Statistics and Probability Letters 168 (2021) 108923

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Asymptotics of certain conditionally identically distributed sequences

Patrizia Berti [a], Emanuela Dreassi [b], Luca Pratelli [c], Pietro Rigo [d,*]

[a] *Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy*
[b] *Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Universita' di Firenze, viale Morgagni 59, 50134 Firenze, Italy*
[c] *Accademia Navale, viale Italia 72, 57100 Livorno, Italy*
[d] *Dipartimento di Scienze Statistiche "P. Fortunati", Universita' di Bologna, via delle Belle Arti 41, 40126 Bologna, Italy*

ARTICLE INFO

Article history:
Received 25 February 2020
Received in revised form 23 August 2020
Accepted 28 August 2020
Available online 9 September 2020

MSC:
60G25
60G09
62F15
62M20

Keywords:
Bayesian nonparametrics
Conditional identity in distribution
Exchangeability
Limit theorem
Predictive distribution
Predictive inference

ABSTRACT

The probability distribution of a sequence $X = (X_1, X_2, \dots)$ of random variables is determined by its predictive distributions $P(X_1 \in \cdot)$ and $P\big(X_{n+1} \in \cdot \mid X_1, \dots, X_n\big)$, $n \geq 1$. Motivated by applications in Bayesian predictive inference, in Berti et al. (2020), a class $\mathcal{C}$ of sequences is introduced by specifying such predictive distributions. Each $X \in \mathcal{C}$ is conditionally identically distributed. The asymptotics of $X \in \mathcal{C}$ is investigated in this paper. Both strong and weak limit theorems are provided. Conditions for X to converge a.s., and for X not to converge in probability, are given in terms of the predictive distributions. A stable CLT is provided as well. Such a CLT is used to obtain approximate credible intervals.

1. Introduction

Throughout, $(S, \mathcal{B})$ is a measurable space and X_n the nth coordinate projection on S^∞, namely, $X_n(s) = s_n$ for each $n \geq 1$ and each sequence $s = (s_1, \dots, s_n, \dots) \in S^\infty$. To avoid needless technicalities, S is assumed to be a Borel subset of a Polish space and $\mathcal{B}$ the Borel σ-field on S. Moreover, $\mathcal{P}$ denotes the collection of all probability measures on $\mathcal{B}$.

Following Dubins and Savage (1965), a *strategy* is a sequence $\sigma = (\sigma_0, \sigma_1, \dots)$ 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 $(X_1, \dots, X_n) = x$. The probabilities σ_0 and $\sigma_n(x)$ are also called the *predictive distributions* of the sequence (X_n).

* Corresponding author.
E-mail addresses: patrizia.berti@unimore.it (P. Berti), emanuela.dreassi@unifi.it (E. Dreassi), pratel@mail.dm.unipi.it (L. Pratelli), pietro.rigo@unibo.it (P. Rigo).

https://doi.org/10.1016/j.spl.2020.108923

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

$$P(X_1 \in \cdot) = \sigma_0 \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid (X_1, \dots, X_n) = x\big) = \sigma_n(x)$$

for all $n \geq 1$ and P-almost all $x \in S^n$. Such a P is denoted as P_σ in the sequel.

The Ionescu-Tulcea theorem plays a role in Bayesian predictive inference. In fact, in a Bayesian framework, to make predictions on (X_n) the statistician needs precisely a strategy σ. At each time $n \geq 1$, having observed $(X_1, \dots, X_n) = x$, 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

In Bayesian predictive inference, the data sequence (X_n) is usually assumed to be exchangeable. In that case, there are essentially two procedures for selecting a strategy σ. Following Berti et al. (2020), we call them the *standard approach* (SA) and the *non-standard approach* (NSA). This terminology is adopted just for the sake of clarity and its only possible motivation is that SA is much more popular than NSA.

According to SA, to obtain σ, the statistician should:

(i) Select a prior π, namely, a probability measure on $\mathcal{P}$;
(ii) Calculate the posterior of π given that $(X_1, \dots, X_n) = x$, 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 cannot be written in closed form but only approximated numerically.

On the other hand, SA is not motivated by prediction alone. Another motivation, possibly the main one, is to make inference on other features of the data distribution, such as a mean, a quantile, a correlation, or more generally some random parameter (possibly, infinite dimensional). In all these cases, the posterior π_n is fundamental. In short, SA is a cornerstone of Bayesian inference, but, when prediction is the main target, is possibly quite involved.

