, x_n\}\big) = 1$ for $i \leq n$, it follows that

$$P_\sigma\Big(X_{n+1} = X_i \text{ for some } i \leq n \mid (X_1, \ldots, X_n) = x\Big) = \sigma_n(x)\big(\{x_1, \ldots, x_n\}\big)$$
$$= (1 - q_{n-1}) + \sum_{i=1}^{n-1} (1 - q_{i-1}) \prod_{j=i}^{n-1} q_j = 1 - \prod_{i=0}^{n-1} q_i.$$

More importantly, choosing q_i suitably, various real situations can be modeled by σ. As an example, if $q \in (0,1)$ is a constant and $q_i = q$ for each $i \geq 0$, one obtains

$$\sigma_n(x) = q^n \sigma_0 + (1-q) \sum_{i=1}^{n} q^{n-i} \delta_{x_i}; \tag{3}$$

see also [1] and [2]. Roughly speaking, this choice of σ makes sense when the inferrer has only vague opinions on the dependence structure of the data, and yet he/she feels that the weight of the i-th observation x_i should be a decreasing function of $n-i$. In this case, $\sigma_n(x)$ is not invariant under permutations of x, so that (X_n) fails to be exchangeable under P_σ.

As another example, take a constant $c>0$ and define $q_i = \frac{i+c}{i+1+c}$. Then, formula (2) yields the predictive distributions of a Dirichlet sequence, i.e.

$$\sigma_n(x) = \frac{c\,\sigma_0 + \sum_{i=1}^n \delta_{x_i}}{n+c}.$$

In the above two examples, q_i does not depend on $(x_1, \ldots, x_i)$. Clearly, much more elaborated strategies can be obtained if q_i actually depends on $(x_1, \ldots, x_i)$. We refer to [10] for examples of this type, including generalized Polya urns and species sampling sequences.

1.3. **Main results.** If a strategy σ is used to make predictions, a meaningful information is the asymptotic behavior of the data sequence (X_n) under P_σ. This paper investigates the asymptotics of (X_n) under P_σ when σ is given by (2). Our main results, formally stated in Section 3, are a strong limit theorem and a stable CLT. Here, we briefly sketch such results.

Consider the probability space $(S^\infty, \mathcal{B}^\infty, P_\sigma)$, where σ is given by (2), and define

$$Q_n = q_{n-1}(X_1, \ldots, X_{n-1}).$$

The strong limit theorem is

- X_n converges a.s. whenever $\alpha \le Q_n \le \beta$ a.s. for all n, where $0 < \alpha \le \beta < 1$ are constants;
- X_n does not converge even in probability whenever σ_0 is nondegenerate, $Q_n > 0$ for all n and $\sum_n (1-Q_n) < \infty$ a.s.

Thus, it may be that X_n is non-trivial and yet it converges a.s. This is a big difference with respect to the exchangeable case. In fact, an exchangeable sequence Y_n converges in probability if and only if $Y_n = Y_1$ a.s. for each n.

Let us turn to the stable CLT. We first recall that stable convergence is a strong form of convergence in distribution; see Subsection 2.2. In particular, stable convergence implies convergence in distribution.

For definiteness, suppose $S=[a,b]$ is a bounded interval of the real line. (Otherwise, as in Section 3, it suffices to replace X_n with $f(X_n)$ where $f: S \to \mathbb{R}$ is a bounded measurable function). Since (X_n) and (X_n^2) are both c.i.d. under P_σ,

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} V \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{a.s.} V^*$$

for some random variables V and V^*; see Subsection 2.1. Our CLT is

- If $\sum_n \{1 - E(Q_n)\} < \infty$ and $Q_n \le Q_{n+1}$ a.s. for all n, then
$$\sqrt{n}\,\big(\overline{X}_n - V\big) \to \mathcal{N}(0, L) \quad \text{stably}$$
where
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad L = V^* - V^2 \overset{a.s.}{=} \lim_n \frac{1}{n}\sum_{i=1}^n \big\{X_i - \overline{X}_n\big\}^2.$$

In a Bayesian framework, the limit V of the sample means can be seen as a random parameter and the above CLT is useful to make inference on V. In particular, it allows to build (approximate) credible intervals for V.

