Article

A Central Limit Theorem for Predictive Distributions

Patrizia Berti [1], Luca Pratelli [2] and Pietro Rigo [3],*

[1] Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio-Emilia, Via Campi 213/B, 41100 Modena, Italy; patrizia.berti@unimore.it

[2] Accademia Navale di Livorno, 57127 Livorno, Italy; pratel@mail.dm.unipi.it

[3] Dipartimento di Scienze Statistiche "P. Fortunati", Università di Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy

* Correspondence: pietro.rigo@unibo.it

Abstract: Let S be a Borel subset of a Polish space and F the set of bounded Borel functions $f : S \to \mathbb{R}$. Let $a_n(\cdot) = P(X_{n+1} \in \cdot \mid X_1, \dots, X_n)$ be the n-th predictive distribution corresponding to a sequence (X_n) of S-valued random variables. If (X_n) is conditionally identically distributed, there is a random probability measure μ on S such that $\int f \, da_n \xrightarrow{a.s.} \int f \, d\mu$ for all $f \in F$. Define $D_n(f) = d_n \{ \int f \, da_n - \int f \, d\mu \}$ for all $f \in F$, where $d_n > 0$ is a constant. In this note, it is shown that, under some conditions on (X_n) and with a suitable choice of d_n, the finite dimensional distributions of the process $D_n = \{ D_n(f) : f \in F \}$ stably converge to a Gaussian kernel with a known covariance structure. In addition, $E\{ \varphi(D_n(f)) \mid X_1, \dots, X_n \}$ converges in probability for all $f \in F$ and $\varphi \in C_b(\mathbb{R})$.

Keywords: bayesian predictive inference; central limit theorem; conditional identity in distribution; exchangeability; predictive distribution; stable convergence

MSC: 60B10; 60G25; 60G09; 60F05; 62F15; 62M20

1. Introduction

All random elements appearing in the sequel are defined on a common probability space, say $(\Omega, \mathcal{A}, P)$. We denote by S a Borel subset of a Polish space and by $\mathcal{B}$ the Borel σ-field on S. We let

$$\mathcal{P} = \{ \text{probability measures on } \mathcal{B} \} \quad \text{and}$$
$$F = \{ \text{real bounded Borel functions on } S \}.$$

Moreover, if $\lambda \in \mathcal{P}$ and $f \in F$, we write $\lambda(f)$ to denote

$$\lambda(f) = \int f \, d\lambda.$$

In other terms, depending on the context, λ is regarded as a function on $\mathcal{B}$ or a function on F. This slight abuse of notation is quite usual (see, e.g., [1,2]) and very useful for the purposes of this note.

Let

$$X = (X_1, X_2, \dots)$$

be a sequence of S-valued random variables and

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad \text{and} \quad \mathcal{F}_n = \sigma(X_1, \dots, X_n).$$

The *predictive distributions* of X are the random probability measures on $(S, \mathcal{B})$ given by

$$a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{F}_n) \qquad \text{for all } n \geq 0.$$

Under some conditions, there is a further random probability measure μ on $(S, \mathcal{B})$ such that

$$\mu(f) \overset{a.s.}{=} \lim_n a_n(f) \qquad \text{for each } f \in F. \tag{1}$$

For instance, condition (1) holds if X is exchangeable. More generally, it holds if X is conditionally identically distributed (c.i.d.), as defined in Section 2. Note also that, since S is separable, condition (1) implies $a_n \to \mu$ weakly. Regarding a_n and μ as measurable functions from Ω into $\mathcal{P}$, one obtains

$$P\big(\{\omega \in \Omega : a_{n,\omega} \to \mu_\omega \text{ weakly}\}\big) = 1.$$

Assume condition (1), fix a sequence d_n of positive constants, and define

$$D_n(f) = d_n \big\{a_n(f) - \mu(f)\big\} \qquad \text{for each } f \in F.$$

This note deals with the process

$$D_n = \big\{D_n(f) : f \in F\big\}.$$

Our goal is to show that, under some conditions on X and with a suitable choice of the constants d_n, the finite-dimensional distributions of D_n stably converge, as $n \to \infty$, to a certain Gaussian limit.

