Bernoulli **29**(2), 2023, 1321–1342
https://doi.org/10.3150/22-BEJ1500

Kernel based Dirichlet sequences

PATRIZIA BERTI[1,a], EMANUELA DREASSI[2,b], FABRIZIO LEISEN[3,c], LUCA PRATELLI[4,d] and PIETRO RIGO[5,e]

[1]*Dipartimento di Matematica Pura ed Applicata "G. Vitali", Università di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy,* [a]*patrizia.berti@unimore.it*
[2]*Dipartimento di Statistica, Informatica, Applicazioni, Università di Firenze, viale Morgagni 59, 50134 Firenze, Italy,* [b]*emanuela.dreassi@unifi.it*
[3]*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK,* [c]*fabrizio.leisen@gmail.com*
[4]*Accademia Navale, viale Italia 72, 57100 Livorno, Italy,* [d]*pratel@mail.dm.unipi.it*
[5]*Dipartimento di Scienze Statistiche "P. Fortunati", Università di Bologna, via delle Belle Arti 41, 40126 Bologna, Italy,* [e]*pietro.rigo@unibo.it*

Let $X = (X_1, X_2, \ldots)$ be a sequence of random variables with values in a standard space $(S, \mathcal{B})$. Suppose

$$X_1 \sim \nu \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big) = \frac{\theta \nu(\cdot) + \sum_{i=1}^n K(X_i)(\cdot)}{n+\theta} \qquad \text{a.s.}$$

where $\theta > 0$ is a constant, ν a probability measure on $\mathcal{B}$, and K a random probability measure on $\mathcal{B}$. Then, X is exchangeable whenever K is a regular conditional distribution for ν given any sub-σ-field of $\mathcal{B}$. Under this assumption, X enjoys all the main properties of classical Dirichlet sequences, including Sethuraman's representation, conjugacy property, and convergence in total variation of predictive distributions. If μ is the weak limit of the empirical measures, conditions for μ to be a.s. discrete, or a.s. non-atomic, or $\mu \ll \nu$ a.s., are provided. Two CLT's are proved as well. The first deals with stable convergence while the second concerns total variation distance.

Keywords: Bayesian nonparametrics; central limit theorem; Dirichlet sequence; exchangeability; predictive distribution; random probability measure; regular conditional distribution

1. Introduction

Throughout, S is a Borel subset of a Polish space and $\mathcal{B}$ the Borel σ-field on S. All random elements are defined on a common probability space, say $(\Omega, \mathcal{A}, P)$. Moreover,

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

is a sequence of random variables with values in $(S, \mathcal{B})$ and

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n).$$

We say that X is a *Dirichlet sequence*, or a *Polya sequence*, if its predictive distributions are of the form

$$P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\, P(X_1 \in \cdot) + \sum_{i=1}^n \delta_{X_i}(\cdot)}{n+\theta} \qquad \text{a.s.}$$

for all $n \geq 1$ and some constant $\theta > 0$. The finite measure $\theta\, P(X_1 \in \cdot)$ is called the *parameter* of X. Here and in the sequel, for each $x \in S$, we denote by δ_x the unit mass at x.

1350-7265

Let $\mathcal{L}_0$ be the class of Dirichlet sequences. As it can be guessed from the definition, each element of $\mathcal{L}_0$ is *exchangeable*. We recall that X is exchangeable if

$$\pi(X_1,\dots,X_n) \sim (X_1,\dots,X_n) \quad \text{for all } n \geq 2 \text{ and all permutations } \pi \text{ of } S^n.$$

A permutation of S^n is meant as a map $\pi : S^n \to S^n$ of the form

$$\pi(x_1,\dots,x_n) = (x_{j_1},\dots,x_{j_n}) \qquad \text{for all } (x_1,\dots,x_n) \in S^n,$$

where $(j_1,\dots,j_n)$ is a fixed permutation of $(1,\dots,n)$. An i.i.d. sequence is obviously exchangeable while the converse is not true. However, the distribution of an exchangeable sequence (with values in a standard space) is a mixture of the distributions of i.i.d. sequences; see Subsection 1.2.

Since Ferguson, Blackwell and Mac Queen, $\mathcal{L}_0$ played a prevailing role in Bayesian statistics. It was for a long time the basic ingredient of Bayesian nonparametrics. And still today, the Bayesian nonparametrics machinery is greatly affected by $\mathcal{L}_0$ and its developments. In addition, $\mathcal{L}_0$ plays a role in various other settings, including population genetics and species sampling. The literature on $\mathcal{L}_0$ is huge and we do not try to summarize it. Without any claim of being exhaustive, we mention a few seminal papers and recent textbooks: [1,8,10,12,14,18,21–24].

The object of this paper is a new class of exchangeable sequences, say $\mathcal{L}$, such that $\mathcal{L} \supset \mathcal{L}_0$. There are essentially two reasons for taking $\mathcal{L}$ into account. First, all main features of $\mathcal{L}_0$ are preserved by $\mathcal{L}$, including the Sethuraman's representation, the conjugacy property and the simple form of predictive distributions. Thus, from the point of view of a Bayesian statistician, $\mathcal{L}$ can be handled as simply as $\mathcal{L}_0$. Second, $\mathcal{L}$ is more flexible than $\mathcal{L}_0$ and allows to model more real situations. For instance, if $X \in \mathcal{L}$, the weak limit of the empirical measures is not forced to be a.s. discrete, but it may be a.s. non-atomic or even a.s. absolutely continuous with respect to a reference measure.

