RATE OF CONVERGENCE FOR PREDICTIVE DISTRIBUTIONS OF EXCHANGEABLE INDICATORS

PATRIZIA BERTI, IRENE CRIMALDI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. Let (X_n) be an exchangeable sequence of indicators and π the probability distribution of $\limsup_n \overline{X}_n$, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then,

$$C_n = \sqrt{n}\,\{\,\overline{X}_n - E(X_{n+1} \mid X_1,\ldots,X_n)\,\}$$

converges stably (in particular, in distribution) provided π does not have a singular continuous part. Moreover, $C_n \overset{P}{\to} 0$ in case π is absolutely continuous with respect to Lebesgue measure, and $\sqrt{n}\,C_n$ converges a.s. under a mild Lipschitz condition on the density of π. Results of this type are useful in Bayesian statistics where π is the prior distribution. Related results are also obtained for the case where the X_n take values in an arbitrary measurable space.

1. Introduction and motivations

A number of real problems reduce to predict the next outcome for a sequence of events, that is, to evaluate

$$E\big(X_{n+1} \mid X_1,\ldots,X_n\big) = P\big(X_{n+1} = 1 \mid X_1,\ldots,X_n\big)$$

where $X_1, X_2, \ldots$ are the indicators of such events.

Here, we focus on those situations where $E\big(X_{n+1} \mid X_1,\ldots,X_n\big)$ can not be calculated in closed form, and one decides to estimate it basing on the available data $X_1,\ldots,X_n$. Related references are [1], [2], [3], [4], [5], [11].

In case (X_n) is an *exchangeable* sequence, as assumed throughout, a reasonable approximation for $E\big(X_{n+1} \mid X_1,\ldots,X_n\big)$ is the observed frequency

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

In line with de Finetti [9], the choice of $\overline{X}_n$ can be defended as follows. Suppose (Z_n) is an exchangeable sequence of random variables, with values in a Polish space S, and $\mathcal{D}$ a class of Borel subsets of S. Then,

$$\sup_{B\in\mathcal{D}} \Big| \frac{1}{n}\sum_{i=1}^n I_{\{Z_i\in B\}} - P\big(Z_{n+1}\in B \mid Z_1,\ldots,Z_n\big)\Big| \overset{a.s.}{\longrightarrow} 0 \tag{1}$$

provided $\mathcal{D}$ is a Glivenko-Cantelli class in the i.i.d. case (that is, provided (1) holds in the particular case where (Z_n) is i.i.d.); see [4]. Roughly speaking, thus, the empirical distribution $\frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ is a consistent estimate of the *predictive distribution* $P\big(Z_{n+1}\in\cdot \mid Z_1,\ldots,Z_n\big)$ for exchangeable data.

2000 *Mathematics Subject Classification.* 60G09, 60B10, 60A10, 62F15.

Key words and phrases. Bayesian predictive inference – Central limit theorem – Empirical distribution – Exchangeability – Predictive distribution – Stable convergence.

Taking (1) as a starting point, the next step is to investigate the rate of convergence. If $S = \{0, 1\}$ and $Z_n = X_n$, this means to investigate whether

$$C(a_n) = a_n \left\{ \overline{X}_n - E\big(X_{n+1} \mid X_1, \ldots, X_n\big) \right\}$$

approaches a limit (in some sense) for suitable constants $a_n > 0$.

This is just the purpose of this paper. Letting $V = \limsup_n \overline{X}_n$ and $W(a_n) = a_n(\overline{X}_n - V)$, exchangeability of (X_n) yields

$$E\big(X_{n+1} \mid X_1, \ldots, X_n\big) = E\big(V \mid X_1, \ldots, X_n\big) \quad \text{a.s..}$$

Hence, $C(a_n) = E\big(W(a_n) \mid X_1, \ldots, X_n\big)$ a.s.. Also, $\sup_n E|W(\sqrt{n})|^r < \infty$ for all $r > 0$, as it is not hard to prove (see the proof of Theorem 2). If $\frac{a_n}{\sqrt{n}} \to 0$, it follows that

$$\begin{aligned} \big(E|C(a_n)|^r\big)^{\frac{1+r}{r}} &\leq E|C(a_n)|^{1+r} \leq E|W(a_n)|^{1+r} \\ \leq \sup_m E|W(\sqrt{m})|^{1+r} \big(\frac{a_n}{\sqrt{n}}\big)^{1+r} &\to 0, \quad \text{as } n \to \infty, \text{ for all } r > 0. \end{aligned}$$

And what about $a_n = \sqrt{n}$? The answer to this (natural) question depends on the law π of V.

Our main result (Theorems 2 and 4) is that $E|C(\sqrt{n})|^r \to 0$, for all $r > 0$, whenever π is absolutely continuous with respect to Lebesgue measure. One consequence is

$$E\big(X_{n+1} \mid X_1, \ldots, X_n\big) = \overline{X}_n + o_P\big(\frac{1}{\sqrt{n}}\big). \tag{2}$$

Under a mild Lipschitz condition on the density of π, one also obtains

$$E\big(X_{n+1} \mid X_1, \ldots, X_n\big) = \overline{X}_n + \frac{D}{n} + o\big(\frac{1}{n}\big) \quad \text{a.s.} \tag{2*}$$

for some real random variable D. In addition, if π does not have a singular continuous part, $C(\sqrt{n})$ converges *stably* in the sense of Renyi; cf. Section 2. In particular, $C(\sqrt{n})$ converges in distribution to the probability measure

$$P(V \notin \Delta)\, \delta_0 + \sum_{v \in \Delta} P(V = v)\, \mathcal{N}(0, v - v^2),$$

where $\Delta = \{v : P(V = v) > 0\}$ and $\mathcal{N}(0, \sigma^2)$ denotes the centered Gaussian law with variance σ^2 (with $\mathcal{N}(0, 0) = \delta_0$).