Finally, under the same assumptions of the previous CLT, it is also shown that
$$\sqrt{n}\,\Big\{\overline{X}_n - E\big(X_{n+1} \mid X_1, \ldots, X_n\big)\Big\} \to \mathcal{N}(0, L) \quad \text{stably.}$$

2. Preliminaries

From now on, $(\Omega, \mathcal{A}, P)$ is a probability space, $(Y_n : n \ge 1)$ a sequence of S-valued random variables on $(\Omega, \mathcal{A}, P)$, and
$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n).$$

2.1. **Conditionally identically distributed random variables.** C.i.d. sequences have been introduced in [4] and [20] and then investigated in various papers; see e.g. [1], [2], [6], [8], [9], [10], [11], [17]. Here, we just recall a few basic facts.

Let $(\mathcal{G}_n : n \ge 0)$ be a filtration on $(\Omega, \mathcal{A}, P)$. Then, (Y_n) is c.i.d. with respect to $(\mathcal{G}_n)$ if is adapted to $(\mathcal{G}_n)$ and
$$P\big(Y_k \in \cdot \mid \mathcal{G}_n\big) = P\big(Y_{n+1} \in \cdot \mid \mathcal{G}_n\big) \quad \text{a.s. for all } k > n \ge 0.$$

If $\mathcal{G}_n = \mathcal{F}_n$, the filtration is not mentioned at all and (Y_n) is just called c.i.d. In this case, by a result in [20], (Y_n) is exchangeable if and only if is stationary and c.i.d.

Asymptotically, a c.i.d. sequence (Y_n) looks like an exchangeable one. We support this claim by three facts.

First, (Y_n) is asymptotically exchangeable, in the sense that
$$(Y_n, Y_{n+1}, \ldots) \to (Z_1, Z_2, \ldots) \text{ in distribution, as } n \to \infty,$$
where $(Z_1, Z_2, \ldots)$ is an exchangeable sequence.

Second, for each bounded measurable function $f : S \to \mathbb{R}$, one obtains
$$\frac{1}{n}\sum_{i=1}^n f(Y_i) \overset{a.s.}{\longrightarrow} V \quad \text{and} \quad E\big\{f(Y_{n+1}) \mid \mathcal{F}_n\big\} \overset{a.s.}{\longrightarrow} V$$
for some real random variable V.

To state the third fact, let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ be the empirical measure. Then, there is a random probability measure μ on $(S, \mathcal{B})$ satisfying
$$\mu_n(B) \overset{a.s.}{\longrightarrow} \mu(B) \quad \text{as } n \to \infty \text{ for every fixed } B \in \mathcal{B}.$$

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

$$E\{\mu(B) \mid \mathcal{F}_n\} = \lim_m E\{\mu_m(B) \mid \mathcal{F}_n\}$$
$$= \lim_m \frac{1}{m} \sum_{i=n+1}^{m} P\big(Y_i \in B \mid \mathcal{F}_n\big) = P\big(Y_{n+1} \in B \mid \mathcal{F}_n\big) \text{ a.s.}$$

Thus, as in the exchangeable case, the predictive distribution $P\big(Y_{n+1} \in \cdot \mid \mathcal{F}_n\big)$ can be written as $E\{\mu(\cdot) \mid \mathcal{F}_n\}$, where μ is the a.s. weak limit of the empirical measures μ_n.

Finally, to complete claim (j) of Subsection 1.2, we report a characterization of c.i.d. sequences in terms of strategies.