1.1. Definition of $\mathcal{L}$

Obviously, the notion of Dirichlet sequence can be extended in various ways. In this paper, for X to be an extended Dirichlet sequence, two conditions are essential. First, X should be exchangeable. Second, the predictive distributions of X should have a known (and possibly simple) structure. Indeed, to define a sequence X via its predictive distributions has various merits. It is technically convenient (see the proof of Theorem 13) and makes the dynamics of X explicit. Furthermore, having the predictive distributions in closed form makes straightforward the Bayesian predictive inference on X; see e.g. [3] and [15]. We also note that, as claimed in [16]: "There are very few models for exchangeable sequences X with an explicit prediction rule".

Let $\mathcal{P}$ be the collection of all probability measures on $\mathcal{B}$ and $\mathcal{C}$ the σ-field over $\mathcal{P}$ generated by the maps $p \mapsto p(A)$ for all $A \in \mathcal{B}$. A *kernel* on $(S,\mathcal{B})$ is a measurable map $K : (S,\mathcal{B}) \to (\mathcal{P},\mathcal{C})$. Thus, $K(x) \in \mathcal{P}$ for each $x \in S$ and $x \mapsto K(x)(A)$ is a $\mathcal{B}$-measurable map for fixed $A \in \mathcal{B}$. Here, $K(x)(A)$ denotes the value attached to the event A by the probability measure $K(x)$. (This notation is possibly heavy but suitable for this paper).

A quite natural extension of $\mathcal{L}_0$, among the possible ones, consists in replacing δ with any kernel K in the predictive distributions of X. If K is arbitrary, however, X may fail to be exchangeable.

More precisely, fix $\nu \in \mathcal{P}$, a constant $\theta > 0$ and a kernel K on $(S,\mathcal{B})$. By the Ionescu-Tulcea theorem, there is a sequence X such that

$$X_1 \sim \nu \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu(\cdot) + \sum_{i=1}^n K(X_i)(\cdot)}{n+\theta} \qquad \text{a.s.} \tag{1}$$

for all $n \geq 1$. Generally, however, X is not exchangeable. As an obvious example, take the trivial kernel $K(x) = \nu^*$ for all $x \in S$, where $\nu^* \in \mathcal{P}$ but $\nu^* \neq \nu$. Then, condition (1) implies that X_2 is not distributed as X_1.

Our starting point is that, for X to be exchangeable, it suffices condition (1) and

$$K \text{ is a regular conditional distribution (r.c.d.) for } \nu \text{ given } \mathcal{G} \tag{2}$$

for some sub-σ-field $\mathcal{G} \subset \mathcal{B}$. We recall that K is a r.c.d. for ν given $\mathcal{G}$ if $K(x) \in \mathcal{P}$ for each $x \in S$, the map $x \mapsto K(x)(A)$ is $\mathcal{G}$-measurable for each $A \in \mathcal{B}$, and

$$\nu(A \cap G) = \int_G K(x)(A)\, \nu(dx) \qquad \text{for all } A \in \mathcal{B} \text{ and } G \in \mathcal{G}.$$

Equivalently, K is a r.c.d. for ν given $\mathcal{G}$ if $K(x) \in \mathcal{P}$ for each $x \in S$ and

$$K(\cdot)(A) = E_\nu(1_A \mid \mathcal{G}), \quad \nu\text{-a.s.,} \quad \text{for all } A \in \mathcal{B}.$$

Since $(S, \mathcal{B})$ is a standard space, for any sub-σ-field $\mathcal{G} \subset \mathcal{B}$, a r.c.d. for ν given $\mathcal{G}$ exists and is ν-essentially unique. See e.g. [7] for more information on r.c.d.'s.

Condition (2) makes the next definition operational.

Say that X is a *kernel based Dirichlet sequence* if it is exchangeable and satisfies condition (1) for some $\nu \in \mathcal{P}$, some constant $\theta > 0$ and some kernel K on $(S, \mathcal{B})$. In particular, X is a kernel based Dirichlet sequence if conditions (1)-(2) hold. In the sequel, $\mathcal{L}$ denotes the collection of all X satisfying conditions (1)-(2).

If $X \in \mathcal{L}$ and $\mathcal{G} = \mathcal{B}$, then $K = \delta$ and $X \in \mathcal{L}_0$. At the opposite extreme, if $\mathcal{G} = \{\emptyset, S\}$, then $K(x) = \nu$ for ν-almost all $x \in S$ and X is i.i.d. Various other examples come soon to the fore. The following are from [3] (even if, when writing [3], we didn't know yet that X is exchangeable).

Example 1. Let $\mathcal{G} = \sigma(\mathcal{H})$, where $\mathcal{H} \subset \mathcal{B}$ is a countable partition of S such that $\nu(H) > 0$ for all $H \in \mathcal{H}$. A r.c.d. for ν given $\mathcal{G}$ is

$$K(x) = \sum_{H \in \mathcal{H}} 1_H(x)\, \nu(\cdot \mid H) = \nu\big[\cdot \mid H(x)\big]$$

where $H(x)$ denotes the only $H \in \mathcal{H}$ such that $x \in H$. Therefore, $X \in \mathcal{L}$ whenever

$$X_1 \sim \nu \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta \nu(\cdot) + \sum_{i=1}^n \nu\big[\cdot \mid H(X_i)\big]}{n + \theta} \qquad \text{a.s.}$$

Note that

$$P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) \ll \nu(\cdot) \qquad \text{a.s.}$$

This fact highlights a sticking difference between $\mathcal{L}$ and $\mathcal{L}_0$. In this example, if ν is non-atomic, the probability distributions of X and Y are singular for any $Y \in \mathcal{L}_0$.