Finally, we make four brief remarks.

(i) To our knowledge, there is no general representation for the predictive distributions $P\big(Z_{n+1} \in \cdot \mid Z_1, \ldots, Z_n\big)$ of an exchangeable sequence (Z_n). Such a representation would be very useful. Results like (2) and (2*) contribute to fill the gap for indicators. The general case, where the Z_n take values in an arbitrary measurable space, is dealt with in Subsection 4.2.

(ii) In Bayesian statistics, π is the *prior* distribution. And priors are typically assumed absolutely continuous with respect to Lebesgue measure (possibly, with smooth densities). The results mentioned above, thus, apply to most Bayesian problems.

(iii) Let $p > 1$ and $c > 0$. Those π which are absolutely continuous with respect to Lebesgue measure, with a density f such that $\big(\int_0^1 f(x)^p \, dx\big)^{\frac{1}{p}} \leq c$ (or such that

$f \le c$), can be characterized via their moments

$$\int x^j \, \pi(dx) = EV^j = P\big(X_1 = \ldots = X_j = 1\big).$$

This is the "Markov moment problem". We refer to [10] for more on this topic.

(iv) The results mentioned above straightforwardly extend to k-step predictions. Let $a_1, \ldots, a_k \in \{0,1\}$. Then, $P\big(X_{n+1} = a_1, \ldots, X_{n+k} = a_k \mid X_1, \ldots, X_n\big)$ can be approximated by $\overline{X}_n^{\sum_i a_i}\big(1 - \overline{X}_n\big)^{k - \sum_i a_i}$ (where the possible indeterminate form 0^0 should be meant as $0^0 = 1$). Moreover, the error

$$\overline{X}_n^{\sum_i a_i}\big(1 - \overline{X}_n\big)^{k - \sum_i a_i} - P\big(X_{n+1} = a_1, \ldots, X_{n+k} = a_k \mid X_1, \ldots, X_n\big)$$

behaves asymptotically as $\big(\overline{X}_n - E\big(X_{n+1} \mid X_1, \ldots, X_n\big)\big)$; see Subsection 4.1.

2. Stable convergence

Let $(\Omega, \mathcal{A}, P)$ be a probability space and S a metric space. We write $\mathcal{B}$ for the Borel σ-field of S and $C_b(S)$ for the set of real bounded continuous functions on S. A *random probability measure on* S, defined on $(\Omega, \mathcal{A}, P)$, is a mapping N on $\Omega \times \mathcal{B}$ such that: (i) $N(\omega, \cdot)$ is a probability measure on $\mathcal{B}$ for $\omega \in \Omega$; (ii) $N(\cdot, B)$ is $\mathcal{A}$-measurable for $B \in \mathcal{B}$. The real random variable $N(\omega, f) = \int f(x)\, N(\omega, dx)$, where f is a bounded $\mathcal{B}$-measurable function on S, is denoted by $N(f)$.

Let us turn to stable convergence. Let (Z_n) be a sequence of S-valued random variables and N a random probability measure on S. Both (Z_n) and N are defined on $(\Omega, \mathcal{A}, P)$. Say that Z_n *converges stably to* N in case

$$E\big(f(Z_n) \mid H\big) \to E\big(N(f) \mid H\big) \quad \text{for all } f \in C_b(S) \text{ and } H \in \mathcal{A} \text{ with } P(H) > 0.$$

If $Z_n \to N$ stably, then Z_n converges in distribution to the probability measure $B \mapsto EN(B)$ on $\mathcal{B}$ (just let $H = \Omega$). Stable convergence has been introduced by Renyi in [13] and subsequently investigated by various authors. A detailed treatment, including some strengthened forms of stable convergence, is in [8].

3. Main results

In the sequel, as in Section 1, $(X_n : n \ge 1)$ is an *exchangeable sequence of indicators* on the probability space $(\Omega, \mathcal{A}, P)$. We let

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \quad V = \limsup_n \overline{X}_n, \quad \Delta = \{v : P(V = v) > 0\}.$$

Also,

$$\pi = P \circ V^{-1}$$

is the probability distribution of V, λ the Lebesgue measure on $(0,1)$, and $\mathcal{N}(a,b)$ the Gaussian law with mean a and variance $b \ge 0$ (with $\mathcal{N}(a,0) = \delta_a$).

We first investigate stable convergence of $E\left(f(W_n) \mid \mathcal{G}_n\right)$, where $f \in C_b(\mathbb{R})$ and

$$W_n = W(\sqrt{n}) = \sqrt{n}\left(\overline{X}_n - V\right), \quad \mathcal{G}_n = \sigma(X_1, \ldots, X_n).$$

To this end, we begin with two introductory examples.

If (X_n) is i.i.d., then $V = v$ a.s. for some $v \in [0,1]$ and a result of Renyi [13] yields $\sqrt{n}\left(\overline{X}_n - v\right) \to \mathcal{N}(0, v - v^2)$ stably. Since $\sqrt{n}\left(\overline{X}_n - v\right)$ is $\mathcal{G}_n$-measurable,

$$E\left(f(W_n) \mid \mathcal{G}_n\right) = f(W_n) \to \mathcal{N}(0, v - v^2) \circ f^{-1} \quad \text{stably for all } f \in C_b(\mathbb{R}).$$

Note that π is discrete in the i.i.d. case (in fact, $\pi = \delta_v$).