Theorem 1. ([8, Theorem 3.1]). *Let σ be a strategy. Then, (X_n) is c.i.d. under P_σ if and only if*

$$\sigma_0(B) = \int \sigma_1(y)(B)\,\sigma_0(dy) \quad \text{and} \quad \sigma_n(x)(B) = \int \sigma_{n+1}(x,y)(B)\,\sigma_n(x)(dy)$$

for all $B \in \mathcal{B}$, all $n \geq 1$ and P_σ-almost all $x \in S^n$.

2.2. **Stable convergence.** Stable convergence is a strong form of convergence in distribution. In a sense, it is intermediate between the latter and convergence in probability.

A *kernel* on S (or a *random probability measure* on S) is a map $K : \Omega \to \mathcal{P}$ such that $\omega \mapsto K(\omega)(B)$ is $\mathcal{A}$-measurable for fixed $B \in \mathcal{B}$. Say that Y_n *converges stably to* K, where K is a kernel on S, if

$$P\big(Y_n \in \cdot \mid H\big) \to E\big(K(\cdot) \mid H\big) \quad \text{weakly}$$
$$\text{for all } H \in \mathcal{A} \text{ with } P(H) > 0.$$

In particular, if $Y_n \to K$ stably, then Y_n converges in distribution to the probability measure $E\big(K(\cdot)\big)$ (just let $H = \Omega$). Further, given any random variable $Y : \Omega \to S$, it is not hard to see that $Y_n \xrightarrow{P} Y$ if and only if Y_n converges stably to the kernel $K = \delta_Y$.

Let $\mathcal{N}(0,b)$ denote the one-dimensional Gaussian law with mean 0 and variance $b \geq 0$ (where $\mathcal{N}(0,0) = \delta_0$). Then, $\mathcal{N}(0,L)$ is a kernel on $\mathbb{R}$ provided L is a real non-negative random variable on $(\Omega, \mathcal{A}, P)$. The next corollary provides conditions for stable convergence toward a kernel of this type. It is a straightforward consequence of [7, Theorem 1].

Corollary 2. *Fix a bounded measurable function $f : S \to \mathbb{R}$ and define*

$$M_n = \frac{1}{n} \sum_{i=1}^{n} f(Y_i) \quad \text{and} \quad Z_n = E\{f(Y_{n+1}) \mid \mathcal{F}_n\}.$$

Suppose (Y_n) c.i.d. and denote by V the a.s. limit of M_n (or, equivalently, the a.s. limit of Z_n). Suppose also that

- **(a)** $\frac{1}{\sqrt{n}} E\{\max_{1 \leq k \leq n} k\,|Z_{k-1} - Z_k|\} \longrightarrow 0,$
- **(b)** $\frac{1}{n} \sum_{k=1}^{n} \{f(Y_k) - Z_{k-1} + k(Z_{k-1} - Z_k)\}^2 \xrightarrow{P} F,$

(c) $\sqrt{n}\,E\{\sup_{k\geq n}|Z_{k-1}-Z_k|\}\longrightarrow 0$,

(d) $n\sum_{k\geq n}(Z_{k-1}-Z_k)^2\overset{P}{\longrightarrow}G$,

where F and G are real nonnegative random variables. Then,

$$\sqrt{n}\,(M_n-Z_n)\to\mathcal{N}(0,F)\quad\text{stably}\quad\text{and}$$
$$\sqrt{n}\,(M_n-V)\to\mathcal{N}(0,F+G)\quad\text{stably.}$$

Proof. Just note that the sequence $(f(Y_n)^2)$ is uniformly integrable (for f is bounded) and

$$E(Z_{n+1}\mid\mathcal{F}_n)=E\{f(Y_{n+2})\mid\mathcal{F}_n\}=E\{f(Y_{n+1})\mid\mathcal{F}_n\}=Z_n\quad\text{a.s.}$$

since (Y_n) is c.i.d. Hence, it suffices to apply [7, Theorem 1]. □

3. Results

We begin with introducing a sequence $(Y_n : n\geq 1)$ of S-valued random variables whose predictive distributions agree with (2).