On the contrary, NSA entails assigning σ_n directly, without passing through π and π_n. Merely, instead of choosing π and then evaluating π_n and σ_n, the statistician just selects his/her predictive distribution σ_n. As noted above, this procedure makes sense because of the Ionescu-Tulcea theorem. See Berti et al. (1997, 2009, 2020), Cifarelli and Regazzini (1996), de Finetti (1937), Dubins and Savage (1965), Fortini et al. (2000), Fortini and Petrone (2012), Hahn et al. (2018), Hill (1993) and Lee et al. (2013).

NSA is in line with de Finetti, Dubins and Savage, among others. Pitman's work should be mentioned as well; see e.g. Pitman (1996, 2006). Moreover, NSA has been recently adopted in Hahn et al. (2018) to obtain a fast online Bayesian prediction via copulas.

From our point of view, NSA has essentially three 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, when prediction is the main target, NSA is much more direct than SA. In this case, 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 $(X_1, \dots, X_n)$. Thirdly, NSA is especially appealing in a nonparametric framework, where selecting a prior with large support is usually difficult.

1.2. Predictive inference with conditionally identically distributed data

If (X_n) is required to be exchangeable, however, 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 Fortini et al. (2000, Theorem 3.1). However, the conditions on σ for making (X_n) exchangeable are quite hard to be checked in real problems. This is one reason for NSA has not yet been developed. Another reason is the lack of constructive procedures for determining σ. It is precisely this lack which makes SA necessary for prediction, even if analytically more involved.

To bypass the gap mentioned in the above paragraph, the exchangeability assumption could be weakened. One option is to assume (X_n) *conditionally identically distributed* (c.i.d.), namely, $X_2 \sim X_1$ and

$$P\big(X_k \in \cdot \mid X_1, \dots, X_n\big) = P\big(X_{n+1} \in \cdot \mid X_1, \dots, X_n\big) \quad \text{a.s. for all } k > n \geq 1.$$

Roughly speaking, the above condition means that, at each time n, the future observations $(X_k : k > n)$ are identically distributed given the past. Such a condition is actually weaker than exchangeability. Indeed, by a result in Kallenberg (1988), (X_n) is exchangeable if and only if is stationary and c.i.d.

There are at least 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 Berti et al. (2012, Theorem 3.1). 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 Section 2.1.
(jjj) A number of meaningful strategies cannot be used if (X_n) is required to be exchangeable, but are available if (X_n) is only asked to be c.i.d. A trivial example is the strategy (2) reported below. Various other examples are in Airoldi et al. (2014), Bassetti et al. (2010) and Berti et al. (2020).

Motivated by (j)–(jjj), in Berti et al. (2020), a few strategies σ which make (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

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

where q_i is a shorthand notation to denote $q_i = q_i(x_1, \ldots, x_i)$ and δ_{x_i} is the unit mass at x_i.

If σ is the above strategy, (X_n) is c.i.d. under P_σ. Note also that

$$\sigma_{n+1}(x, y) = q_n(x)\,\sigma_n(x) + \big\{1 - q_n(x)\big\}\,\delta_y \qquad \text{for all } n \geq 0,\ x \in S^n \text{ and } y \in S.$$

Thus, when a new observation y becomes available, $\sigma_{n+1}(x, y)$ is just a recursive update of $\sigma_n(x)$. Further, if σ_0 vanishes on singletons,

$$\prod_{i=0}^{n-1} q_i = P_\sigma\Big(X_{n+1} \neq X_i \text{ for all } i \leq n \mid (X_1, \ldots, X_n) = x\Big)$$

and this equation may help to attach the q_i some interpretation.

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

$$\sigma_n(x) = q^n \sigma_0 + (1 - q) \sum_{i=1}^{n} q^{n-i} \delta_{x_i}; \tag{2}$$

see also Airoldi et al. (2014) and Bassetti et al. (2010). Roughly speaking, this choice of σ makes sense when the statistician has only vague opinions on the dependence structure of the data, and yet he/she feels that the weight of the ith 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 (1) yields the predictive distributions of a Dirichlet sequence, i.e.

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

Various other examples of strategies which can be written by (1), including generalized Polya urns and species sampling sequences, are obtained in Berti et al. (2020).

1.3. Main results

Obviously, 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 (1). Our main results, formally stated in Section 3, are a strong limit theorem, a stable CLT and some of its consequences. Here, we briefly sketch such results.