Suppose now that (X_n) is a *Polya sequence*, that is, $P(X_1 = 1) = \frac{u}{u+v}$ and

$$E\big(X_{n+1} \mid \mathcal{G}_n\big) = \frac{u + \sum_{i=1}^n X_i}{u+v+n} \quad \text{a.s.}$$

for some reals $u, v > 0$. When u, v are rationals, this probability assessment describes a well known urn scheme. In any case, π is a beta distribution with parameters u, v and

$$E\big(f(W_n) \mid \mathcal{G}_n\big) \overset{a.s.}{\to} \mathcal{N}(0, V - V^2)(f) \quad \text{for all } f \in C_b(\mathbb{R}). \tag{3}$$

Condition (3) has been first proved in Example 6 of [8] (with convergence in probability in the place of a.s. convergence) and then in Corollary 4.2 of [7].

One conjecture is that (3) holds whenever $\pi \ll \lambda$ (and not only in the Polya case). Provided this is true, further, the discrete and absolutely continuous cases could be unified. Next result realizes this programme.

Theorem 1. *Let $f \in C_b(\mathbb{R})$. If π does not have a singular continuous part, then $E\big(f(W_n) \mid \mathcal{G}_n\big)$ converges stably to the random probability measure*

$$M_f = I_{\{V \notin \Delta\}} \delta_{\mathcal{N}(0,V-V^2)(f)} + I_{\{V \in \Delta\}} \mathcal{N}(0, V - V^2) \circ f^{-1}.$$

Moreover, condition (3) *holds whenever $\pi \ll \lambda$.*

Proof. Let N denote the random probability measure $N = \mathcal{N}(0, V - V^2)$.

First, suppose $\pi \ll \lambda$. In order to prove (3), it can be assumed $\Omega = \{0,1\}^\infty$, $\mathcal{A}$ the Borel σ-field and X_n the canonical projections. In this case, (X_n) is a Polya sequence under some probability measure P_0 on $\mathcal{A}$. Let π_0 be the distribution of V under P_0 (recall that π_0 is a beta distribution). Since $\pi \ll \lambda$ and λ is equivalent to π_0, then $\pi \ll \pi_0$ and de Finetti's representation theorem implies $P \ll P_0$. Thus,

$$\sup_{A \in \mathcal{A}} | P\big((X_{n+1}, \ldots) \in A \mid \mathcal{G}_n\big) - P_0\big((X_{n+1}, \ldots) \in A \mid \mathcal{G}_n\big) | \to 0, \quad P\text{-a.s.},$$

by Blackwell-Dubins result on merging [6]. Given $f \in C_b(\mathbb{R})$, define

$$U_n = |E_{P_0}\big(f(W_n) \mid \mathcal{G}_n\big) - N(f)|, \quad V_n = |E\big(f(W_n) \mid \mathcal{G}_n\big) - E_{P_0}\big(f(W_n) \mid \mathcal{G}_n\big)|.$$

By [7], since (X_n) is Polya under P_0, then $U_n \to 0$, P_0-a.s.. By Blackwell-Dubins result on merging, $V_n \to 0$, P-a.s.. Since $P \ll P_0$, one obtains

$$|E\big(f(W_n) \mid \mathcal{G}_n\big) - N(f)| \leq U_n + V_n \to 0, \quad P\text{-a.s.}.$$

Thus, condition (3) holds whenever $\pi \ll \lambda$.

Next, suppose π does not have a singular continuous part. Fix $f \in C_b(\mathbb{R})$, $-1 \leq f \leq 1$, and let $A = \{V \in \Delta\}$. Since $W_n \to N$ stably (see [5], Theorem 3.1),

$$E\big(M_f(g) \mid A \cap H\big) = E\big(N(g \circ f) \mid A \cap H\big) = \lim_n E\big(g \circ f(W_n) \mid A \cap H\big)$$

provided $g \in C_b(\mathbb{R})$, $H \in \mathcal{A}$ and $P(A \cap H) > 0$. It follows that

$$I_A f(W_n) + I_{A^c} N(f) \to M_f \quad \text{stably.}$$

In order to prove $E\big(f(W_n) \mid \mathcal{G}_n\big) \to M_f$ stably, thus, it suffices showing that

$$E\Big|E\big(f(W_n) \mid \mathcal{G}_n\big) - I_A f(W_n) - I_{A^c} N(f)\Big| \to 0.$$

Write $\Delta = \{v_1, v_2, \ldots\}$. Since $|f| \le 1$, one obtains

$$E\Big|E\big(I_A f(W_n) \mid \mathcal{G}_n\big) - I_A f(W_n)\Big|$$
$$= E\Big|\sum_j f\left(\sqrt{n}(\overline{X}_n - v_j)\right)\left(P(V = v_j \mid \mathcal{G}_n) - I_{\{V=v_j\}}\right)\Big|$$
$$\le \sum_j E\Big|P(V = v_j \mid \mathcal{G}_n) - I_{\{V=v_j\}}\Big|$$
$$\le \sum_{j=1}^m E\Big|P(V = v_j \mid \mathcal{G}_n) - I_{\{V=v_j\}}\Big| + 2\sum_{j>m} P(V = v_j) \quad \text{for all } m.$$

By martingale convergence, $E|P(V = v_j \mid \mathcal{G}_n) - I_{\{V=v_j\}}| \to 0$ for fixed j, and thus

$$\limsup_n E\Big|E\big(I_A f(W_n) \mid \mathcal{G}_n\big) - I_A f(W_n)\Big| \le 2 \limsup_m \sum_{j>m} P(V = v_j) = 0.$$

It remains to see that

$$E\Big|E\big(I_{A^c} f(W_n) \mid \mathcal{G}_n\big) - I_{A^c} N(f)\Big| \to 0.$$