Fix $\sigma_0\in\mathcal{P}$, a constant $q_0\in[0,1]$ and the measurable functions $q_n : S^n\to[0,1]$, $n\geq 1$. Moreover, on some probability space $(\Omega,\mathcal{A},P)$, take random variables $(T_n : n\geq 1)$ and $(U_{i,j} : j\in\mathbb{N},\ 1\leq i\leq j)$ such that

- (T_n) is an i.i.d. sequence of S-valued random variables with $T_1\sim\sigma_0$;
- $(U_{i,j})$ is an i.i.d. array of $[0,1]$-valued random variables with $U_{1,1}$ uniformly distributed on $[0,1]$;
- (T_n) is independent of $(U_{i,j})$.

Next, define (Y_n) as follows. Let $Y_1=T_1$. At step 2, let $Q_1=q_0$ and define $Y_2=T_2$ or $Y_2=Y_1$ according to whether $U_{1,1}\leq Q_1$ or $U_{1,1}>Q_1$. At step $n+1$, after $Y_1,\ldots,Y_n$ have been defined, let

$$Q_{i+1}=q_i(Y_1,\ldots,Y_i)\qquad\text{for } i=0,\ldots,n-1$$

and then define

$$Y_{n+1}=T_{n+1}\quad\text{if}\quad U_{i,n}\leq Q_i\quad\text{for all } i,$$
$$Y_{n+1}=Y_i\quad\text{if}\quad U_{i,n}>Q_i\text{ and }U_{j,n}\leq Q_j\quad\text{for some } i\text{ and all } j>i.$$

The predictive distributions of (Y_n) are actually given by (2). Recall that

$$\mathcal{F}_0=\{\emptyset,\Omega\}\quad\text{and}\quad\mathcal{F}_n=\sigma(Y_1,\ldots,Y_n).$$

Lemma 3. *$Y_1\sim\sigma_0$ and*

$$P\big(Y_{n+1}\in\cdot\mid\mathcal{F}_n\big)=\sigma_0\prod_{i=1}^n Q_i+\delta_{Y_n}(1-Q_n)+\sum_{i=1}^{n-1}\delta_{Y_i}(1-Q_i)\prod_{j=i+1}^n Q_j$$

a.s. for each $n\geq 1$.

Proof. It is clear that $Y_1=T_1\sim\sigma_0$. Fix $n\geq 1$, $B\in\mathcal{B}$, and let

$$\mathcal{G}_n=\sigma(Y_1,\ldots,Y_n,U_{1,n},\ldots,U_{n,n}),\quad A_n=\{U_{i,n}\leq Q_i\text{ for all } i\}.$$

Since $A_n \in \mathcal{G}_n$ and T_{n+1} is independent of $\mathcal{G}_n$,

$$P\big(A_n \cap \{Y_{n+1} \in B\} \mid \mathcal{F}_n\big) = E\Big(1_{A_n} P\big(T_{n+1} \in B \mid \mathcal{G}_n\big) \mid \mathcal{F}_n\Big)$$
$$= \sigma_0(B)\, P(A_n \mid \mathcal{F}_n) = \sigma_0(B) \prod_{i=1}^n Q_i \quad \text{a.s.}$$

Similarly,

$$P\big(U_{n,n} > Q_n,\, Y_{n+1} \in B \mid \mathcal{F}_n\big) = 1_B(Y_n)\, P\big(U_{n,n} > Q_n \mid \mathcal{F}_n\big) = \delta_{Y_n}(B)\,(1 - Q_n) \quad \text{a.s.}$$