Consider the probability space $(S^\infty, \mathcal{B}^\infty, P_\sigma)$, where σ is given by (1), and define

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

The strong limit theorem is

- X_n converges a.s. whenever $\alpha \leq Q_n \leq \beta$ a.s. for all n, where $0 < \alpha \leq \beta < 1$ are constants;
- X_n does not converge even in probability whenever σ_0 is non-degenerate, $Q_n > 0$ 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. Overall, this greater flexibility of c.i.d. sequences, with respect to the exchangeable ones, may be useful in real problems.

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

Suppose $S = \mathbb{R}$ (so that the X_n are real valued) and $E(X_1^2) = \int t^2\,\sigma_0(dt) < \infty$. Since (X_n) is 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 Section 2.1. Define

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \quad L_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \quad \text{and} \quad L = V^* - V^2 \overset{a.s.}{=} \lim_n L_n.$$

Also, denote by $\mathcal{N}(0, b)$ the Gaussian law with mean 0 and variance $b \geq 0$ (where $\mathcal{N}(0, 0) = \delta_0$). In this notation, the CLT is:

- Suppose $Q_n \leq Q_{n+1}$ a.s. for all n. Then,

  $$\sqrt{n}\left(\overline{X}_n - V\right) \xrightarrow{stably} \mathcal{N}(0, L)$$

  provided (at least) one of the following two conditions is satisfied:
- $\sum_n \|1 - Q_n\| < \infty$, where $\|\cdot\|$ denotes the sup-norm, namely

  $$\|1 - Q_n\| = \sup_{x \in \mathbb{R}^{n-1}} |1 - q_{n-1}(x)|;$$

- $S \subset \mathbb{R}$ is a bounded subset and $\sum_n \{1 - E(Q_n)\} < \infty$.

In a Bayesian framework, the limit V of the sample means can be seen as a random parameter (see Section 2.1) and the above CLT can be used to make inference on V. In particular, it allows to build (approximate) credible intervals for V. In fact, since convergence is stable, one obtains

$$1_{\{L_n > 0\}} \frac{\sqrt{n}\left(\overline{X}_n - V\right)}{\sqrt{L_n}} \xrightarrow{stably} \mathcal{N}(0, 1)$$

whenever $Q_1 > 0$, σ_0 is non-degenerate and the assumptions of the CLT are satisfied. Therefore, for $\alpha \in (0, 1)$ and large n, an approximate $(1 - \alpha)$-credible interval for V is $\overline{X}_n \pm u_{\alpha/2}\sqrt{L_n/n}$ where $u_{\alpha/2}$ is the quantile of order $(1 - \alpha/2)$ of the standard normal distribution. See Example 7 for details.

Finally, under the same assumptions of the CLT, we also prove that

$$\sqrt{n}\left\{\overline{X}_n - E\left(X_{n+1} \mid X_1, \ldots, X_n\right)\right\} \xrightarrow{stably} \mathcal{N}(0, L).$$

2. Preliminaries

From now on, $(\Omega, \mathcal{A}, P)$ is a probability space, $(Y_n : n \geq 1)$ a sequence of S-valued random variables on $(\Omega, \mathcal{A}, P)$, and

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

Moreover, 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 $B \in \mathcal{B}$.

2.1. Conditionally identically distributed random variables

As recalled in Section 1.2, (Y_n) is c.i.d. if

$$P\left(Y_k \in \cdot \mid \mathcal{F}_n\right) = P\left(Y_{n+1} \in \cdot \mid \mathcal{F}_n\right) \quad \text{a.s. for all } k > n \geq 0.$$

C.i.d. sequences have been introduced in Berti et al. (2004) and Kallenberg (1988) and then investigated in various papers; see e.g. Airoldi et al. (2014), Bassetti et al. (2010), Berti et al. (2009, 2012, 2013, 2020), Cassese et al. (2019) and Fortini et al. (2018).

The asymptotic behavior of a c.i.d. sequence (Y_n) is similar to that of an exchangeable one. We support this claim by three facts.