To this end, it can be assumed $P(A^c) > 0$. Denote $Q(\cdot) = P(\cdot \mid A^c)$. On noting that $|f| \le 1$ and

$$E_Q\big(f(W_n) \mid \mathcal{G}_n\big) = \frac{E\big(I_{A^c} f(W_n) \mid \mathcal{G}_n\big)}{P(A^c \mid \mathcal{G}_n)}, \quad Q\text{-a.s.},$$

one obtains

$$E\Big|E\big(I_{A^c} f(W_n) \mid \mathcal{G}_n\big) - I_{A^c} N(f)\Big|$$
$$\le E\Big|I_A\, E\big(I_{A^c} f(W_n) \mid \mathcal{G}_n\big)\Big| + E_Q\Big|E\big(I_{A^c} f(W_n) \mid \mathcal{G}_n\big) - N(f)\Big|$$
$$\le E\big(I_A\, P(A^c \mid \mathcal{G}_n)\big) + E_Q\Big|P(A^c \mid \mathcal{G}_n) E_Q\big(f(W_n) \mid \mathcal{G}_n\big) - N(f)\Big|$$
$$\le E\big(I_A\, P(A^c \mid \mathcal{G}_n)\big) + E_Q\Big|P(A \mid \mathcal{G}_n)\Big| + E_Q\Big|E_Q\big(f(W_n) \mid \mathcal{G}_n\big) - N(f)\Big|.$$

By martingale convergence,

$$E\big(I_A\, P(A^c \mid \mathcal{G}_n)\big) + E_Q\Big|P(A \mid \mathcal{G}_n)\Big| = E\big(I_A\, P(A^c \mid \mathcal{G}_n)\big) + \frac{E\big(I_{A^c}\, P(A \mid \mathcal{G}_n)\big)}{P(A^c)} \to 0.$$

Finally, since π does not have a singular continuous part, the distribution of V under Q is absolutely continuous with respect to λ. Also, (X_n) is still exchangeable under Q. Hence, the first part of this proof yields

$$E_Q\Big|E_Q\big(f(W_n) \mid \mathcal{G}_n\big) - N(f)\Big| \to 0.$$

□

Incidentally, the previous proof shows that condition (3) holds, even if (X_n) is not exchangeable, provided the law of (X_n) is absolutely continuous with respect to the law of a Polya sequence. For proving (3), in fact, we only used $P \ll P_0$.

Theorem 1 also sheds light on the rate of convergence of $\big\{\overline{X}_n - E\big(X_{n+1} \mid \mathcal{G}_n\big)\big\}$, which is our main purpose. Recall that $E\big(X_{n+1} \mid \mathcal{G}_n\big) = E\big(V \mid \mathcal{G}_n\big)$ a.s. and define

$$C_n = C(\sqrt{n}) = \sqrt{n}\left\{\overline{X}_n - E\big(X_{n+1} \mid \mathcal{G}_n\big)\right\} = E\big(W_n \mid \mathcal{G}_n\big) \quad \text{a.s..}$$

Theorem 2. *If π does not have a singular continuous part, then C_n converges stably to the random probability measure*

$$M = I_{\{V \notin \Delta\}} \delta_0 + I_{\{V \in \Delta\}} \mathcal{N}(0, V - V^2).$$

Moreover $E|C_n|^r \to 0$, for all $r > 0$, whenever $\pi \ll \lambda$.

The following lemma, needed for proving Theorem 2, is certainly known. Since we do not know of any reference, however, we give it a proof.

Lemma 3. *Let (Y_n) be a sequence of real i.i.d. random variables on a common probability space, with $EY_1^{2k} < \infty$ and $EY_1 = 0$, $k = 1, 2, \ldots$. Then,*

$$\sup_n E\left\{ \left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \right)^{2k} \right\} \leq \gamma_k \, EY_1^{2k}$$

for some constant γ_k depending on k only.

Proof. Let $S_0 = 0$ and $S_n = \sum_{i=1}^n Y_i$. Then, S_n is a martingale with quadratic variation $[S]_0 = 0$ and $[S]_n = \sum_{i=1}^n (S_i - S_{i-1})^2 = \sum_{i=1}^n Y_i^2$. By the well known Burkholder-Davis-Gundy inequality, there is a universal constant γ_k such that $E\big(\max_{0 \leq j \leq n} S_j^{2k}\big) \leq \gamma_k \, E\big([S]_n^k\big)$. For such a γ_k and any integer n, one obtains

$$E\left\{ \left(\frac{S_n}{\sqrt{n}} \right)^{2k} \right\} \leq \frac{\gamma_k}{n^k} E\Big\{ \Big(\sum_{i=1}^n Y_i^2 \Big)^k \Big\} = \frac{\gamma_k}{n^k} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n E\big(Y_{i_1}^2 \ldots Y_{i_k}^2 \big) \leq \gamma_k \, EY_1^{2k}.$$

□

Proof of Theorem 2. Let $\mathcal{T}$ denote the tail σ-field of (X_n). By exchangeability of (X_n) and Lemma 3, for each integer $k \geq 1$ there is a constant γ_k satisfying

$$\sup_n EW_n^{2k} = \sup_n E\big(E(W_n^{2k} \mid \mathcal{T})\big) \leq \gamma_k \, E\big(E(X_1^{2k} \mid \mathcal{T})\big) = \gamma_k \, EX_1^{2k} < \infty.$$

Further, $EC_n^{2k} \leq EW_n^{2k}$ since $C_n = E\big(W_n \mid \mathcal{G}_n\big)$ a.s.. Hence, both the sequences $(|W_n|^r)$ and $(|C_n|^r)$ are uniformly integrable for all real $r > 0$.