Finally, if $i < n$ and $A_{i,n} = \{U_{i,n} > Q_i \text{ and } U_{j,n} \le Q_j \text{ for } j = i+1, \ldots, n\}$, one obtains

$$P\big(A_{i,n} \cap \{Y_{n+1} \in B\} \mid \mathcal{F}_n\big) = 1_B(Y_i)\, P\big(A_{i,n} \mid \mathcal{F}_n\big)$$
$$= \delta_{Y_i}(B)\,(1 - Q_i) \prod_{j=i+1}^n Q_j \quad \text{a.s.}$$

□

One consequence of Lemma 3 is that

$$P\Big((Y_1, Y_2, \ldots) \in \cdot\Big) = P_\sigma\Big((X_1, X_2, \ldots) \in \cdot\Big)$$

where the strategy σ is given by (2). Since (X_n) is c.i.d. under P_σ (by [10]) it follows that (Y_n) is c.i.d. as well. More importantly, to fix the asymptotic behavior of (X_n) under P_σ, we may work with (Y_n).

Our first result is the following.

Theorem 4. *If $\alpha \le Q_n \le \beta$ a.s. for each n, where $0 < \alpha \le \beta < 1$ are constants, then Y_n converges a.s.*

Proof. Since S is a Borel subset of a Polish space, each probability measure on $\mathcal{B}$ is tight. Hence, by [5, Theorem 2.2], it suffices to show that $f(Y_n)$ converges a.s. for each bounded continuous function $f : S \to \mathbb{R}$.

Fix a bounded continuous $f : S \to \mathbb{R}$ and define $\Delta_m = E\big\{f(Y_{m+1}) \mid \mathcal{F}_m\big\} - f(Y_m)$. Then,

$$\frac{\Delta_{m+1}}{Q_{m+1}} = \frac{E\big\{f(Y_{m+2}) \mid \mathcal{F}_{m+1}\big\} - f(Y_{m+1})}{Q_{m+1}}$$
$$= \int f\, d\sigma_0 \prod_{i=1}^m Q_i + f(Y_m)(1 - Q_m) + \sum_{i=1}^{m-1} f(Y_i)(1 - Q_i) \prod_{j=i+1}^m Q_j - f(Y_{m+1})$$
$$= E\big\{f(Y_{m+1}) \mid \mathcal{F}_m\big\} - f(Y_{m+1}) = \Delta_m + f(Y_m) - f(Y_{m+1}).$$

Summing over $m = 1, \ldots, n$,

$$\Delta_{n+1}/Q_{n+1} + \sum_{m=1}^{n-1} \Delta_{m+1}/Q_{m+1} = \sum_{m=1}^n \Delta_m + f(Y_1) - f(Y_{n+1})$$

or equivalently

$$\sum_{m=2}^n \Delta_m \big(1/Q_m - 1\big) = -\Delta_{n+1}/Q_{n+1} + \Delta_1 + f(Y_1) - f(Y_{n+1}).$$

Next, since (Y_n) is c.i.d. and Q_j is $\mathcal{F}_{j-1}$-measurable, then

$$E\Big\{\Delta_i\,\big(\frac{1}{Q_i}-1\big)\,\Delta_j\,\big(\frac{1}{Q_j}-1\big)\Big\}$$
$$= E\Big\{\Delta_i\,\big(\frac{1}{Q_i}-1\big)\,\big(\frac{1}{Q_j}-1\big)\,E(\Delta_j \mid \mathcal{F}_{j-1})\Big\} = 0 \quad \text{for all } i<j.$$

Therefore,

$$E\Big\{\Big(\sum_{m=2}^n \Delta_m\,\big(\frac{1}{Q_m}-1\big)\Big)^2\Big\} = \sum_{m=2}^n E\Big\{\Delta_m^2\,\big(\frac{1}{Q_m}-1\big)^2\Big\}.$$