To be more precise, we recall that a *kernel* on $(S, \mathcal{B})$ is a measurable map $\alpha : S \to \mathcal{P}$. This means that $\alpha(x) \in \mathcal{P}$, for each $x \in S$, and the function $x \mapsto \alpha(x)(A)$ is $\mathcal{B}$-measurable for each $A \in \mathcal{B}$. In what follows, we write

$$\alpha(x)(f) = \int f(y)\, \alpha(x)(dy) \qquad \text{for all } x \in S \text{ and } f \in F.$$

Next, as in [3], suppose the predictive distributions of X satisfy the recursive equation

$$a_{n+1} = q_n\, a_n + (1 - q_n)\, \alpha(X_{n+1}) \qquad \text{a.s. for all } n \geq 0, \tag{2}$$

where $q_0, q_1, \ldots \in (0, 1)$ are constants and α is a kernel on $(S, \mathcal{B})$. Moreover, let

$$\nu(\cdot) = P\big(X_1 \in \cdot\big)$$

be the marginal distribution of X_1. Under condition (2), X is c.i.d. whenever α is a regular conditional distribution for ν given a sub-σ-field $\mathcal{G} \subset \mathcal{B}$; see ([3] Section 5). Hence, we assume

$$\alpha(\cdot)(A) = E_\nu(1_A \mid \mathcal{G}), \quad \nu\text{-a.s.}, \tag{3}$$

for all $A \in \mathcal{B}$ and some sub-σ-field $\mathcal{G} \subset \mathcal{B}$. For instance, condition (3) holds if

$$\alpha(x) = \delta_x \qquad \text{for all } x \in S$$

where δ_x denotes the unit mass at the point x (just let $\mathcal{G} = \mathcal{B}$). In addition, we assume

$$\sum_n (1 - q_n)^2 < \infty \quad \text{and} \quad \lim_n d_n \sup_{k \geq n}(1 - q_{k-1}) = 0$$

where

$$d_n = \left(\sum_{k \geq n}(1 - q_k)^2\right)^{-1/2}.$$

In this framework, it is shown that

$$\big(D_n(f_1), \ldots, D_n(f_p)\big) \longrightarrow \mathcal{N}_p(0, \Sigma) \qquad \text{stably} \tag{4}$$

for all $p \geq 1$ and all $f_1, \ldots, f_p \in F$, where Σ is the random covariance matrix with entries

$$\sigma_{jk} = \int \alpha(x)(f_j)\,\alpha(x)(f_k)\,\mu(dx) - \mu(f_j)\,\mu(f_k).$$

We actually prove something more than (4). Let $C_b(\mathbb{R})$ denote the set of real bounded continuous functions on $\mathbb{R}$. Then, it is shown that

$$E\{\varphi(D_n(f)) \mid \mathcal{F}_n\} \xrightarrow{P} \mathcal{N}(0, \sigma^2)(\varphi) \tag{5}$$

for all $f \in F$ and $\varphi \in C_b(\mathbb{R})$, where

$$\sigma^2 = \int \alpha(x)(f)^2\,\mu(dx) - \mu(f)^2.$$

Based on (5), it is not hard to deduce condition (4).

Before concluding the Introduction, several remarks are in order.

(i) A remarkable special case is $\alpha(x) = \delta_x$ for all $x \in S$. Indeed, Equation (2) holds with $\alpha = \delta$ in some meaningful situations, including Dirichlet sequences; see ([3] Section 4) for other examples. Thus, suppose $\alpha = \delta$. Then, the above formulae reduce to $\sigma_{jk} = \mu(f_j f_k) - \mu(f_j)\,\mu(f_k)$ and $\sigma^2 = \mu(f^2) - \mu(f)^2$. Moreover, if ν is non-atomic and

$$\prod_{j=0}^{n} q_j \to 0 \quad \text{and} \quad \sum_n \prod_{j=0}^{n} q_j = \infty,$$

then μ takes the form

$$\mu \stackrel{a.s.}{=} \sum_n V_n\,\delta_{Y_n}$$

where (V_n) and (Y_n) are independent sequences and (Y_n) is i.i.d. with $Y_1 \sim \nu$; see ([3] Theorem 20) and [4] for details.

(ii) Let $l^\infty(G)$ be the set of real bounded functions on G, where G is any subset of F. For instance, if $S = \mathbb{R}$, one could take $G = \{1_{(-\infty,x]} : x \in \mathbb{R}\}$. In view of (4), a natural question is whether D_n has a limit in distribution when $l^\infty(G)$ is equipped with a suitable distance. As an example, $l^\infty(G)$ could be equipped with the uniform distance (as in [1,2]) or with some weaker distance (as in [5]). Even if natural, this question is neglected in this note. We hope and plan to investigate it in a forthcoming paper.

(iii) For fixed $f \in F$, condition (4) provides some information on the convergence rate of $a_n(f)$ to $\mu(f)$. Define $L_n = u_n\,|a_n(f) - \mu(f)|$ where $u_n > 0$ is any sequence of constants. Then, condition (4) yields $L_n \xrightarrow{P} 0$ whenever $u_n/d_n \to 0$. Furthermore, $L_n \xrightarrow{P} \infty$ provided $u_n/d_n \to \infty$ and $\sigma^2 > 0$ a.s.

(iv) The condition $\lim_n d_n \sup_{k \geq n}(1 - q_{k-1}) = 0$ is just a technical assumption which guarantees that, asymptotically, there are no dominating terms. In a sense, this condition is analogous to the weak Lindeberg's condition in the classical CLT for independent summands.