Next, suppose π does not have a singular continuous part. Let $g : \mathbb{R} \to \mathbb{R}$ be a function satisfying $|g(x) - g(y)| \leq |x - y|$ and $|g(x)| \leq 1$ for all x, y, and let $H \in \mathcal{A}$ with $P(H) > 0$. To prove $C_n \to M$ stably, it is enough to see that $E\big(g(C_n) \mid H\big) \to E\left(M(g) \mid H\right)$. Since (W_n) is uniformly integrable, given $\epsilon > 0$, there is $c > 0$ such that

$$\sup_n E\left(|W_n| \, I_{\{|W_n| > c\}}\right) < \frac{\epsilon \, P(H)}{4} \quad \text{and} \quad c^2 > \frac{1}{\epsilon}.$$

Define $f(x) = x$ for $|x| \leq c$, $f(x) = c$ for $x > c$, and $f(x) = -c$ for $x < -c$, and let $U_n = E\left(f(W_n) \mid \mathcal{G}_n\right)$. Since g is Lipschitz continuous and $C_n = E\big(W_n \mid \mathcal{G}_n\big)$ a.s.,

$$|g(C_n) - g(U_n)| \leq |C_n - U_n| \leq 2 \, E\big(|W_n| \, I_{\{|W_n| > c\}} \mid \mathcal{G}_n\big) \quad \text{a.s.,}$$

and this implies

$$E\big(|g(C_n) - g(U_n)| \mid H\big) \leq \frac{2}{P(H)} E\big(|W_n| \, I_{\{|W_n| > c\}}\big) < \frac{\epsilon}{2} \quad \text{for all } n.$$

Since $\mathcal{N}(0, V - V^2)(f) = 0$, then $M_f = \delta_0 = M$ on $\{V \notin \Delta\}$, where M_f is the random probability measure appearing in Theorem 1. Further, since $|g| \leq 1$,

$0 \le V \le 1$ and $c^2 > \frac{1}{\epsilon}$, one obtains

$$\begin{aligned}|\mathcal{N}(0,V-V^2)(g\circ f)-\mathcal{N}(0,V-V^2)(g)\,| &\le \mathcal{N}(0,V-V^2)(|g\circ f-g|)\\ &\le 2\,\mathcal{N}(0,V-V^2)(\{x:|x|>c\}) \le \frac{2(V-V^2)}{c^2} \le \frac{2}{c^2}\frac{1}{4} < \frac{\epsilon}{2}.\end{aligned}$$

To sum up, one can estimate as follows

$$\begin{aligned}&|E\big(g(C_n)\mid H\big)-E\big(M(g)\mid H\big)|-|E\big(g(U_n)\mid H\big)-E\big(M_f(g)\mid H\big)|\\ &\le |E\big(g(C_n)\mid H\big)-E\big(g(U_n)\mid H\big)|+|E\big(M_f(g)\mid H\big)-E\big(M(g)\mid H\big)|\\ &< \frac{\epsilon}{2}+E\big(I_{\{V\in\Delta\}}|M_f(g)-M(g)|\mid H\big)\\ &= \frac{\epsilon}{2}+E\big(I_{\{V\in\Delta\}}\,|\mathcal{N}(0,V-V^2)(g\circ f)-\mathcal{N}(0,V-V^2)(g)|\mid H\big)\\ &\le \frac{\epsilon}{2}+\frac{\epsilon}{2}P(V\in\Delta\mid H)\le\epsilon.\end{aligned}$$

Since Theorem 1 yields $E\big(g(U_n)\mid H\big)\to E\big(M_f(g)\mid H\big)$, one obtains

$$\limsup_n |E\big(g(C_n)\mid H\big)-E\big(M(g)\mid H\big)|\le\epsilon.$$

Therefore, $C_n \to M$ stably. In particular, if $\pi \ll \lambda$, then $C_n \to M = \delta_0$ stably, that is, $C_n \overset{P}{\to} 0$. Hence $E|C_n|^r \to 0$, because of uniform integrability of $(|C_n|^r)$, for all $r > 0$. This concludes the proof. □

At least two remarks on Theorem 2 are in order.

First, if π has a singular continuous part, we suspect that C_n converges stably to a non null limit. But we have not a proof.

Second,

$$\sqrt{n}\,C_n = n\,\big\{\,\overline{X}_n - E\big(X_{n+1}\mid\mathcal{G}_n\big)\,\big\}$$

converges a.s. in case (X_n) is a Polya sequence. A conjecture is that $\sqrt{n}\,C_n$ converges a.s. whenever $\pi \ll \lambda$. This is actually true, as we now prove, under some conditions on the density. Say that a real function f on $(0,1)$ is *almost Lipschitz* in case $x \mapsto f(x)x^a(1-x)^b$ is Lipschitz on $(0,1)$ for some reals a, $b < 1$.

Theorem 4. *If π admits an almost Lipschitz density with respect to λ, then $\sqrt{n}\,C_n \overset{a.s.}{\to} D$ for some real random variable D.*

A few technical facts, needed for proving Theorem 4, are collected in the following lemma.

Lemma 5. *Let $\Omega = \{0,1\}^\infty$, $\mathcal{A}$ the Borel σ-field on Ω and X_n the canonical projections. Let P_0 be the probability on $\mathcal{A}$ which makes (X_n) a Polya sequence (for some u, $v > 0$). If $\pi \ll \lambda$, there is a nonnegative Borel function h on $[0,1]$ such that $h(V)$ is a density of P with respect to P_0. Moreover,*

$$P(V\in B) = c\int_B h(x)\,x^{u-1}\,(1-x)^{v-1}\,dx$$

for each Borel set $B \subset [0,1]$, where $c > 0$ is a constant.