Further, since $\alpha \le Q_m \le \beta$ a.s., one obtains

$$\begin{aligned}\frac{(1-\beta)^2}{\beta^2}\sum_{m=2}^n E(\Delta_m^2) &\le \sum_{m=2}^n E\Big\{\Delta_m^2\,\big(\frac{1}{Q_m}-1\big)^2\Big\}\\ &= E\Big\{\Big(\sum_{m=2}^n \Delta_m\,\big(\frac{1}{Q_m}-1\big)\Big)^2\Big\}\\ &= E\Big\{\Big(-\frac{\Delta_{n+1}}{Q_{n+1}} + \Delta_1 + f(Y_1) - f(Y_{n+1})\Big)^2\Big\}\\ &\le \Big(\frac{2\,\sup|f|}{\alpha} + 4\,\sup|f|\Big)^2.\end{aligned}$$

Hence, $E\{\sum_{n=2}^\infty \Delta_n^2\} = \sum_{n=2}^\infty E(\Delta_n^2) < \infty$, so that $\Delta_n \xrightarrow{a.s.} 0$. To conclude the proof, just recall that $E\{f(Y_{n+1}) \mid \mathcal{F}_n\} \xrightarrow{a.s.} V$ for some real random variable V; see Subsection 2.1. Therefore, $f(Y_n) \xrightarrow{a.s.} V$. □

Incidentally we note that, as apparent from the previous proof, the assumption $Q_n \ge \alpha$ a.s. for all n can be weakened into $\liminf_n E(Q_n^{-2}) < \infty$. We also note that, when the strategy σ is given by (3) (i.e., when $Q_n = q$ for all n and some constant $0<q<1$) Theorem 4 implies that Y_n converges a.s.

In Theorem 4, the Q_n are separated from 0 and 1 and Y_n converges a.s. Things change drastically if Q_n approaches 1 quickly enough.

Theorem 5. *Y_n does not converge in probability provided σ_0 is nondegenerate, $Q_n > 0$ for all n and $\sum_n(1-Q_n) < \infty$ a.s.*

Proof. Let d be the distance on S. It suffices to show that $d(Y_n, Y_{n+1})$ does not converge to 0 in probability. Since σ_0 is nondegenerate, there is $\epsilon > 0$ such that $P\big(d(T_1,T_2) > \epsilon\big)$ is strictly positive. Define

$$H_n = \{U_{i,n} \le Q_i \text{ for each } i \le n \text{ and } U_{i,n-1} \le Q_i \text{ for each } i < n\}.$$

Since $(Q_1,\ldots,Q_n)$ is a function of $(Y_1,\ldots,Y_{n-1})$, then (T_n, T_{n+1}) is independent of H_n. Hence,

$$\begin{aligned}P\Big(d(Y_n,Y_{n+1}) > \epsilon\Big) &\ge P\Big(H_n \cap \{d(T_n,T_{n+1}) > \epsilon\}\Big) = P\big(d(T_1,T_2) > \epsilon\big)\,P(H_n)\\ &= P\big(d(T_1,T_2) > \epsilon\big)\,E\Big\{\prod_{i=1}^n Q_i\,\prod_{i=1}^{n-1} Q_i\Big\}.\end{aligned}$$

Finally, $Q_n > 0$ for all n and $\sum_n (1 - Q_n) < \infty$ a.s. implies that $\prod_{i=1}^n Q_i \overset{a.s.}{\longrightarrow} Q$, where Q is a random variable such that $Q > 0$ a.s. Therefore,

$$\liminf_n P\Big(d(Y_n, Y_{n+1}) > \epsilon\Big) \geq P\big(d(T_1, T_2) > \epsilon\big)\, E(Q^2) > 0.$$

□

We finally turn to the CLT. Fix a bounded measurable function $f : S \to \mathbb{R}$ and define

$$M_n = \frac{1}{n} \sum_{i=1}^n f(Y_i) \quad \text{and} \quad Z_n = E\big\{f(Y_{n+1}) \mid \mathcal{F}_n\big\}.$$