(v) From a Bayesian point of view, μ can be seen as a random parameter of the data sequence X. This is quite clear if X is exchangeable, for, in this case, X is conditionally i.i.d. given μ. If X is only c.i.d., the role of μ is not as crucial, but μ still contributes to specify the probability distribution of X; see ([3] Section 2.1). Thus, in a Bayesian framework, conditions (4)–(5) may be useful to make (asymptotic) inference about μ. To this end, an alternative could be proving a limit theorem for $W_n = w_n\,(\mu_n - \mu)$, where w_n is a suitable constant and $\mu_n = (1/n)\sum_{j=1}^n \delta_{X_j}$ the empirical measure. However, D_n has two advantages with respect to W_n. It usually converges at a better rate and the variance of the limit distribution is smaller; see, e.g., Example 3.

(vi) Conditions (4)–(5) are our main results. They can be motivated in at least two ways. Firstly, from the theoretical perspective, conditions (4)–(5) fit into the results concerning the asymptotic behavior of conditional expectations (see, e.g., [6–8] and references therein). Secondly, from the practical perspective, conditions (4)–(5) play a role in all those fields where predictive distributions are basic objects. The main example is Bayesian predictive inference. Indeed, the predictive distributions investigated in this note have been introduced in connection with Bayesian prediction problems; see [3]. Another example is the asymptotic behavior of certain urn schemes. Related subjects, where (4)–(5) are potentially useful, are empirical processes for dependent data, Glivenko-Cantelli-type theorems and merging of opinions. Without any claim of being exhaustive, a list of references is: [3,5,9–21].

2. Preliminaries

In this note, $\mathcal{N}_p(0, C)$ denotes the Gaussian law on the Borel sets of $\mathbb{R}^p$ with mean 0 and covariance matrix C, where C is symmetric and semidefinite positive. If $p = 1$ and $c \geq 0$ is a scalar, we write $\mathcal{N}(0, c)$ instead of $\mathcal{N}_1(0, c)$ and

$$\mathcal{N}(0, c)(\varphi) = \int \varphi(x)\, \mathcal{N}(0, c)(dx)$$

for all bounded measurable $\varphi : \mathbb{R} \to \mathbb{R}$. Note that, if Σ is a random covariance matrix, $\mathcal{N}_p(0, \Sigma)$ is a random probability measure on the Borel sets of $\mathbb{R}^p$.

Let us briefly recall *stable convergence*. Let $\mathcal{A}^+ = \{H \in \mathcal{A} : P(H) > 0\}$. Fix a random probability measure K on $(S, \mathcal{B})$ and define

$$\lambda_H(A) = E\{K(A) \mid H\} \qquad \text{for all } A \in \mathcal{B} \text{ and } H \in \mathcal{A}^+.$$

Each λ_H is a probability measure on $\mathcal{B}$. Then, X_n *converges stably to* K, written $X_n \to K$ stably, if

$$P(X_n \in \cdot \mid H) \longrightarrow \lambda_H \text{ weakly for all } H \in \mathcal{A}^+.$$

In particular, X_n converges in distribution to λ_Ω. However, stable convergence is stronger than convergence in distribution. To see this, take a further random variable $X : \Omega \to S$. Then, $X_n \xrightarrow{P} X$ if, and only if, $X_n \to \delta_X$ stably. Thus, stable convergence is strictly connected to convergence in probability. Moreover, $(X_n, X) \to K \times \delta_X$ stably whenever $X_n \to K$ stably. Therefore, if X_n converges stably, (X_n, X) still converges stably for *any* S-valued random variable X.

We next turn to conditional identity in distribution. Say that X is *conditionally identically distributed* (c.i.d.) if

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

Thus, at each time n, the future observations $(X_k : k > n)$ are identically distributed given the past. This is actually weaker than exchangeability. Indeed, X is exchangeable if, and only if, it is stationary and c.i.d.

C.i.d. sequences were introduced in [9,22] and then investigated in various papers; see, e.g., [3–5,11,23–29].

The asymptotics of c.i.d. sequences is similar to that of exchangeable ones. To see this, suppose X is c.i.d. and define the empirical measures

$$\mu_n = \frac{1}{n} \sum_{j=1}^n \delta_{X_j}.$$

Then, there is a random probability measure μ on $(S, \mathcal{B})$ such that

$$\mu(A) \overset{a.s.}{=} \lim_m \mu_m(A) \qquad \text{for each fixed } A \in \mathcal{B}.$$