Proof. Let $\mathcal{N}_0 = \{A \in \mathcal{A} : P_0(A) = 0\}$ and $\mathcal{S}$ the symmetric σ-field on $\Omega = \{0,1\}^\infty$. Since $\pi \ll \lambda$, then $P \ll P_0$. Fix a version f of $\frac{dP}{dP_0}$ and a finite permutation ϕ of

Ω. Let φ be the finite permutation such that $\varphi \circ \phi(\omega) = \omega$ for all $\omega \in \Omega$. By exchangeability of both P and P_0, one obtains

$$\int_A f \, dP_0 = P(A) = P(\varphi^{-1} A) = \int (I_A \circ \varphi) \, f \, dP_0 = \int_A (f \circ \phi) \, dP_0$$

for all $A \in \mathcal{A}$. Hence, $\{f \neq f \circ \phi\} \in \mathcal{N}_0$. Since finite permutations are countably many, there is a nonnegative $\mathcal{S}$-measurable function f_1 on Ω satisfying $\{f \neq f_1\} \in \mathcal{N}_0$. Since f_1 is $\mathcal{S}$-measurable and P_0 exchangeable,

$$\sigma(f_1) \subset \mathcal{S} \subset \sigma\big(\sigma(V) \cup \mathcal{N}_0\big).$$

Hence, $\{f_1 \neq f_2\} \in \mathcal{N}_0$ for some nonnegative $\sigma(V)$-measurable function f_2 on Ω. Such f_2 is a version of $\frac{dP}{dP_0}$ and $f_2 = h(V)$ for some nonnegative Borel function h. Finally, it suffices noting that the distribution of V under P_0 is beta with parameters u, v. □

Proof of Theorem 4. Since $\sqrt{n}\, C_n$ is a $\mathcal{G}_n$-martingale, it suffices proving that $\sup_n \sqrt{n}\, E|C_n| < \infty$. To this end, it can be assumed $\Omega = \{0,1\}^\infty$, $\mathcal{A}$ the Borel σ-field and X_n the canonical projections.

Since π has an almost Lipschitz density, there is a version g of $\frac{d\pi}{d\lambda}$ such that $x \mapsto g(x) x^a (1-x)^b$ is Lipschitz on $(0,1)$ for some a, $b < 1$. Let P_0 be the probability on $\mathcal{A}$ which makes (X_n) a Polya sequence with $u = 1 - a$ and $v = 1 - b$. By Lemma 5, some version of $\frac{dP}{dP_0}$ is of the form $h(V)$ where h is a nonnegative Lipschitz function on $(0,1)$.

Using such version, C_n can be written as

$$C_n = E\big(W_n \mid \mathcal{G}_n\big) = \frac{E_0\big(h(V)\, W_n \mid \mathcal{G}_n\big)}{E_0\big(h(V) \mid \mathcal{G}_n\big)}, \quad P\text{-a.s.},$$

where E_0 denotes expectation under P_0. Thus,

$$E|C_n| = E_0\Big\{\, h(V)\, \frac{|E_0\big(h(V)\, W_n \mid \mathcal{G}_n\big)|}{E_0\big(h(V) \mid \mathcal{G}_n\big)} \,\Big\} = E_0 \,|\, E_0\big(h(V)\, W_n \mid \mathcal{G}_n\big) \,|.$$

Let

$$V_n = E_0\big(V \mid \mathcal{G}_n\big) = E_0\big(X_{n+1} \mid \mathcal{G}_n\big) = \frac{u + \sum_{i=1}^n X_i}{u + v + n}.$$

Then, $\sqrt{n}\,|\, E_0\big(W_n \mid \mathcal{G}_n\big) \,| = n\, |\overline{X}_n - V_n| \leq u + v$, P_0-a.s.. Since h is Lipschitz (and thus bounded) on $(0,1)$ and $P_0\big(0 < V_n < 1,\, 0 < V < 1\big) = 1$ for all n, it follows that

$$\begin{aligned} E|C_n| &\leq \; E_0 \,|\, h(V_n)\, E_0(W_n \mid \mathcal{G}_n) \,| + \; E_0 \,|\, E_0\big((h(V) - h(V_n))\, W_n \mid \mathcal{G}_n\big) \,| \\ &\leq \frac{(u+v)\, \sup h}{\sqrt{n}} + c\, E_0\big\{\, E_0\big(|(V - V_n)\, W_n| \mid \mathcal{G}_n\big) \,\big\} \end{aligned}$$

where c is the Lipschitz constant of h. Letting $U_n = \sqrt{n}\,(V - V_n)$, one also obtains

$$\sqrt{n}\, E|C_n| \leq (u+v)\, \sup h + c\, E_0\big\{ E_0\big(|U_n W_n| \mid \mathcal{G}_n\big) \,\big\} = (u+v)\, \sup h + c\, E_0|U_n W_n|.$$

As noted in the proof of Theorem 2, $E_0 C_n^2 \leq E_0 W_n^2 \leq d$ for all n and some constant d. Since $U_n = C_n - W_n$, it follows that $E_0 U_n^2 \leq 2\,(E_0 C_n^2 + E_0 W_n^2) \leq 4\, d$ and

$$\sqrt{n}\, E|C_n| \leq (u+v)\, \sup h \,+\, c\, \sqrt{E_0 U_n^2\, E_0 W_n^2} \leq (u+v)\, \sup h \,+\, 2\, c\, d$$

for all n. This concludes the proof. □

4. Miscellaneous results

The results obtained so far admit some generalizations.

4.1. **k-step predictions.** Let $a_1, \dots, a_k \in \{0,1\}$ and $a = \sum_{i=1}^k a_i$. Then,

$$P\big(X_{n+1} = a_1, \dots, X_{n+k} = a_k \mid \mathcal{G}_n\big) = E\big(V^a\,(1-V)^{k-a} \mid \mathcal{G}_n\big)$$

is well approximated by $\overline{X}_n^a\,\big(1-\overline{X}_n\big)^{k-a}$ (where the possible indeterminate form 0^0 is meant as $0^0 = 1$). In addition, the asymptotic behaviour of

$$T_n = \sqrt{n}\,\Big\{\,\overline{X}_n^a\,\big(1-\overline{X}_n\big)^{k-a} \;-\; E\big(V^a\,(1-V)^{k-a} \mid \mathcal{G}_n\big)\,\Big\}$$

is quite similar to that of C_n.