Since (Y_n) is c.i.d., there is a real random variable V such that

$$M_n \overset{a.s.}{\longrightarrow} V \quad \text{and} \quad Z_n \overset{a.s.}{\longrightarrow} V.$$

Our last result deals with

$$C_n = \sqrt{n}\,(M_n - Z_n) \quad \text{and} \quad W_n = \sqrt{n}\,(M_n - V).$$

Indeed, both C_n and W_n are often involved in the CLT for dependent data; see e.g. [4], [6], [7], [8]. Note also that, in the special case where (Y_n) is i.i.d. (namely, when $Q_n = 1$ for all n) one obtains $C_n = W_n = \sqrt{n}\,\big\{M_n - E(f(Y_1))\big\}$.

Theorem 6. *Suppose* $\sum_n \big\{1 - E(Q_n)\big\} < \infty$ *and* $Q_n \leq Q_{n+1}$ *a.s. for all* n*. Then, for each bounded measurable function* $f : S \to \mathbb{R}$*, one obtains*

$$C_n \to \mathcal{N}(0, L) \quad \textit{stably} \quad \textit{and} \quad W_n \to \mathcal{N}(0, L) \quad \textit{stably}$$

where

$$L = V^* - V^2 \quad \textit{with} \quad V^* = \lim_n \frac{1}{n} \sum_{i=1}^n f^2(Y_i) \quad \textit{a.s.}$$

Proof. By Corollary 2, it suffices to prove conditions (a)-(d) with $F = V^* - V^2$ and $G = 0$.

First note that

$$Q_n\, Z_{n-1} = Z_n - f(Y_n)(1 - Q_n) \qquad \text{a.s.}$$

Letting $c = 2\, \sup|f|$, it follows that

$$|Z_n - Z_{n-1}| = (1 - Q_n)\, |f(Y_n) - Z_{n-1}| \leq c\,(1 - Q_n) \qquad \text{a.s.}$$

Since $Q_n \leq Q_{n+1}$ a.s. for all n, one also obtains

$$n\,\big\{1 - E(Q_n)\big\} = j\,\big\{1 - E(Q_n)\big\} + (n - j)\,\big\{1 - E(Q_n)\big\}$$
$$\leq j\,\big\{1 - E(Q_n)\big\} + \sum_{i=j+1}^n \big\{1 - E(Q_i)\big\} \quad \text{for each } j < n.$$

Hence, $\sum_n \big\{1 - E(Q_n)\big\} < \infty$ implies

$$\limsup_n\, n\,\big\{1 - E(Q_n)\big\} = 0.$$

We next prove conditions (a) and (c). As to (c),

$$\sqrt{n}\; E\big\{\sup_{k \geq n} |Z_{k-1} - Z_k|\,\big\} \leq c\,\sqrt{n}\; E\big\{\sup_{k \geq n}\,(1 - Q_k)\big\} = c\,\sqrt{n}\, E(1 - Q_n) \to 0.$$

As to (a), since

$$n\,|Z_{n-1}-Z_n|\le c\,n\,(1-Q_n)\le c\sum_{i=1}^n(1-Q_i)\quad\text{a.s.},$$

one obtains

$$\frac{1}{\sqrt{n}}\,E\big\{\max_{1\le k\le n}k\,|Z_{k-1}-Z_k|\big\}\le\frac{c}{\sqrt{n}}\,E\Big\{\sum_{k=1}^n(1-Q_k)\Big\}\le\frac{c}{\sqrt{n}}\sum_{k=1}^\infty\big\{1-E(Q_k)\big\}\to 0.$$