Example 2. Let $S = \mathbb{R}^2$ and $\mathcal{G} = \sigma(f)$ where $f(u, v) = u$ for all $(u, v) \in \mathbb{R}^2$. Let $\mathcal{B}_0$ be the Borel σ-field on $\mathbb{R}$ and $\mathcal{N}(u, 1)$ the Gaussian law on $\mathcal{B}_0$ with mean u and variance 1. Fix a probability measure r on $\mathcal{B}_0$ and define

$$\nu(A \times B) = \int_A \mathcal{N}(u, 1)(B)\, r(du) \qquad \text{for all } A,\, B \in \mathcal{B}_0$$

where $\mathcal{N}(u,1)(B)$ denotes the value attached to B by $\mathcal{N}(u,1)$. Then, a r.c.d. for ν given $\mathcal{G}$ is

$$K(u,v) = \delta_u \times \mathcal{N}(u,1) \qquad \text{for all } (u,v) \in \mathbb{R}^2.$$

Hence, letting $X_i = (U_i, V_i)$, one obtains $X \in \mathcal{L}$ provided $(U_1, V_1) \sim \nu$ and

$$P\big(U_{n+1} \in A, V_{n+1} \in B \mid \mathcal{F}_n\big) = \frac{\theta\nu(A \times B) + \sum_{i=1}^n 1_A(U_i)\,\mathcal{N}(U_i,1)(B)}{n+\theta} \qquad \text{a.s.}$$

Example 3. Let $f : S \to S$ be a measurable map. If ν is f-invariant, that is $\nu = \nu \circ f^{-1}$, it may be reasonable to take

$$\mathcal{G} = \big\{A \in \mathcal{B} : f^{-1}(A) = A\big\}.$$

As a trivial example, if $S = \mathbb{R}$, $f(x) = -x$ and ν is symmetric, then

$$K(x) = \frac{\delta_x + \delta_{-x}}{2}$$

is a r.c.d. for ν given $\mathcal{G}$. Hence, $X \in \mathcal{L}$ whenever $X_1 \sim \nu$ and

$$P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{2\,\theta\nu + \sum_{i=1}^n(\delta_{X_i} + \delta_{-X_i})}{2\,(n+\theta)} \qquad \text{a.s.}$$

This example is related to [3,9] and [17]. We will take up it again in forthcoming Example 17.

1.2. Sethuraman's representation and conjugacy for $\mathcal{L}_0$

Before going on, a few basic properties of $\mathcal{L}_0$ are to be recalled.

A *random probability measure* on $(S,\mathcal{B})$ is a measurable map $\mu : (\Omega,\mathcal{A}) \to (\mathcal{P},\mathcal{C})$.

Let X be exchangeable. Since $(S,\mathcal{B})$ is a standard space, there is a random probability measure μ on $(S,\mathcal{B})$ such that

$$\mu(A) \overset{a.s.}{=} \lim_n \frac{1}{n}\sum_{i=1}^n 1_A(X_i) \overset{a.s.}{=} \lim_n P\big(X_{n+1} \in A \mid \mathcal{F}_n\big)$$

for each fixed $A \in \mathcal{B}$. Moreover, X is i.i.d. conditionally on μ, in the sense that

$$P\big(X \in B \mid \mu\big) = \mu^\infty(B) \qquad \text{a.s. for all } B \in \mathcal{B}^\infty$$

where $\mu^\infty = \mu \times \mu \times \ldots$; see e.g. [6, p. 2090].

Suppose now that $X \in \mathcal{L}_0$ and define

$$\mathcal{D}(C) = P(\mu \in C) \qquad \text{for all } C \in \mathcal{C}.$$

Such a $\mathcal{D}$ is a probability measure on $\mathcal{C}$, called the *Dirichlet prior*, and admits the following representation. Define a random probability measure μ^* on $(S,\mathcal{B})$ as

$$\mu^* = \sum_j V_j\,\delta_{Z_j},$$

where (Z_j) and (V_j) are independent sequences, (Z_j) is i.i.d. with $Z_1 \sim \nu$, and (V_j) has the stick-breaking distribution with parameter θ; see Section 2. Then,

$$\mathcal{D}(C) = P(\mu^* \in C) \qquad \text{for all } C \in \mathcal{C}.$$

Thus, $\mathcal{D}$ can be also regarded as the probability distribution of μ^*. This fact, proved by Sethuraman [23], is fundamental in applications; see e.g. [11].

Finally, we recall the conjugacy property of $\mathcal{L}_0$. Write $\mathcal{D}(\lambda)$ (instead of $\mathcal{D}$) if $X \in \mathcal{L}_0$ has parameter λ. In this notation, if X has parameter $\theta\nu$, then

$$P(\mu \in C \mid \mathcal{F}_n) = \mathcal{D}\Big(\theta\nu + \sum_{i=1}^{n} \delta_{X_i}\Big)(C) \qquad \text{a.s. for all } C \in \mathcal{C} \text{ and } n \geq 1.$$

Roughly speaking, the posterior distribution of μ given $(X_1, \ldots, X_n)$ is still of the Dirichlet type but the parameter turns into $\theta\nu + \sum_{i=1}^{n} \delta_{X_i}$. Once again, this fact plays a basic role in applications.