Corollary 6. *If π does not have a singular continuous part, then T_n converges stably to the random probability measure*

$$M(\sigma^2) = I_{\{V\notin\Delta\}}\delta_0 + I_{\{V\in\Delta\}}\mathcal{N}(0,\sigma^2) = \mathcal{N}(0,\, I_{\{V\in\Delta\}}\,\sigma^2),$$

where $\sigma^2 = k^2V^{2k-1}(1-V)$ if $a = k$, $\sigma^2 = k^2V(1-V)^{2k-1}$ if $a = 0$, and

$$\sigma^2 = (a-kV)^2\,V^{2a-1}(1-V)^{2(k-a)-1} \quad \text{if } 0 < a < k.$$

In particular, $T_n \overset{P}{\to} 0$ in case $\pi \ll \lambda$.

Proof. Letting $f(x) = x^a(1-x)^{k-a}$, Lagrange theorem yields

$$\begin{aligned} T_n &= \sqrt{n}\,E\big(f(\overline{X}_n) - f(V) \mid \mathcal{G}_n\big) = \sqrt{n}\,E\big(\,f'(V_n)\,(\overline{X}_n - V) \mid \mathcal{G}_n\big) \\ &= f'(\overline{X}_n)\,E(W_n \mid \mathcal{G}_n) \;+\; E\big((f'(V_n) - f'(\overline{X}_n))\,W_n \mid \mathcal{G}_n\big) \quad \text{a.s.} \end{aligned}$$

where V_n is between $\overline{X}_n$ and V. Let $M = \mathcal{N}(0,\, I_{\{V\in\Delta\}}V(1-V))$. Since $C_n \to M$ stably (by Theorem 2), f' is continuous and $\overline{X}_n \overset{a.s.}{\to} V$, one obtains

$$f'(\overline{X}_n)\,E(W_n \mid \mathcal{G}_n) = f'(\overline{X}_n)\,C_n \;\to\; \mathcal{N}(0,\, I_{\{V\in\Delta\}}f'(V)^2V(1-V)) = M(\sigma^2) \quad \text{stably.}$$

Thus, it remains only to see that $E\big((f'(V_n) - f'(\overline{X}_n))\,W_n \mid \mathcal{G}_n\big) \overset{P}{\to} 0$. Let $R_n = f'(V_n) - f'(\overline{X}_n)$. Then, $R_n \overset{a.s.}{\to} 0$ and $R_n^2 \le 4\max_{0\le x\le 1} f'(x)^2$ for all n. Since $\sup_n EW_n^2 < \infty$, it follows that

$$E\big|\,E\big(R_n\,W_n \mid \mathcal{G}_n\big)\,\big| \le E|\,R_n\,W_n\,| \le \sqrt{EW_n^2\,ER_n^2} \;\to\; 0.$$

□

4.2. **General state space.** Let (Z_n) be an exchangeable sequence of random variables, defined on $(\Omega, \mathcal{A}, P)$ and taking values in the measurable space $(S, \mathcal{B})$. Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ denote the empirical measure and

$$\mathcal{G}_n = \sigma(Z_1, \dots, Z_n).$$

Given $B \in \mathcal{B}$, let us consider

$$C_n^* = \sqrt{n}\,\big\{\,\mu_n(B) - E\big(I_{\{Z_{n+1}\in B\}} \mid \mathcal{G}_n\big)\,\big\}.$$

After Section 3, we know something about $\sqrt{n}\,\big\{\mu_n(B) - E\big(I_{\{Z_{n+1}\in B\}} \mid \mathcal{G}_n^B\big)\big\}$ where $\mathcal{G}_n^B = \sigma(I_{\{Z_1\in B\}}, \dots, I_{\{Z_n\in B\}})$. But this is not enough for C_n^*, since the asymptotic behaviour of

$$\sqrt{n}\,\big\{\,E\big(I_{\{Z_{n+1}\in B\}} \mid \mathcal{G}_n^B\big) - E\big(I_{\{Z_{n+1}\in B\}} \mid \mathcal{G}_n\big)\,\big\}$$

is unknown (to us). The arguments of Section 3, however, give some help.

Let $Z=(Z_1,Z_2,\ldots)$ and ν a probability measure on $\mathcal{B}^\infty$ such that

$$C_n^* \overset{P}{\to} 0 \quad \text{whenever } Z\sim\nu. \tag{4}$$

Theorem 7. *Suppose $P(Z\in\cdot)\ll\nu$, where $Z=(Z_1,Z_2,\ldots)$ and ν is a probability on $\mathcal{B}^\infty$ satisfying* (4)*. Then, $E|C_n^*|^r\to 0$ for all $r>0$. In particular,*

$$E\big(I_{\{Z_{n+1}\in B\}}\mid\mathcal{G}_n\big)=\mu_n(B)+o_P(\frac{1}{\sqrt{n}}).$$

To avoid repetitions, we just give a sketch of the proof.

Proof of Theorem 7. Define $V_B=\limsup_n\mu_n(B)$, $W_n^*=\sqrt{n}(\mu_n(B)-V_B)$ and note that $C_n^*=E\big(W_n^*\mid\mathcal{G}_n\big)$ a.s.. As in the proof of Theorem 2, the sequences $(|W_n^*|^r)$ and $(|C_n^*|^r)$ can be shown to be uniformly integrable for all $r>0$. Thus, it suffices proving that $E|C_n^*|\to 0$. It can be assumed $(\Omega,\mathcal{A})=(S^\infty,\mathcal{B}^\infty)$ and Z_n the canonical projections. Let $P_0=\nu$ and f a version of $\frac{dP}{dP_0}$. As in the proof of Theorem 4, $E|C_n^*|=E_0\,|\,E_0\big(f\,W_n^*\mid\mathcal{G}_n\big)\,|$ where E_0 denotes expectation under P_0. Since (W_n^*) is uniformly integrable, given $\epsilon>0$, there is $c>0$ such that

$$E_0\,|\,E_0\big(f\,I_{\{f>c\}}\,W_n^*\mid\mathcal{G}_n\big)\,|\le E_0\big(f\,I_{\{f>c\}}\,|W_n^*|\big)=E\big(I_{\{f>c\}}\,|W_n^*|\big)<\epsilon\quad\text{for all } n.$$