It follows that

$$E\{\mu(A) \mid \mathcal{F}_n\} = \lim_m E\{\mu_m(A) \mid \mathcal{F}_n\}$$

$$= \lim_m \frac{1}{m} \sum_{j=n+1}^{m} P(X_j \in A \mid \mathcal{F}_n) = P(X_{n+1} \in A \mid \mathcal{F}_n) \qquad \text{a.s.}$$

for all $n \geq 0$ and $A \in \mathcal{B}$. Therefore, as in the exchangeable case, the predictive distributions can be written as

$$a_n(\cdot) = P(X_{n+1} \in \cdot \mid \mathcal{F}_n) = E\{\mu(\cdot) \mid \mathcal{F}_n\} \qquad \text{a.s.}$$

Using the martingale convergence theorem, this implies

$$\mu(f) \overset{a.s.}{=} \lim_n E\{\mu(f) \mid \mathcal{F}_n\} = \lim_n a_n(f) \qquad \text{for all } f \in F.$$

Furthermore, X is asymptotically exchangeable, in the sense that the probability distribution of the shifted sequence $(X_n, X_{n+1}, \ldots)$ converges weakly to an exchangeable probability measure on $(S^\infty, \mathcal{B}^\infty)$.

Finally, we state a technical result to be used later on.

Lemma 1. *Let (Y_n) be a sequence of real integrable random variables, adapted to the filtration $(\mathcal{F}_n)$, and*

$$Z_n = E(Y_{n+1} \mid \mathcal{F}_n).$$

Let V be a real non-negative random variable and $0 < b_1 < b_2 < \ldots$ an increasing sequence of constants, such that $b_n \uparrow \infty$ and $b_n/b_{n+1} \to 1$. Suppose (Y_n^2) is uniformly integrable, $Z_n \overset{a.s.}{\longrightarrow} Z$ for some random variable Z, and define

$$T_n = b_n (Z_n - Z).$$

Then,

$$E\{\varphi(T_n) \mid \mathcal{F}_n\} \overset{P}{\longrightarrow} \mathcal{N}(0, V)(\varphi) \qquad \text{for all } \varphi \in C_b(\mathbb{R})$$

provided

$$b_n^2 \sum_{k \geq n} (Z_k - Z_{k-1})^2 \overset{P}{\longrightarrow} V; \tag{6}$$

$$\lim_n b_n E\left\{ \sup_{k \geq n} |Z_k - Z_{k-1}| \right\} = 0; \tag{7}$$

$$\sum_{k \geq n} E\left| E(Z_{k+1} \mid \mathcal{F}_k) - Z_k \right| = o(1/b_n). \tag{8}$$

Proof. Just repeat the proof of ([10] Theorem 1) with b_n in the place of $\sqrt{n}$. $\quad\square$

3. Main Result

Let us go back to the notation of Section 1. Recall that $q_n \in (0, 1)$ is a constant for each $n \geq 0$ and $d_n = \left(\sum_{k \geq n} (1 - q_k)^2 \right)^{-1/2}$. We aim to prove the following CLT.

Theorem 1. *Assume conditions* (2)–(3) *and*

$$\sum_n (1 - q_n)^2 < \infty \quad and \quad \lim_n d_n \sup_{k \geq n} (1 - q_{k-1}) = 0.$$

Then, there is a random probability measure μ on $(S, \mathcal{B})$ such that

$$\mu(f) \overset{a.s.}{=} \lim_n a_n(f) \qquad and \qquad E\{\varphi(D_n(f)) \mid \mathcal{F}_n\} \overset{P}{\longrightarrow} \mathcal{N}(0, \sigma^2)(\varphi)$$

for all $f \in F$ and $\varphi \in C_b(\mathbb{R})$, where

$$\sigma^2 = \int \alpha(x)(f)^2 \, \mu(dx) - \mu(f)^2.$$

As a consequence,

$$\left(D_n(f_1), \ldots, D_n(f_p) \right) \longrightarrow \mathcal{N}_p(0, \Sigma) \qquad stably$$

for all $p \geq 1$ and all $f_1, \ldots, f_p \in F$ where the covariance matrix Σ has entries

$$\sigma_{jk} = \int \alpha(x)(f_j) \, \alpha(x)(f_k) \, \mu(dx) - \mu(f_j) \, \mu(f_k).$$

Proof. Due to conditions (2)–(3), X is c.i.d.; see ([3] Section 5). Hence, as noted in Section 2, there is a random probability measure μ on $(S, \mathcal{B})$ such that

$$a_n(f) \overset{a.s.}{=} E\{\mu(f) \mid \mathcal{F}_n\} \qquad \text{for all } f \in F.$$

By martingale convergence, it follows that $a_n(f) \overset{a.s.}{\longrightarrow} \mu(f)$ for all $f \in F$.

We next prove condition (5). Fix $f \in F$ and define

$$b_n = d_n, \quad Y_n = a_n(f), \quad Z = \mu(f) \quad \text{and} \quad V = \sigma^2.$$

Then, (Y_n^2) is uniformly integrable (for f is bounded) and b_n satisfies the conditions of Lemma 1. Moreover,

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

so that $Z_n \overset{a.s.}{\longrightarrow} Z$. Therefore, Lemma 1 applies. Hence, to prove (5), it suffices to check conditions (6)–(8).