(k) If $S = \mathbb{R}$ (so that the Y_n are real valued) and $E\{|Y_1|\} < \infty$, then

$$\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{a.s.} V \quad \text{and} \quad E\left(Y_{n+1} \mid \mathcal{F}_n\right) \xrightarrow{a.s.} V$$

for some real random variable V. Note that, since (Y_n^2) is still c.i.d., if $E(Y_1^2)<\infty$ there is a random variable V^* such that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \xrightarrow{a.s.} V^* \quad \text{and} \quad E\big(Y_{n+1}^2 \mid \mathcal{F}_n\big) \xrightarrow{a.s.} V^*.$$

(kk) Let $\mu_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i}$ be the empirical measure. There is a kernel μ on S satisfying

$$\mu_n(B) \xrightarrow{a.s.} \mu(B) \quad \text{as } n\to\infty \text{ for every fixed } B\in\mathcal{B}.$$

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

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

In other terms, as in the exchangeable case, the predictive distribution $P\big(Y_{n+1}\in\cdot\mid\mathcal{F}_n\big)$ can be written as $E\big\{\mu(\cdot)\mid\mathcal{F}_n\big\}$.

(kkk) (Y_n) is asymptotically exchangeable, in the sense that

$$(Y_n, Y_{n+1},\ldots)\to(Y_1^*,Y_2^*,\ldots) \text{ in distribution, as } n\to\infty,$$

where (Y_n^*) is an exchangeable sequence. Moreover, (Y_n^*) is directed by the kernel μ of (kk), namely

$$P\big(Y_1^*\in B_1,\ldots,Y_k^*\in B_k\big)=E\Big\{\prod_{i=1}^{k}\mu(B_i)\Big\} \quad \text{for all } k\geq 1 \text{ and } B_1,\ldots,B_k\in\mathcal{B}.$$

The kernel μ is a meaningful random parameter, even if it does not completely determine the probability law of (Y_n); see Berti et al. (2020, Section 8). In fact, $\mu(B)$ is the long run frequency of the events $\{Y_n\in B\}$. On noting that

$$P\big(Y_{n+1}\in B\mid\mathcal{F}_n\big)=E\big\{\mu(B)\mid\mathcal{F}_n\big\}\xrightarrow{a.s.}\mu(B),$$

$\mu(B)$ can be also regarded as the asymptotically optimal predictor of the event {the next observation belongs to B}. In addition, μ is the directing measure of the exchangeable limit sequence (Y_n^*). Therefore, as in the exchangeable case, it may be reasonable to make inference on μ. Similar considerations can be repeated for the random variable V involved in (k). In fact, if $S=\mathbb{R}$ and $E\{|Y_1|\}<\infty$, V agrees with the mean of μ, that is, $V(\omega)=\int t\,\mu(\omega)(dt)$ for almost all $\omega\in\Omega$.

2.2. Stable convergence

Let K be a kernel on S. Say that Y_n *converges stably to* K if

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

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 $H=\Omega$).

In case $S=\mathbb{R}$, a remarkable kernel is $\mathcal{N}(0,L)$ where L is any real non-negative random variable on $(\Omega,\mathcal{A},P)$. (Recall that, for each $b\geq 0$, $\mathcal{N}(0,b)$ denotes the Gaussian law with mean 0 and variance b). The next corollary provides conditions for stable convergence toward a kernel of this type.

Corollary 1. *Let (Y_n) be a c.i.d. sequence of real random variables such that $E(Y_1^2)<\infty$. Define $Z_n=E\big(Y_{n+1}\mid\mathcal{F}_n\big)$, $\overline{Y}_n=\frac{1}{n}\sum_{k=1}^{n}Y_k$ and $V\overset{a.s.}{=}\lim_n\overline{Y}_n$. Suppose that*

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

where F and G are real non-negative random variables. Then,

$$\sqrt{n}\,(\overline{Y}_n-Z_n)\xrightarrow{stably}\mathcal{N}(0,F) \quad \textit{and} \quad \sqrt{n}\,(\overline{Y}_n-V)\xrightarrow{stably}\mathcal{N}(0,F+G).$$

Proof. This is a straightforward consequence of Berti et al. (2011, 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 (1).

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 $(U_{i,j})$.

Using (T_n) and $(U_{i,j})$ as building blocks, 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 $U_{1,1} \le Q_1$ or $U_{1,1} > Q_1$. At step $n+1$, after $Y_1, \ldots, Y_n$ have been defined, let

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

and then define

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

The predictive distributions of such (Y_n) are actually given by (1).

Lemma 2. *$Y_1 \sim \sigma_0$ 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 $n \ge 1$ (recall that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$).