It remains to prove conditions (b) and (d) with $F=V^*-V^2$ and $G=0$. On the other hand, since

$$\frac{1}{n}\sum_{k=1}^n\big\{f(Y_k)-Z_{k-1}\big\}^2\xrightarrow{a.s.}V^*-V^2,$$

conditions (b) and (d) are actually true with $F=V^*-V^2$ and $G=0$ provided

$$n\,(Z_n-Z_{n-1})\xrightarrow{a.s.}0.$$

Since

$$E\Big\{\sum_n(1-Q_n)\Big\}=\sum_n\big\{1-E(Q_n)\big\}<\infty,$$

then $\sum_n(1-Q_n)<\infty$ a.s. Hence, arguing as above,

$$n\,(1-Q_n)\le j\,(1-Q_n)+\sum_{i=j+1}^\infty(1-Q_i)\quad\text{a.s. for each } j<n.$$

Therefore,

$$\limsup_n\,n\,|Z_n-Z_{n-1}|\le c\,\limsup_n\,n\,(1-Q_n)=0\quad\text{a.s.}$$

and this concludes the proof. □

References

[1] Airoldi E.M., Thiago Costa, Bassetti F., Leisen F., Guindani M. (2014) Generalized species sampling priors with latent beta reinforcements, *J.A.S.A.*, 109, 1466-1480.

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

[3] Berti P., Regazzini E., Rigo P. (1997) Well-calibrated, coherent forecasting systems, *Theory Probab. Appl.*, 42, 82-102.

[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) Almost sure weak convergence of random probability measures, *Stochastics (formerly: Stochastics and Stochastic Reports)*, 78, 91-97.

[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. (2011) A central limit theorem and its applications to multicolor randomly reinforced urns, *J. Appl. Probab.*, 48, 527-546.

[8] Berti P., Pratelli L., Rigo P. (2012) Limit theorems for empirical processes based on dependent data, *Electronic J. Probab.*, 17, 1-18.

[9] Berti P., Pratelli L., Rigo P. (2013) Exchangeable sequences driven by an absolutely continuous random measure, *Ann. Probab.*, 41, 2090-2102.

[10] Berti P., Dreassi E., Pratelli L., Rigo P. (2019) A predictive approach to Bayesian nonparametrics, *submitted*, currently available at: http://local.disia.unifi.it/dreassi/mg.pdf

[11] Cassese A., Zhu W., Guindani M., Vannucci M. (2019) A Bayesian nonparametric spiked process prior for dynamic model selection, *Bayesian Analysis*, 14, 553-572.

[12] Cifarelli D.M., Regazzini E. (1996) De Finetti's contribution to probability and statistics, *Statist. Science*, 11, 253-282.

[13] de Finetti B. (1937) La prevision: ses lois logiques, ses sources subjectives, *Ann. Inst. H. Poincare*, 7, 1-68.

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

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

[16] Fortini S., Petrone S. (2012) Predictive construction of priors in Bayesian nonparametrics, *Brazilian J. Probab. Statist.*, 26, 423-449.

[17] Fortini S., Petrone S., Sporysheva P. (2018) On a notion of partially conditionally identically distributed sequences, *Stoch. Proc. Appl.*, 128, 819-846.

[18] Hahn P.R., Martin R., Walker S.G. (2018) On recursive Bayesian predictive distributions, *J.A.S.A.*, 113, 1085-1093.

[19] Hill B.M. (1993) Parametric models for A_n: splitting processes and mixtures, *J. Royal Stat. Soc.* B, 55, 423-433.

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

[21] Martin R., Tokdar S.T. (2011) Semiparametric inference in mixture models with predictive recursion marginal likelihhod, *Biometrika*, 98, 567-582.

[22] Newton M.A., Zhang Y. (1999) A recursive algorithm for nonparametric analysis with missing data, *Biometrika*, 86, 15-26.

[23] Tokdar S.T., Martin R., Ghosh J.K. (2009) Consistency of a recursive estimate of mixing distributions, *Ann. Statist.*, 37, 2502-2522.

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

Emanuela Dreassi, Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Universita' di Firenze, viale Morgagni 59, 50134 Firenze, Italy
E-mail address: emanuela.dreassi@unifi.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 Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy
E-mail address: pietro.rigo@unipv.it