1.3. Our contribution

As claimed above, this paper aims to introduce and investigate the class $\mathcal{L}$.

Our first result is that conditions (1)-(2) suffice for exchangeability of X. Thus, each $X \in \mathcal{L}$ is a kernel based Dirichlet sequence, as defined in Subsection 1.1.

The next step is to develop some theory for $\mathcal{L}$. The obvious hope is that, at least to a certain extent, such a theory is parallel to that of $\mathcal{L}_0$. This is exactly the case. Essentially all main results concerning $\mathcal{L}_0$ extend nicely to $\mathcal{L}$. To illustrate, we assume $X \in \mathcal{L}$ and we mention a few facts.

- Up to replacing δ with K, the Sethuraman's representation remains exactly the same. Precisely, $P(\mu \in C) = P(\mu^* \in C)$ for all $C \in \mathcal{C}$, where

  $$\mu^* = \sum_j V_j\, K(Z_j)$$

  and (V_j) and (Z_j) are as in Subsection 1.2.
- The predictive distributions converge in total variation, that is

  $$\sup_{A \in \mathcal{B}} \Big| P\big(X_{n+1} \in A \mid \mathcal{F}_n\big) - \mu(A) \Big| \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty.$$

- If $X \in \mathcal{L}_0$, it is well known that μ is a.s. discrete. This result extends to $\mathcal{L}$ as follows. Denote by D_1, D_2, D_3 the collections of elements of $\mathcal{P}$ which are, respectively, discrete, non-atomic, or absolutely continuous with respect to ν. Then, for each $1 \leq j \leq 3$,

  $$P(\mu \in D_j) = 1 \quad \Leftrightarrow \quad K(x) \in D_j \text{ for } \nu\text{-almost all } x \in S.$$

  Since $\delta_x \in D_1$ for all $x \in S$, the classical result is recovered. But now, with a suitable K, one obtains $P(\mu \in D_2) = 1$ or $P(\mu \in D_3) = 1$. This fact may be useful in applications.
- The conjugacy property of $\mathcal{L}_0$ is still available. For each $n \geq 1$, let

  $$V^{(n)} = \big(V_j^{(n)} : j \geq 1\big) \quad \text{and} \quad Z^{(n)} = \big(Z_j^{(n)} : j \geq 1\big)$$

  be two sequences such that

(i) $V^{(n)}$ and $Z^{(n)}$ are conditionally independent given $\mathcal{F}_n$;
(ii) $V^{(n)}$ has the stick-breaking distribution, with parameter $n+\theta$, conditionally on $\mathcal{F}_n$;
(iii) $Z^{(n)}$ is i.i.d., conditionally on $\mathcal{F}_n$, with

$$P(Z_1^{(n)} \in \cdot \mid \mathcal{F}_n) = P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu(\cdot) + \sum_{i=1}^n K(X_i)(\cdot)}{n+\theta} \qquad \text{a.s.}$$

Then,

$$P(\mu \in \cdot \mid \mathcal{F}_n) = P(\mu_n^* \in \cdot \mid \mathcal{F}_n)$$

where

$$\mu_n^* = \sum_j V_j^{(n)} K\big(Z_j^{(n)}\big).$$

Again, if $K=\delta$, this result reduces to the classical one.

- A stable CLT holds true. Let $S=\mathbb{R}^p$ and $\int \|x\|^2\,\nu(dx) < \infty$, where $\|\cdot\|$ is the Euclidean norm. Suppose that K has mean 0, in the sense that

$$\int y_i\, K(x)(dy) = 0 \qquad \text{for all } x \in \mathbb{R}^p \text{ and } i = 1,\dots,p$$

where y_i denotes the i-th coordinate of a point $y \in \mathbb{R}^p$. Then, $n^{-1/2}\sum_{i=1}^n X_i$ converges stably (in particular, in distribution) to the Gaussian kernel $\mathcal{N}_p(0,\Sigma)$, where Σ is the (random) covariance matrix

$$\Sigma = \left(\int y_i\, y_j\, \mu(dy) : 1 \le i, j \le p \right).$$

Moreover, under some additional conditions, $n^{-1/2}\sum_{i=1}^n X_i$ converges in total variation as well.

This is a brief summary of our main results. Before closing the introduction, however, two remarks are in order.

First, to prove such results, we often exploit the fact that

$$\big(K(X_n) : n \ge 1\big) \quad \text{is a classical Dirichlet sequence with values in } (\mathcal{P}, \mathcal{C}). \qquad (3)$$

Condition (3) is not surprising. We give a simple proof of it, based on predictive distributions, but condition (3) could be also obtained via some known results on $\mathcal{L}_0$.

Second, the above results are potentially useful in Bayesian nonparametrics. Define in fact

$$\Pi(C) = P(\mu \in C) = P(\mu^* \in C) \qquad \text{for all } C \in \mathcal{C}.$$

Such a Π is a new prior to be used in Bayesian nonparametrics. In real problems, working with Π is as simple as working with the classical Dirichlet prior $\mathcal{D}$. In both cases, the posterior can be easily evaluated. Unlike $\mathcal{D}$, however, Π can be chosen such that $\Pi(C)=1$ for some meaningful sets C of probability measures. For instance, $C=D_j$ with D_j defined as above for $j=1,2,3$. Or else, C the set of invariant probability measures under a countable class of measurable transformations; see forthcoming Example 17. Finally, just because of its definition, $\mathcal{L}$ is particularly suitable in Bayesian predictive inference. And predicting future observations is one of the main tasks of Bayesian nonparametrics.