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

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

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

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

One consequence of Lemma 2 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 (1). Since (X_n) is c.i.d. under P_σ (by Berti et al., 2020) 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 3. *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 Berti et al. (2006, 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,

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

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

$$E\Big\{\Delta_i\,\big(\frac{1}{Q_i}-1\big)\,\Delta_j\,\big(\frac{1}{Q_j}-1\big)\Big\}=E\Big\{\Delta_i\,\big(\frac{1}{Q_i}-1\big)\,\big(\frac{1}{Q_j}-1\big)\,E(\Delta_j\mid\mathcal{F}_{j-1})\Big\}=0$$

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

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

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

As an example, Theorem 3 implies that Y_n converges a.s. when the strategy σ is given by (2) (i.e., when $Q_n=q$ for all n and some constant $0<q<1$).

In Theorem 3, 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 4. *Y_n does not converge in probability provided σ_0 is non-degenerate, $Q_n>0$ 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 non-degenerate, 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,

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

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

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

We next turn to the CLT. Suppose $S=\mathbb{R}$ (so that the Y_n are real valued) and $E(Y_1^2)<\infty$. Define $\overline{Y}_n=\frac{1}{n}\sum_{i=1}^{n}Y_i$ and $Z_n=E\big(Y_{n+1}\mid\mathcal{F}_n\big)$. Since (Y_n) is c.i.d., there are random variables V and V^* such that

$$\overline{Y}_n\xrightarrow{a.s.}V,\quad Z_n\xrightarrow{a.s.}V,\quad\frac{1}{n}\sum_{i=1}^{n}Y_i^2\xrightarrow{a.s.}V^*\quad\text{and}\quad E\big(Y_{n+1}^2\mid\mathcal{F}_n\big)\xrightarrow{a.s.}V^*.$$

Some CLT's for dependent data are concerned with

$$C_n=\sqrt{n}\,(\overline{Y}_n-Z_n)\quad\text{and}\quad W_n=\sqrt{n}\,(\overline{Y}_n-V);$$

see e.g. Berti et al. (2004, 2009, 2011, 2012). Note that, in the special case where (Y_n) is i.i.d. (namely, when $Q_n=1$ for all n) one obtains $C_n=W_n=\sqrt{n}\,\big(\overline{Y}_n-E(Y_1)\big)$.

Theorem 5. *Suppose $S=\mathbb{R}$, $E(Y_1^2)<\infty$, and $Q_n\le Q_{n+1}$ a.s. for all n. Define $L=V^*-V^2\overset{a.s.}{=}\lim_n\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y}_n)^2$. Then,*

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

provided (at least) one of the following two conditions is satisfied:

- $\sum_n \|1 - Q_n\| < \infty$, *where* $\|\cdot\|$ *denotes the sup-norm;*
- $S \subset \mathbb{R}$ *is a bounded subset and* $\sum_n \{1 - E(Q_n)\} < \infty$.

Proof. By Corollary 1, it suffices to prove conditions (a)–(d) with $F = L$ and $G = 0$.

First note that $Q_n Z_{n-1} = Z_n - Y_n(1 - Q_n)$ a.s., which in turn implies

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

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

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

Next, suppose $\sum_n \|1 - Q_n\| < \infty$. Fix $j < n$. Since

$$E\{(Y_n - Z_{n-1})^2\} = E(Y_n^2) - E(Z_{n-1}^2) \le E(Y_n^2) = E(Y_1^2),$$

one obtains

$$\begin{aligned}
n^2 E\{(Z_n - Z_{n-1})^2\} &= E\Big\{\Big(n(1 - Q_n)(Y_n - Z_{n-1})\Big)^2\Big\} \\
&\le E\Big\{\Big(j(1 - Q_n) + \sum_{i=j+1}^{n} (1 - Q_i)\Big)^2 (Y_n - Z_{n-1})^2\Big\} \\
&\le \|j(1 - Q_n) + \sum_{i=j+1}^{n} (1 - Q_i)\|^2 E\{(Y_n - Z_{n-1})^2\} \\
&\le E(Y_1^2)\Big\{j\,\|1 - Q_n\| + \sum_{i=j+1}^{n} \|1 - Q_i\|\Big\}^2.
\end{aligned}$$

Therefore,

$$\limsup_n n^2 E\{(Z_n - Z_{n-1})^2\} \le E(Y_1^2)\Big\{\sum_{i=j+1}^{\infty} \|1 - Q_i\|\Big\}^2 \quad \text{for each } j.$$