Let $c = \sup|f|$. Since $E(Z_{k+1} \mid \mathcal{F}_k) = Z_k$ a.s., condition (8) is trivially true. Moreover, condition (2) implies

$$
\begin{aligned}
Z_k - Z_{k-1} &= a_k(f) - a_{k-1}(f) \\
&= q_{k-1} \, a_{k-1}(f) + (1 - q_{k-1}) \, \alpha(X_k)(f) - a_{k-1}(f) \\
&= (1 - q_{k-1}) \{ \alpha(X_k)(f) - a_{k-1}(f) \} \qquad \text{a.s. for all } k \geq 1.
\end{aligned}
$$

Hence, condition (7) holds, since

$$d_n E \left\{ \sup_{k \geq n} |Z_k - Z_{k-1}| \right\} \leq 2 \, c \, d_n \sup_{k \geq n} (1 - q_{k-1}) \longrightarrow 0.$$

It remains to prove condition (6), namely

$$d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 \left\{ \alpha(X_k)(f) - a_{k-1}(f) \right\}^2 \overset{P}{\longrightarrow} \sigma^2.$$

First note that, since $a_{k-1}(f)^2 \xrightarrow{a.s.} \mu(f)^2$ as $k \to \infty$, one obtains

$$d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 a_{k-1}(f)^2 = \frac{\sum_{k \geq n} (1 - q_{k-1})^2 a_{k-1}(f)^2}{\sum_{k \geq n} (1 - q_k)^2} \xrightarrow{a.s.} \mu(f)^2.$$

Next, define

$$R_k = \alpha(X_k)(f)^2 \quad \text{and} \quad M_n = d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 \{R_k - E(R_k \mid \mathcal{F}_{k-1})\}.$$

Then,

$$\begin{aligned} E(M_n^2) &= d_n^4 \sum_{k \geq n} (1 - q_{k-1})^4 E\{(R_k - E(R_k \mid \mathcal{F}_{k-1}))^2\} \\ &\leq 4 c^4 d_n^4 \sum_{k \geq n} (1 - q_{k-1})^4 \\ &\leq 4 c^4 d_n^2 \sup_{k \geq n} (1 - q_{k-1})^2 \cdot d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 \\ &\longrightarrow 0. \end{aligned}$$

Moreover,

$$E(R_k \mid \mathcal{F}_{k-1}) = E\left\{ \int \alpha(x)(f)^2 \mu(dx) \mid \mathcal{F}_{k-1} \right\} \xrightarrow{a.s.} \int \alpha(x)(f)^2 \mu(dx).$$

Therefore,

$$d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 R_k = M_n + d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 E(R_k \mid \mathcal{F}_{k-1}) \xrightarrow{P} \int \alpha(x)(f)^2 \mu(dx).$$

By the same argument, it follows that

$$d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 \alpha(X_k)(f) a_{k-1}(f) \xrightarrow{P} \mu(f) \int \alpha(x)(f) \mu(dx).$$

In addition, as proved in the Claim below,

$$\int \alpha(x)(f) \mu(dx) \stackrel{a.s.}{=} \mu(f).$$

Collecting all pieces together, one finally obtains

$$d_n^2 \sum_{k \geq n} (1 - q_{k-1})^2 \{\alpha(X_k)(f) - a_{k-1}(f)\}^2 \xrightarrow{P} \mu(f)^2 + \int \alpha(x)(f)^2 \mu(dx) - 2 \mu(f)^2 = \sigma^2.$$

Hence, condition (6) holds.

This concludes the proof of (5). We next prove that (5) $\Rightarrow$ (4). Let $p \geq 1$ and $f_1, \ldots, f_p \in F$. Fix $u_1, \ldots, u_p \in \mathbb{R}$ and define

$$U_n = \sum_{j=1}^{p} u_j D_n(f_j) \quad \text{and} \quad \sigma_u^2 = \sum_{j,k} u_j u_k \sigma_{jk}.$$

Moreover, for each $H \in \mathcal{A}^+$, define the probability measure

$$\lambda_H(A) = E\{\mathcal{N}(0, \sigma_u^2)(A) \mid H\} \quad \text{for each Borel set } A \subset \mathbb{R}.$$

We have to show that

$$P(U_n \in \cdot \mid H) \longrightarrow \lambda_H \text{ weakly for each } H \in \mathcal{A}^+. \tag{9}$$

To this end, call ϕ_H the characteristic function of λ_H, namely

$$\phi_H(t) = E\left(\int e^{itx}\mathcal{N}(0, \sigma_u^2)(dx) \mid H\right) = E\left(e^{-t^2\sigma_u^2/2} \mid H\right) \quad \text{for all } t \in \mathbb{R}.$$