2. Preliminaries

For all $\lambda \in \mathcal{P}$ and bounded measurable $f : S \to \mathbb{R}$, the notation $\lambda(f)$ stands for $\lambda(f) = \int f\,d\lambda$. Moreover, $\mathcal{N}_p(0,\Sigma)$ denotes the p-dimensional Gaussian law (on the Borel σ-field of $\mathbb{R}^p$) with mean 0 and covariance matrix Σ.

Let $\theta > 0$ be a constant, (W_n) an i.i.d. sequence with $W_1 \sim \text{beta}(1,\theta)$ and

$$T_1 = W_1, \quad T_n = W_n \prod_{i=1}^{n-1}(1 - W_i) \text{ for } n > 1.$$

A sequence (V_n) of real random variables has the *stick-breaking distribution with parameter* θ if $(V_n) \sim (T_n)$. Note that $V_n > 0$ for all n and $\sum_n V_n = 1$ a.s.

Stable convergence is a strong form of convergence in distribution. Let N be a random probability measure on $(S,\mathcal{B})$. Then, X_n *converges to* N *stably* if

$$E\big[N(f) \mid H\big] = \lim_n E\big[f(X_n) \mid H\big]$$

for all bounded continuous $f : S \to \mathbb{R}$ and all $H \in \mathcal{A}$ with $P(H) > 0$. In particular, X_n converges in distribution to the probability measure $A \mapsto E\big[N(A)\big]$.

We next report an useful characterization of exchangeability due to [13]; see also [5] and [3]. Let $\mathcal{F}_0 = \{\emptyset,\Omega\}$ be the trivial σ-field and

$$\sigma_n(x) = P\big[X_{n+1} \in \cdot \mid (X_1,\dots,X_n) = x\big] \qquad \text{for all } x \in S^n.$$

Theorem 4 ([13, Theorem 3.1]). *The sequence X is exchangeable if and only if*

$$P\big[(X_{n+1},X_{n+2}) \in \cdot \mid \mathcal{F}_n\big] = P\big[(X_{n+2},X_{n+1}) \in \cdot \mid \mathcal{F}_n\big] \qquad a.s.$$

for all $n \ge 0$ and

$$\sigma_n(x) = \sigma_n(\pi(x))$$

for all $n \ge 2$, all permutations π of S^n, and almost all $x \in S^n$. (Here, "almost all" is with respect to the marginal distribution of $(X_1,\dots,X_n)$).

We conclude this section with two technical lemmas. Let

$$\sigma(K) = \big\{\{x \in S : K(x) \in C\} : C \in \mathcal{C}\big\}$$

be the σ-field over S generated by the kernel K.

Lemma 5 (Lemma 10 of [7]). *Under condition (2), there is a set $F \in \sigma(K)$ such that $\nu(F) = 1$ and*

$$K(x)(B) = \delta_x(B) \qquad \textit{for all } B \in \sigma(K) \textit{ and } x \in F.$$

Proof. This is basically [7, Lem. 10] but we give a proof to make the paper self-contained. The atoms of the σ-field $\sigma(K)$ are sets of the form

$$B(x) = \big\{y \in S : K(y) = K(x)\big\} \qquad \text{for all } x \in S.$$

Hence, each $B \in \sigma(K)$ can be written as

$$B = \bigcup_{x \in B} B(x).$$

Moreover, by [7, Lem. 10], there is a set $F \in \sigma(K)$ such that $\nu(F) = 1$ and

$$K(x)\big(B(x)\big) = 1 \qquad \text{for all } x \in F.$$

Having noted these facts, fix $x \in F$ and $B \in \sigma(K)$. If $x \in B$, then

$$K(x)(B) \geq K(x)\big(B(x)\big) = 1.$$

If $x \notin B$, since $B^c \in \sigma(K)$, then $K(x)(B) = 1 - K(x)(B^c) = 0$. Hence, $K(x)(B) = \delta_x(B)$. □

Lemma 6. *Under condition (2), there is a set $F \in \sigma(K)$ such that $\nu(F) = 1$ and*

$$\int_A K(y)(B)\,K(x)(dy) = K(x)(A)\,K(x)(B) \qquad \textit{for all } x \in F \textit{ and } A, B \in \mathcal{B}.$$

Moreover,

$$\int_A K(y)(B)\,\nu(dy) = \int_B K(y)(A)\,\nu(dy) \qquad \textit{for all } A, B \in \mathcal{B}.$$

Proof. Let F be as in Lemma 5. Fix $x \in F$ and $A, B \in \mathcal{B}$. Define

$$G = \big\{y \in S : K(y)(B) = K(x)(B)\big\}$$

and note that $x \in G$ and $G \in \sigma(K)$. Since $x \in G$, then $\delta_x(G) = 1$. Since $G \in \sigma(K)$ and $x \in F$, Lemma 5 implies

$$K(x)(G) = \delta_x(G) = 1.$$

Therefore,

$$\int_A K(y)(B)\,K(x)(dy) = K(x)(B) \int_A K(x)(dy) = K(x)(A)\,K(x)(B).$$

Finally,

$$\int_A K(y)(B)\,\nu(dy) = \int_A E_\nu(1_B \mid \mathcal{G})\,d\nu = \int_B E_\nu(1_A \mid \mathcal{G})\,d\nu = \int_B K(y)(A)\,\nu(dy).$$ □