Thus, $\sum_n \|1 - Q_n\| < \infty$ yields $\lim_n n^2 E\{(Z_n - Z_{n-1})^2\} = 0$, which in turn implies

$$\begin{aligned}
E\Big\{n \sum_{k\ge n} (Z_{k-1} - Z_k)^2\Big\} &= n \sum_{k\ge n} E\{(Z_k - Z_{k-1})^2\} \\
&= n \sum_{k\ge n} \frac{1}{k^2} k^2 E\{(Z_k - Z_{k-1})^2\} \longrightarrow 0.
\end{aligned}$$

Hence, condition (d) holds with $G = 0$. Conditions (a) and (c) can be checked in a similar way. In fact,

$$\begin{aligned}
\frac{1}{n} E\Big\{\max_{1\le k\le n} k\,|Z_{k-1} - Z_k|\Big\}^2 &\le \frac{1}{n} E\Big\{\max_{1\le k\le n} k^2 (Z_{k-1} - Z_k)^2\Big\} \\
&\le \frac{1}{n} \sum_{k=1}^{n} k^2 E\{(Z_{k-1} - Z_k)^2\} \longrightarrow 0 \qquad \text{and}
\end{aligned}$$

$$n\,E\Big\{\sup_{k\ge n} |Z_{k-1} - Z_k|\Big\}^2 \le n\,E\Big\{\sup_{k\ge n} (Z_{k-1} - Z_k)^2\Big\} \le E\Big\{n \sum_{k\ge n} (Z_{k-1} - Z_k)^2\Big\} \longrightarrow 0.$$

As to condition (b), first note that

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

Hence, (b) holds with $F = L$ since $\lim_n n^2 E\{(Z_n - Z_{n-1})^2\} = 0$.

This concludes the proof if $\sum_n \|1 - Q_n\| < \infty$. If $S \subset \mathbb{R}$ is bounded, say $S \subset [a, b]$, essentially the same argument applies. But in this case, since

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

the condition $\sum_n \|1 - Q_n\| < \infty$ can be weakened into $\sum_n \{1 - E(Q_n)\} < \infty$. More precisely, arguing as above, $\sum_n \{1 - E(Q_n)\} < \infty$ implies $\lim_n n\{1 - E(Q_n)\} = 0$ and $\lim_n n(Z_n - Z_{n-1}) \overset{a.s.}{=} 0$, which in turn imply conditions (a)–(d) with $F = L$ and $G = 0$. □

From a Bayesian point of view, the limit V of the sample means can be seen as a random parameter (see Section 2.1) and Theorem 5 may be exploited to make inference on V. To this end, we first highlight a consequence of Theorem 5.

Theorem 6. *Let $L_n = \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$ and $R_n = 1_{\{L_n>0\}} \frac{\sqrt{n}\left(\overline{Y}_n - V\right)}{\sqrt{L_n}}$. Then, $R_n \overset{stably}{\longrightarrow} \mathcal{N}(0, 1)$ provided $Q_1 > 0$, σ_0 is non-degenerate, and the conditions of Theorem 5 are satisfied.*

Proof. Suppose first $L > 0$ a.s. Let K be the kernel on $\mathbb{R}^2$ given by

$$K(A \times B) = \delta_L(A)\mathcal{N}(0, L)(B) \quad \text{for all Borel sets } A,\ B \subset \mathbb{R}.$$

Then, $(L_n, W_n) \overset{stably}{\longrightarrow} K$ since $L_n \overset{a.s.}{\longrightarrow} L$ and $W_n \overset{stably}{\longrightarrow} \mathcal{N}(0, L)$; see Berti et al. (2011, Lemma 1). Fix $H \in \mathcal{A}$ with $P(H) > 0$ and define the function

$$f(x, y) = 1_{(0,\infty)}(x)\,\frac{y}{\sqrt{x}} \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

The set of discontinuity points of f is $D = \{(x, y) \in \mathbb{R}^2 : x = 0\}$ and

$$E\big(K(D) \mid H\big) = P(L = 0 \mid H) = 0.$$

Therefore,

$$P\big(R_n \in \cdot \mid H\big) = P\Big(f(L_n, W_n) \in \cdot \mid H\Big) \overset{weakly}{\longrightarrow} E\Big(K(f \in \cdot) \mid H\Big).$$

Finally, since $L > 0$ a.s. and $K(f \in \cdot) = \mathcal{N}(0, 1)(\cdot)$ on the set $\{L > 0\}$, one obtains

$$E\Big(K(f \in \cdot) \mid H\Big) = E\Big(K(f \in \cdot)\,1_{\{L>0\}} \mid H\Big) = \mathcal{N}(0, 1)(\cdot).$$