Letting $f = \sum_{j=1}^{p} u_j f_j$, one obtains

$$U_n = D_n(f) \quad \text{and} \quad \sigma_u^2 = \int \alpha(x)(f)^2\,\mu(dx) - \mu(f)^2.$$

Therefore, condition (5) yields

$$E\left(e^{it\,U_n}\right) = E\left(E\left\{e^{it\,D_n(f)} \mid \mathcal{F}_n\right\}\right) \longrightarrow E\left(e^{-t^2\sigma_u^2/2}\right) = \phi_\Omega(t)$$

for each $t \in \mathbb{R}$. Hence, condition (9) holds for $H = \Omega$. Next, suppose $H \in \bigcup_n \mathcal{F}_n$ and $P(H) > 0$. Then, for large n, one obtains

$$E\left(1_H\,e^{it\,U_n}\right) = E\left(1_H\,E\left\{e^{it\,D_n(f)} \mid \mathcal{F}_n\right\}\right).$$

Hence, for each $t \in \mathbb{R}$, condition (5) still implies

$$P(H)\,\phi_H(t) = E\left(1_H\,e^{-t^2\sigma_u^2/2}\right) = \lim_n E\left(1_H\,E\left\{e^{it\,D_n(f)} \mid \mathcal{F}_n\right\}\right) = \lim_n E\left(1_H\,e^{it\,U_n}\right).$$

Therefore, condition (9) holds whenever $H \in \bigcup_n \mathcal{F}_n$ and $P(H) > 0$. Based on this fact, by standard arguments, condition (9) easily follows for each $H \in \mathcal{A}^+$.

To conclude the proof of the Theorem, it remains only to show that:

Claim: $\int \alpha(x)(f)\,\mu(dx) \overset{a.s.}{=} \mu(f)$ for all $f \in F$.

Proof of the Claim: By (3), α is a regular conditional distribution for ν given a sub-σ-field of $\mathcal{B}$, where ν is the marginal distribution of X_1. Therefore, as proved in ([3] Lemma 6), there is a set $A \in \mathcal{B}$ such that $\nu(A) = 1$ and

$$\int \alpha(z)(f)\,\alpha(x)(dz) = \alpha(x)(f) \qquad \text{for all } x \in A \text{ and } f \in F.$$

Since X is c.i.d. (and, thus, identically distributed) one also obtains $P(X_n \in A) = \nu(A) = 1$ for all $n \geq 1$.

Having noted these facts, fix $f \in F$. Since $a_0 = \nu$ and α is a regular conditional distribution for ν,

$$\int \alpha(x)(f)\,a_0(dx) = a_0(f).$$

Moreover, if $\int \alpha(x)(f)\,a_n(dx) = a_n(f)$ a.s. for some $n \geq 0$, then

$$\begin{aligned}
\int \alpha(x)(f)\,a_{n+1}(dx) &= q_n \int \alpha(x)(f)\,a_n(dx) + (1-q_n)\int \alpha(x)(f)\,\alpha(X_{n+1})(dx) \\
&= q_n\,a_n(f) + (1-q_n)\,\alpha(X_{n+1})(f) \\
&= a_{n+1}(f) \quad \text{a.s.}
\end{aligned}$$

By induction, one obtains $\int \alpha(x)(f)\,a_n(dx) = a_n(f)$ a.s. for each $n \geq 0$. Hence,

$$\int \alpha(x)(f)\,\mu(dx) = \lim_n \int \alpha(x)(f)\,a_n(dx) = \lim_n a_n(f) = \mu(f) \qquad \text{a.s.}$$

$\square$

We do not know whether $E\{\varphi(D_n(f)) \mid \mathcal{F}_n\}$ converges a.s. (and not only in probability) under the conditions of Theorem 1. However, it can be shown that $E\{\varphi(D_n(f)) \mid \mathcal{F}_n\}$ converges a.s. under slightly stronger conditions on q_n.

Under conditions (2)–(3), for Theorem 1 to work, it suffices that

$$\lim_n n^b \left(1 - q_n\right) = c \qquad \text{for some } b > 1/2 \text{ and } c > 0. \tag{10}$$

In addition, if (10) holds, then

$$\frac{n^{b-1/2}}{d_n} \to \frac{c}{\sqrt{2b-1}}.$$

Hence, letting $D_n^* = n^{b-1/2}(a_n - \mu)$, one obtains

$$\left(D_n^*(f_1), \ldots, D_n^*(f_p)\right) \longrightarrow \mathcal{N}_p\left(0, \frac{c^2}{2b-1}\Sigma\right) \qquad \text{stably,}$$

for all $p \geq 1$ and all $f_1, \ldots, f_p \in F$, provided conditions (2), (3) and (10) hold.

We close this note with some examples.