3. Results

Recall that $\mathcal{L}$ is the class of sequences satisfying conditions (1)-(2) for some $\nu \in \mathcal{P}$ and some constant $\theta > 0$. In this section, $X \in \mathcal{L}$ and μ is a random probability measure on $(S, \mathcal{B})$ such that

$$\mu(A) \overset{a.s.}{=} \lim_n \frac{1}{n} \sum_{i=1}^n 1_A(X_i) \overset{a.s.}{=} \lim_n P\big(X_{n+1} \in A \mid \mathcal{F}_n\big) \qquad \text{for all } A \in \mathcal{B}.$$

Existence of μ depends on X is exchangeable and $(S, \mathcal{B})$ is a standard space; see Subsection 1.2.

Our starting point is the following.

Theorem 7. *Under condition* (1)*, X is exchangeable if and only if*

$$\int_A K(y)(B)\,\nu(dy) = \int_B K(y)(A)\,\nu(dy) \tag{4}$$

and

$$\int_A K(y)(B)\,K(x)(dy) = \int_B K(y)(A)\,K(x)(dy) \tag{5}$$

for all $A, B \in \mathcal{B}$ and ν-almost all $x \in S$. In particular, X is exchangeable whenever $X \in \mathcal{L}$ (because of Lemma 6).

Proof. For all $A, B \in \mathcal{B}$, condition (1) implies

$$\begin{aligned} P(X_1 \in A,\, X_2 \in B) &= E\Big\{1_A(X_1)\,P(X_2 \in B \mid \mathcal{F}_1)\Big\} \\ &= E\Big\{1_A(X_1)\,\frac{\theta\nu(B) + K(X_1)(B)}{1+\theta}\Big\} \\ &= \frac{\theta}{1+\theta}\,\nu(B)\,\nu(A) \,+\, \frac{1}{1+\theta}\int_A K(y)(B)\,\nu(dy). \end{aligned}$$

Therefore,

$$\text{condition (4)} \quad \Longleftrightarrow \quad (X_1, X_2) \sim (X_2, X_1).$$

Similarly, under (1), one obtains

$$\begin{aligned} P\big(X_2 \in A,\, X_3 \in B \mid \mathcal{F}_1\big) &= E\Big\{1_A(X_2)\,P(X_3 \in B \mid \mathcal{F}_2) \mid \mathcal{F}_1\Big\} \\ &= E\Big\{1_A(X_2)\,\frac{\theta\nu(B) + K(X_1)(B) + K(X_2)(B)}{2+\theta} \mid \mathcal{F}_1\Big\} \\ &= \frac{1+\theta}{2+\theta}\,P(X_2 \in B \mid \mathcal{F}_1)\,P(X_2 \in A \mid \mathcal{F}_1) \,+\, \frac{1}{2+\theta}\,E\Big\{1_A(X_2)\,K(X_2)(B) \mid \mathcal{F}_1\Big\} \quad \text{a.s.} \end{aligned}$$

and

$$E\Big\{1_A(X_2)\,K(X_2)(B) \mid \mathcal{F}_1\Big\} = \frac{\theta}{1+\theta}\int_A K(y)(B)\,\nu(dy) + \frac{1}{1+\theta}\int_A K(y)(B)\,K(X_1)(dy) \quad \text{a.s.}$$

Next, if X is exchangeable, condition (4) follows from $(X_1, X_2) \sim (X_2, X_1)$. Moreover, $P\big(X_2 \in A,\, X_3 \in B \mid \mathcal{F}_1\big) = P\big(X_2 \in B,\, X_3 \in A \mid \mathcal{F}_1\big)$ a.s. implies

$$E\Big\{1_A(X_2)\,K(X_2)(B) \mid \mathcal{F}_1\Big\} = E\Big\{1_B(X_2)\,K(X_2)(A) \mid \mathcal{F}_1\Big\} \quad \text{a.s.}$$

Therefore, (5) follows from (4) and the above condition.

Conversely, assume conditions (4)-(5). Define

$$\sigma_n(x) = \frac{\theta\nu + \sum_{i=1}^n K(x_i)}{n+\theta} \qquad \text{for all } n \geq 1 \text{ and } x = (x_1, \ldots, x_n) \in S^n.$$

By (1), $P(X_{n+1} \in \cdot \mid \mathcal{F}_n) = \sigma_n(X_1, \dots, X_n)$ a.s. Moreover, $\sigma_n(x) = \sigma_n(\pi(x))$ for all $n \geq 2$, all permutations π of S^n and all $x \in S^n$. Hence, by Theorem 4, it suffices to show that

$$P\big[(X_{n+1}, X_{n+2}) \in \cdot \mid \mathcal{F}_n\big] = P\big[(X_{n+2}, X_{n+1}) \in \cdot \mid \mathcal{F}_n\big] \qquad \text{a.s. for all } n \geq 0.$$

For $n = 0$, the above condition is equivalent to (4) (recall that $\mathcal{F}_0$ is the trivial σ-field). Therefore, it is enough to show that

$$\int_A \sigma_{n+1}(x,y)(B)\,\sigma_n(x)(dy) = \int_B \sigma_{n+1}(x,y)(A)\,\sigma_n(x)(dy) \tag{6}$$

for all $n \geq 1$, all $A, B \in \mathcal{B}$ and almost all $x \in S^n$ (where "almost all" refers to the marginal distribution of $(X_1, \dots, X_n)$).