Using such c, define $U_n=f\,I_{\{f\le c\}}-E_0\big(f\,I_{\{f\le c\}}\mid\mathcal{G}_n\big)$. Then,

$$\begin{aligned}E|C_n^*|&<\epsilon+E_0\,|\,E_0\big(f\,I_{\{f\le c\}}\,W_n^*\mid\mathcal{G}_n\big)\,|\\ &\le\epsilon+E_0|\,E_0\big(f\,I_{\{f\le c\}}\mid\mathcal{G}_n\big)\,E_0\big(W_n^*\mid\mathcal{G}_n\big)\,|+E_0|\,E_0\big(U_n\,W_n^*\mid\mathcal{G}_n\big)\,|\\ &\le\epsilon+c\,E_0|\,E_0\big(W_n^*\mid\mathcal{G}_n\big)\,|+\sqrt{E_0U_n^2\,E_0W_n^{*\,2}}.\end{aligned}$$

By (4), $E_0\big(W_n^*\mid\mathcal{G}_n\big)\overset{P_0}{\to}0$. Since the sequence $\big(E_0\big(W_n^*\mid\mathcal{G}_n\big)\big)$ is uniformly integrable under P_0, then $E_0|E_0\big(W_n^*\mid\mathcal{G}_n\big)|\to 0$. Thus, to conclude the proof, it suffices noting that $\sup_n E_0W_n^{*\,2}<\infty$ and

$$\lim_n E_0U_n^2=\lim_n E_0\big\{\big(f\,I_{\{f\le c\}}-E_0\big(f\,I_{\{f\le c\}}\mid\mathcal{G}_n\big)\big)^2\big\}=0$$

by martingale convergence. □

Various examples of ν satisfying (4) are available in the *Bayesian nonparametrics* framework; see e.g. [12] and references therein. One of the most popular is the law of a Ferguson-Dirichlet sequence. If Z is such a sequence,

$$E\big(I_{\{Z_{n+1}\in B\}}\mid\mathcal{G}_n\big)=\frac{a\,P(Z_1\in B)+n\,\mu_n(B)}{a+n}\quad\text{a.s.}$$

for some $a>0$, and thus $|C_n^*|\le\frac{a}{\sqrt{n}}$. Note that Ferguson-Dirichlet sequences reduce to Polya's for $S=\{0,1\}$.

Let ν denote the law of a Ferguson-Dirichlet sequence. Characterizing those Z such that $P(Z\in\cdot)\ll\nu$ is quite easy in case S is finite (and $\mathcal{B}$ the power set of S). Suppose in fact $S=\{x_1,\ldots,x_k,x_{k+1}\}$ and $P(Z_1=x)>0$ for all $x\in S$. Define $V_x=\limsup_n\mu_n\{x\}$. Then, $P(Z\in\cdot)\ll\nu$ if and only if $(V_{x_1},\ldots,V_{x_k})$ has an absolutely continuous distribution, with respect to Lebesgue measure, on the set $\{(u_1,\ldots,u_k):u_i>0\text{ for all } i \text{ and }\sum_{i=1}^k u_i<1\}$. Note that, in case of indicators ($S=\{0,1\}$ and $0<P(Z_1=1)<1$), one obtains $P(Z\in\cdot)\ll\nu$ if and only if $V=V_1$ has an absolutely continuous distribution with respect to Lebesgue measure on $(0,1)$.

For general state spaces, instead, we do not know of reasonably simple characterizations of $P(Z \in \cdot) \ll \nu$.

References

[1] Algoet P.H. (1992) Universal schemes for prediction, gambling and portfolio selection, *Ann. Probab.*, 20, 901-941.

[2] Algoet P.H. (1995) Universal prediction schemes (Correction), *Ann. Probab.*, 23, 474-478.

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

[4] Berti P., Mattei A. and Rigo P. (2002) Uniform convergence of empirical and predictive measures, *Atti Sem. Mat. Fis. Univ. Modena*, L, 465-477.

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

[6] Blackwell D. and Dubins L.E.(1962) Merging of opinions with increasing information, *Ann. Math. Statist.*, 33, 882-886.

[7] Crimaldi I. (2007) Almost sure conditional convergence for a generalized Polya urn, *submitted*, currently available at: `http://www.dm.unibo.it/~crimaldi/articoli-inviati.html`

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

[9] de Finetti B. (1937) La prevision: ses lois logiques, ses sources subjectives, *Annales Instit. Poincare*, 7, 1-68.

[10] Diaconis P. and Freedman D. (2004) The Markov moment problem and de Finetti's theorem, Parts I and II, *Math. Zeitschrift*, 247, 183-199 and 201-212.

[11] Morvai G. and Weiss B. (2005) Forward estimation for ergodic time series, *Ann. Inst. H. Poincare Probab. Statist.*, 41, 859-870.

[12] Pitman J. (1996) Some developments of the Blackwell-MacQueen urn scheme, In: *Statistics, Probability and Game Theory* (Ferguson, Shapley and Mac Queen Eds.), IMS Lecture Notes Monogr. Ser., 30, 245-267.

[13] Renyi A. (1963) On stable sequences of events, *Sankhya* A, 25, 293-302.

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`

Irene Crimaldi, Dipartimento di Matematica, Universita' di Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
E-mail address: `crimaldi@dm.unibo.it`

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

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