Example 1. *Let*

$$q_n = \frac{n + \theta_n}{n + 1 + \theta_{n+1}}$$

where (θ_n) is a bounded increasing sequence with $\theta_0 > 0$. Then, X is c.i.d. (because of (2)–(3)) but is exchangeable if and only if $\theta_n = \theta_0$ for all n. In any case, since condition (10) holds with $b = c = 1$, Theorem 1 applies and d_n can be replaced by $\sqrt{n}$. Letting $D_n^* = \sqrt{n}\,(a_n - \mu)$, it follows that

$$\left(D_n^*(f_1), \ldots, D_n^*(f_p)\right) \longrightarrow \mathcal{N}_p(0, \Sigma) \qquad \text{stably.}$$

It is worth noting that, in the special case $\theta_n = \theta_0$ for all n, the predictive distributions of X reduce to

$$a_n = \frac{\theta_0\, \nu + \sum_{i=1}^n \alpha(X_i)}{n + \theta_0}.$$

Therefore, X is a Dirichlet sequence if $\alpha = \delta$. The general case, where α is any kernel satisfying condition (3), is investigated in [30]. It turns out that X satisfies most properties of Dirichlet sequences. In particular, μ has the same distribution as

$$\mu^* = \sum_n V_n\, \alpha(Y_n),$$

where (V_n) and (Y_n) are independent sequences, (Y_n) is i.i.d. with $Y_1 \sim \nu$, and (V_n) has the stick breaking distribution. Nevertheless, as shown in the next example, X can behave quite differently from a Dirichlet sequence.

Example 2 (Example 1 continued). *Let $\mathcal{H}$ be a countable partition of S such that $H \in \mathcal{B}$ and $\nu(H) > 0$ for all $H \in \mathcal{H}$. Define*

$$\alpha(x) = \sum_{H \in \mathcal{H}} 1_H(x)\, \nu(\cdot \mid H) = \nu(\cdot \mid H_x) \qquad \text{for all } x \in S$$

where H_x is the only element of the partition $\mathcal{H}$, such that $x \in H$. Then, α is a regular conditional distribution for ν given $\sigma(\mathcal{H})$ (i.e., condition (3) holds). If the q_n are as in Example 1 with $\theta_n = \theta_0$ for all n, one obtains

$$a_n = \frac{\theta_0\, \nu + \sum_{i=1}^n \nu(\cdot \mid H_{X_i})}{n + \theta_0}.$$

Therefore,

$$a_n \ll \nu \qquad \text{for all } n \geq 0. \tag{11}$$

This is a striking difference with respect to Dirichlet sequences. For instance, if ν is non-atomic, condition (11) yields

$$P(X_i = X_j \text{ for some } i \neq j) = 0$$

while $P(X_i = X_j \text{ for some } i \neq j) = 1$ if X is a Dirichlet sequence. Note also that, for each $f \in F$,

$$\sigma^2 = \int \alpha(x)(f)^2 \, \mu(dx) - \mu(f)^2 = \sum_{H \in \mathcal{H}} \nu(f \mid H)^2 \, \mu(H) - \mu(f)^2$$

while $\sigma^2 = \mu(f^2) - \mu(f)^2$ if X is a Dirichlet sequence. Other choices of α, which make X quite different from a Dirichlet sequence, are in [30].

Example 3. *A meaningful special case is $\sum_n (1 - q_n) < \infty$. In this case,*

$$\prod_{j=0}^{\infty} q_j := \lim_n \prod_{j=0}^{n} q_j$$

exists and is strictly positive. Hence, μ admits the representation

$$\mu = \nu \prod_{j=0}^{\infty} q_j + \sum_{i=1}^{\infty} \alpha(X_i) \, (1 - q_{i-1}) \prod_{j=i}^{\infty} q_j.$$

As an example, under conditions (2)–(3), Theorem 1 applies whenever

$$q_n = \exp\{-(c + n)^{-2}\} \qquad \text{for some constant } c > 0.$$

With this choice of q_n, one obtains $(1 - q_n)(c + n)^2 \to 1$, so that $\sum_n (1 - q_n) < \infty$ and μ can be written as above. Note also that

$$\lim_n \frac{d_n}{(c + n)^{3/2}} = \sqrt{3}.$$

Therefore, for fixed $f \in F$, the rate of convergence of $a_n(f)$ to $\mu(f)$ is $n^{-3/2}$ and not the usual $n^{-1/2}$.