It remains to show that $L > 0$ a.s. Under each of the conditions of Theorem 5, one obtains $\sum_n (1-Q_n)\left(1+|Y_n|+Y_n^2\right) < \infty$ a.s. Therefore,

$$V \overset{a.s.}{=} \lim_n E(Y_{n+1} \mid \mathcal{F}_n) = E(Y_1) \prod_{i=1}^{\infty} Q_i + \sum_{i=1}^{\infty} Y_i\,(1 - Q_i) \prod_{k=i+1}^{\infty} Q_k \quad \text{and}$$

$$V^* \overset{a.s.}{=} \lim_n E(Y_{n+1}^2 \mid \mathcal{F}_n) = E(Y_1^2) \prod_{i=1}^{\infty} Q_i + \sum_{i=1}^{\infty} Y_i^2\,(1 - Q_i) \prod_{k=i+1}^{\infty} Q_k.$$

In other terms, $L = V^* - V^2$ agrees with the variance of the kernel

$$\mu = \sigma_0 \prod_{i=1}^{\infty} Q_i + \sum_{i=1}^{\infty} \delta_{Y_i}\,(1 - Q_i) \prod_{k=i+1}^{\infty} Q_k.$$

Moreover, $\prod_{i=1}^{\infty} Q_i > 0$ a.s. due to $\sum_n (1 - Q_n) < \infty$ and $Q_n \geq Q_1 > 0$ for all n a.s. Thus, $L > 0$ a.s. follows from σ_0 non-degenerate. □

Among other things, Theorem 6 can be used to obtain credible intervals.

Example 7 (*Credible Intervals for V*)**.** Fix $\alpha \in (0, 1)$ and define

$$I_n(\alpha) = \Big(\overline{Y}_n - u_{\alpha/2}\sqrt{L_n/n}\,,\ \overline{Y}_n + u_{\alpha/2}\sqrt{L_n/n}\Big)$$

where $u_{\alpha/2}$ is the quantile of order $(1 - \alpha/2)$ of the standard normal distribution. Under the assumptions of Theorem 6,

$$\lim_n P\Big(V \in I_n(\alpha) \mid H\Big) = \lim_n P\big(-u_{\alpha/2} < R_n < u_{\alpha/2} \mid H\big) = 1 - \alpha$$

for each $H \in \mathcal{A}$ with $P(H) > 0$. Therefore, for large n, $I_n(\alpha)$ is an approximate $(1 - \alpha)$-credible interval for V.

It is worth noting that $I_n(\alpha)$ is an approximate credible interval conditionally on *every fixed* $H \in \mathcal{A}$ with $P(H) > 0$. In principle, one could profit of this fact by choosing some H related to the inferential procedure. An example could be $H = \{F \in A\}$, where F is a vector of covariates and A any measurable set. If $P(F \in A) > 0$, then $I_n(\alpha)$ is an approximate $(1 - \alpha)$-credible interval for V conditionally on $F \in A$, namely, $P\big(V \in I_n(\alpha) \mid F \in A\big)$ is close to $1 - \alpha$ for large n. Similarly, given $B \in \mathcal{B}$, another option could be $H = \{V \in B\}$ or even $H = \{F \in A,\ V \in B\}$ provided $P(H) > 0$.

Finally, we briefly discuss some of the assumptions made in the previous results.

Remark 8. In Theorem 3, the assumption $Q_n \geq \alpha$ a.s. for all n can be weakened into $\liminf_n E(Q_n^{-2}) < \infty$. Instead, $\alpha \leq Q_n \leq \beta$ a.s. cannot be replaced by $\alpha \leq E(Q_n) \leq \beta$. As an example, fix $B \in \mathcal{B}$ with $\sigma_0(B) \in (0, 1)$ and define

$$\alpha = (1/2)\,\sigma_0(B) + \sigma_0(B^c), \quad Q_1 = 1, \quad Q_n = (1/2)\,1_{\{Y_1 \in B\}} + 1_{\{Y_1 \notin B\}} \text{ for } n > 1.$$

Then, $Y_n = T_n$ a.s. for all n on the set $\{Y_1 \notin B\}$. However, T_n does not converge in probability since (T_n) is i.i.d. and σ_0 is non-degenerate. Therefore, Y_n fails to converge in probability, even if $E(Q_n) = \alpha \in (0, 1)$ for all $n > 1$.