Fix $m \geq 1$ and $A \in \mathcal{B}$. If $X_i \sim \nu$ for $i = 1, \dots, m$, then

$$E\Big\{K(X_i)(A)\Big\} = \int K(y)(A)\,\nu(dy) = \nu(A) \quad \text{for } i = 1, \dots, m,$$

where the second equality is by (4) (applied with $B = S$). Hence,

$$P(X_{m+1} \in A) = E\Big\{P(X_{m+1} \in A \mid \mathcal{F}_m)\Big\} = \frac{\theta\nu(A)}{m+\theta} + \frac{\sum_{i=1}^m E\Big\{K(X_i)(A)\Big\}}{m+\theta} = \nu(A).$$

By induction, it follows that $X_i \sim \nu$ for all $i \geq 1$.

Finally, fix $n \geq 1$ and $A, B \in \mathcal{B}$. By (5), there is a set $M \in \mathcal{B}$ such that $\nu(M) = 1$ and

$$\int_A K(y)(B)\,K(x)(dy) = \int_B K(y)(A)\,K(x)(dy) \qquad \text{for all } x \in M.$$

Thanks to this fact and condition (4), if $x = (x_1, \dots, x_n) \in M^n$, one obtains

$$\int_A K(y)(B)\,\sigma_n(x)(dy) = \frac{\theta \int_A K(y)(B)\,\nu(dy) + \sum_{i=1}^n \int_A K(y)(B)\,K(x_i)(dy)}{n+\theta}$$
$$= \frac{\theta \int_B K(y)(A)\,\nu(dy) + \sum_{i=1}^n \int_B K(y)(A)\,K(x_i)(dy)}{n+\theta} = \int_B K(y)(A)\,\sigma_n(x)(dy).$$

It follows that

$$\int_A \sigma_{n+1}(x,y)(B)\,\sigma_n(x)(dy) = \int_A \frac{\theta\nu(B) + \sum_{i=1}^n K(x_i)(B) + K(y)(B)}{n+1+\theta}\,\sigma_n(x)(dy)$$
$$= \frac{n+\theta}{n+1+\theta}\,\sigma_n(x)(B)\,\sigma_n(x)(A) + \frac{\int_A K(y)(B)\,\sigma_n(x)(dy)}{n+1+\theta}$$
$$= \frac{n+\theta}{n+1+\theta}\,\sigma_n(x)(B)\,\sigma_n(x)(A) + \frac{\int_B K(y)(A)\,\sigma_n(x)(dy)}{n+1+\theta}$$
$$= \int_B \sigma_{n+1}(x,y)(A)\,\sigma_n(x)(dy).$$

Therefore, equation (6) holds for each $x \in M^n$. To conclude the proof, it suffices to note that, since $\nu(M) = 1$ and $X_i \sim \nu$ for all i,

$$P\big((X_1,\dots,X_n) \in M^n\big) = 1. \qquad \square$$

In view of Theorem 7, X is a kernel based Dirichlet sequence, as defined in Subsection 1.1, if and only if conditions (1) and (4)-(5) hold. Since (2) $\Rightarrow$ (4)-(5) (because of Lemma 6), a sufficient condition for X to be a kernel based Dirichlet sequence is that $X \in \mathcal{L}$. We do not know whether (4)-(5) $\Rightarrow$ (2). In the sequel, however, we always assume $X \in \mathcal{L}$, namely, we always assume conditions (1)-(2).

The next step is to develop some theory for $\mathcal{L}$. To this end, the following result is useful.

Theorem 8. *If $X \in \mathcal{L}$, the sequence $\big(K(X_n) : n \geq 1\big)$ is a Dirichlet sequence with values in $(\mathcal{P}, \mathcal{C})$ and parameter the image measure $\theta\, \nu \circ K^{-1}$.*

Proof. By Lemma 5, there is a set $F \in \sigma(K)$ such that

$$\nu(F) = 1 \quad \text{and} \quad K(x)(B) = \delta_x(B) \qquad \text{for all } B \in \sigma(K) \text{ and } x \in F.$$

Since $P(X_n \in F) = \nu(F) = 1$ for all n, it follows that

$$P\big(X_{n+1} \in\ B \mid \mathcal{F}_n\big) = \frac{\theta\nu(B) + \sum_{i=1}^n \delta_{X_i}(B)}{n+\theta} \qquad \text{for all } B \in \sigma(K) \text{ a.s.}$$

Having noted this fact, define

$$\mathcal{K}_n = \sigma\big[K(X_1),\dots,K(X_n)\big].$$

Since $\mathcal{K}_n \subset \mathcal{F}_n$ and $P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big)$ is $\mathcal{K}_n$-measurable,

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

Finally, fix $C \in \mathcal{C}$ and define $B = \{K \in C\}$. Since $B \in \sigma(K)$, one obtains

$$P\Big[K(X_{n+1}) \in C \mid \mathcal{K}_n\Big] = P\Big(X_{n+1} \in B \mid \mathcal{K}_n\Big) = P\Big(X_{n+1} \in B \mid \mathcal{F}_n\Big)$$
$$= \frac{\theta\nu(B) + \sum_{i=1}^n \delta_{X_i}(B)}{n+\theta} = \frac{\theta\nu \circ K^{-1}(C) + \sum_{i=1}^n \delta_{K(X_i)}(C)}{n+\theta} \qquad \text{a.s.} \qquad \square$$

We next turn to a Sethuraman-like representation for $\mathcal{L}$. Let μ^* be the random probability measure on $(S, \mathcal{B})$ defined as

$$\mu^* = \sum_j V_j\, K(Z_j),$$

where (Z_j) and (V_j) are independent sequences, (Z_j) is i.i.d. with $Z_1 \sim \nu$, and (V_j) has the stick-breaking distribution with parameter θ; see Section 2.