References

1. Dudley, R.M. *Uniform Central Limit Theorems*; Cambridge University Press: Cambridge, UK, 1999.
2. van der Vaart, A.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
3. Berti, P.; Dreassi, E.; Pratelli, L.; Rigo, P. A class of models for Bayesian predictive inference. *Bernoulli* **2021**, *27*, 702–726. [CrossRef]
4. Berti, P.; Dreassi, E.; Pratelli, L.; Rigo, P. Asymptotics of certain conditionally identically distributed sequences. *Statist. Prob. Lett.* **2021**, *168*, 108923. [CrossRef]
5. Berti, P.; Pratelli, L.; Rigo, P. Limit theorems for empirical processes based on dependent data. *Electron. J. Probab.* **2012**, *17*, 1–18. [CrossRef]
6. Crimaldi, I.; Pratelli, L. Convergence results for conditional expectations. *Bernoulli* **2005**, *11*, 737–745. [CrossRef]
7. Goggin, E.M. Convergence in distribution of conditional expectations. *Ann. Probab.* **1994**, *22*, 1097–1114. [CrossRef]

8.   Lan, G.; Hu, Z.C.; Sun, W. Products of conditional expectation operators: Convergence and divergence. *J. Theore. Probab.* **2021**, *34*, 1012–1028. [CrossRef]

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

10.   Berti, P.; Crimaldi, I.; Pratelli, L.; Rigo, P. A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.* **2011**, *48*, 527–546. [CrossRef]

11.   Berti, P.; Pratelli, L.; Rigo, P. Exchangeable sequences driven by an absolutely continuous random measure. *Ann. Probab.* **2013**, *41*, 2090–2102. [CrossRef]

12.   Blackwell, D.; Dubins, L.E. Merging of opinions with increasing information. *Ann. Math. Statist.* **1962**, *33*, 882–886. [CrossRef]

13.   Cifarelli, D.M.; Regazzini, E. De Finetti's contribution to probability and statistics. *Statist. Sci.* **1996**, *11*, 253–282. [CrossRef]

14.   Cifarelli, D.M.; Dolera, E.; Regazzini, E. Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Int. J. Approx. Reason.* **2016**, *78*, 138–152. [CrossRef]

15.   Dolera, E.; Regazzini, E. Uniform rates of the Glivenko-Cantelli convergence and their use in approximating Bayesian inferences. *Bernoulli* **2019**, *25*, 2982–3015. [CrossRef]

16.   Fortini, S.; Ladelli, L.; Regazzini, E. Exchangeability, predictive distributions and parametric models. *Sankhyā Indian J. Stat. Ser. A* **2000**, *62*, 86–109.

17.   Hahn, P.R.; Martin, R.; Walker, S.G. On recursive Bayesian predictive distributions. *J. Am. Stat. Assoc.* **2018**, *113*, 1085–1093. [CrossRef]

18.   Morvai, G.; Weiss, B. On universal algorithms for classifying and predicting stationary processes. *Probab. Surv.* **2021**, *18*, 77–131. [CrossRef]

19.   Pitman, J. Some developments of the Blackwell-MacQueen urn scheme. *Stat. Probab. Game Theory IMS Lect. Notes Mon. Ser.* **1996**, *30*, 245–267.

20.   Pitman, J. *Combinatorial Stochastic Processes*; Lectures from the XXXII Summer School in Saint-Flour; Springer: Berlin/Heidelberg, Germany, 2006.

21.   Regazzini, E. Old and recent results on the relationship between predictive inference and statistical modeling either in nonparametric or parametric form. In *Bayesian Statistics 6*; Oxford University Press: Oxford, UK, 1999; pp. 571–588.

22.   Kallenberg, O. Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.* **1988**, *16*, 508–534. [CrossRef]

23.   Airoldi, E.M.; Costa, T.; Bassetti, F.; Leisen, F.; Guindani, M. Generalized species sampling priors with latent beta reinforcements. *J. Am. Stat. Assoc.* **2014**, *109*, 1466–1480. [CrossRef]

24.   Bassetti, F.; Crimaldi, I.; Leisen, F. Conditionally identically distributed species sampling sequences. *Adv. Appl. Probab.* **2010**, *42*, 433–459. [CrossRef]

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

26.   Fong, E.; Holmes, C.; Walker, S.G. Martingale posterior distributions. *arXiv* **2021**, arXiv:2103.15671v1.

27.   Fortini, S.; Petrone, S. Predictive construction of priors in Bayesian nonparametrics. *Braz. J. Probab. Statist.* **2012**, *26*, 423–449. [CrossRef]

28.   Fortini, S.; Petrone, S.; Sporysheva, P. On a notion of partially conditionally identically distributed sequences. *Stoch. Proc. Appl.* **2018**, *128*, 819–846. [CrossRef]

29.   Fortini, S.; Petrone, S. Quasi-Bayes properties of a procedure for sequential learning in mixture models. *J. R. Stat. Soc.* B **2020**, *82*, 1087–1114. [CrossRef]

30.   Berti, P.; Dreassi, E.; Leisen, F.; Pratelli, L.; Rigo, P. Kernel based Dirichlet sequences. *arXiv* **2021**, arXiv:2106.00114.