Similarly, in Theorem 5, $Q_n \leq Q_{n+1}$ a.s. cannot be weakened into $E(Q_n) \leq E(Q_{n+1})$. However, Theorem 5 admits a few alternative versions. One of these versions is the following. If $S \subset \mathbb{R}$ is a bounded subset, $E\big\{\sup_n n\,(1 - Q_n)\big\} < \infty$ and $n\,(1 - Q_n) \xrightarrow{L_2} M$, for some random variable M, then

$$C_n \xrightarrow{stably} \mathcal{N}\Big(0,\ (1-M)^2 L\Big) \quad \text{and} \quad W_n \xrightarrow{stably} \mathcal{N}\Big(0,\ \big((1-M)^2 + M^2\big)\,L\Big).$$

For instance, letting $M = 1$, such a version applies to Dirichlet sequences (i.e., $Q_n = \frac{n-1+c}{n+c}$).

References

Airoldi, E.M., Costa, Thiago, Bassetti, F., Leisen, F., Guindani, M., 2014. Generalized species sampling priors with latent beta reinforcements. J. Am. Stat. Assoc. 109, 1466–1480.

Bassetti, F., Crimaldi, I., Leisen, F., 2010. Conditionally identically distributed species sampling sequences. Adv. Appl. Probab. 42, 433–459.

Berti, P., Crimaldi, I., Pratelli, L., Rigo, P., 2009. Rate of convergence of predictive distributions for dependent data. Bernoulli 15, 1351–1367.

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.

Berti, P., Dreassi, E., Pratelli, L., Rigo, P., 2020. A class of models for Bayesian predictive inference. Bernoulli available at: http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers (in press).

Berti, P., Pratelli, L., Rigo, P., 2004. Limit theorems for a class of identically distributed random variables. Ann. Probab. 32, 2029–2052.

Berti, P., Pratelli, L., Rigo, P., 2006. Almost sure weak convergence of random probability measures. Stochastics (Formerly: Stoch. Stoch. Rep.) 78, 91–97.

Berti, P., Pratelli, L., Rigo, P., 2012. Limit theorems for empirical processes based on dependent data. Electron. J. Probab. 17, 1–18.

Berti, P., Pratelli, L., Rigo, P., 2013. Exchangeable sequences driven by an absolutely continuous random measure. Ann. Probab. 41, 2090–2102.

Berti, P., Regazzini, E., Rigo, P., 1997. Well-calibrated, coherent forecasting systems. Theory Probab. Appl. 42, 82–102.

Cassese, A., Zhu, W., Guindani, M., Vannucci, M., 2019. A Bayesian nonparametric spiked process prior for dynamic model selection. Bayesian Anal. 14, 553–572.

Cifarelli, D.M., Regazzini, E., 1996. De Finetti's contribution to probability and statistics. Statist. Sci. 11, 253–282.

de Finetti, B., 1937. La prevision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincare 7, 1–68.

Dubins, L.E., Savage, L.J., 1965. How to Gamble if you must: Inequalities for Stochastic Processes. McGraw Hill.

Fortini, S., Ladelli, L., Regazzini, E., 2000. Exchangeability, predictive distributions and parametric models. Sankhya A 62, 86–109.

Fortini, S., Petrone, S., 2012. Predictive construction of priors in Bayesian nonparametrics. Braz. J. Probab. Statist. 26, 423–449.

Fortini, S., Petrone, S., Sporysheva, P., 2018. On a notion of partially conditionally identically distributed sequences. Stochastic Process. Appl. 128, 819–846.

Hahn, P.R., Martin, R., Walker, S.G., 2018. On recursive Bayesian predictive distributions. J. Am. Stat. Assoc. 113, 1085–1093.

Hill, B.M., 1993. Parametric models for A_n: splitting processes and mixtures. J. R. Stat. Soc. Ser. B Stat. Methodol. 55, 423–433.

Kallenberg, O., 1988. Spreading and predictable sampling in exchangeable sequences and processes. Ann. Probab. 16, 508–534.

Lee, J., Quintana, F.A., Muller, P., Trippa, L., 2013. Defining predictive probability functions for species sampling models. Statist. Sci. 28, 209–222.

Pitman, J., 1996. Some developments of the Blackwell–MacQueen urn scheme. In: Statistics, Probability and Game Theory. In: IMS Lect. Notes Mon. Series, vol. 30, pp. 245–267.

Pitman, J., 2006. Combinatorial Stochastic Processes. In: Lectures from the XXXII Summer School in Saint-Flour, Springer, Berlin.