Theorem 9. *If $X \in \mathcal{L}$, then*

$$P(\mu \in C) = P(\mu^* \in C) \qquad \textit{for all } C \in \mathcal{C}.$$

Proof. Let μ_0 and μ_0^* be the restrictions of μ and μ^* on $\sigma(K)$. Then, $\mu_0 \sim \mu_0^*$ by [23] and since $\big(K(X_n) : n \geq 1\big)$ is a classical Dirichlet sequence. Hence,

$$\big(\mu(g_1), \dots, \mu(g_k)\big) \sim \big(\mu^*(g_1), \dots, \mu^*(g_k)\big)$$

whenever $g_1, \dots, g_k : S \to \mathbb{R}$ are bounded and $\sigma(K)$-measurable. In addition, for fixed $A \in \mathcal{B}$, one obtains

$$\int K(x)(A)\,\mu(dx) = \lim_n \frac{\sum_{i=1}^n K(X_i)(A)}{n} = \lim_n P(X_{n+1} \in A \mid \mathcal{F}_n) = \mu(A) \qquad \text{a.s.}$$

Similarly, Lemma 6 (applied with $B = S$) implies

$$\int K(x)(A)\,K(Z_j)(dx) = K(Z_j)(A) \qquad \text{a.s. for all } j \geq 1.$$

Thus,

$$\begin{aligned} \int K(x)(A)\,\mu^*(dx) &= \sum_j V_j \int K(x)(A)\,K(Z_j)(dx) \\ &= \sum_j V_j\,K(Z_j)(A) = \mu^*(A) \qquad \text{a.s.} \end{aligned}$$

Having noted these facts, fix $k \geq 1$, $A_1, \dots, A_k \in \mathcal{B}$, and define $g_i(x) = K(x)(A_i)$ for all $x \in S$ and $i = 1, \dots, k$. Then,

$$\big(\mu(A_1), \dots, \mu(A_k)\big) \overset{a.s.}{=} \big(\mu(g_1), \dots, \mu(g_k)\big) \sim \big(\mu^*(g_1), \dots, \mu^*(g_k)\big) \overset{a.s.}{=} \big(\mu^*(A_1), \dots, \mu^*(A_k)\big).$$

This concludes the proof. □

Theorem 9 plays for $\mathcal{L}$ the same role played by [23] for $\mathcal{L}_0$. Among other things, it provides a simple way to approximate the probability distribution of μ and to obtain its posterior distribution; see forthcoming Theorem 13 and its proof. For a further implication, define

$$D_1 = \{p \in \mathcal{P} : p \text{ discrete}\}, \quad D_2 = \{p \in \mathcal{P} : p \text{ non-atomic}\}, \quad D_3 = \{p \in \mathcal{P} : p \ll \nu\}.$$

Then, Theorem 9 implies the following result.

Theorem 10. *If $j \in \{1,2,3\}$ and $X \in \mathcal{L}$, then $P(\mu \in D_j) \in \{0,1\}$ and*

$$P(\mu \in D_j) = 1 \quad \Leftrightarrow \quad K(x) \in D_j \text{ for } \nu\text{-almost all } x \in S.$$

In addition,

$$\sup_{A \in \mathcal{B}} \big| P\big(X_{n+1} \in A \mid \mathcal{F}_n\big) - \mu(A) \big| \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty. \tag{7}$$

Proof. Fix $j \in \{1,2,3\}$ and define $a_j = \nu\{x : K(x) \in D_j\}$. If $a_j = 1$, Theorem 9 yields

$$P(\mu \in D_j) = P(\mu^* \in D_j) = P\big(K(Z_i) \in D_j \text{ for all } i \geq 1\big) = 1.$$

Similarly, if $a_j < 1$,

$$P(\mu \in D_j) \le P\big(K(Z_i) \in D_j \text{ for } 1 \le i \le n\big) = a_j^n \longrightarrow 0 \qquad \text{as } n \to \infty.$$

It remains to prove (7). Define the random probability measure

$$\lambda_n = \frac{1}{n}\sum_{i=1}^n K(X_i).$$

To prove (7), it is enough to show that $\lim_n \sup_{A\in\mathcal{B}} |\lambda_n(A) - \mu(A)| \overset{a.s.}{=} 0$ and this limit relation is actually true if $X \in \mathcal{L}_0$; see e.g. [21, Prop. 11]. Hence, since $\big(K(X_n) : n \ge 1\big)$ is a classical Dirichlet sequence, one obtains

$$\sup_{A\in\sigma(K)} |\lambda_n(A) - \mu(A)| \overset{a.s.}{\longrightarrow} 0.$$

Now, we argue as in the proof of Theorem 9. Precisely, for each $A \in \mathcal{B}$, Lemma 6 (applied with $B = S$) yields

$$\int K(x)(A)\,\lambda_n(dx) = \frac{1}{n}\sum_{i=1}^n \int K(x)(A)\,K(X_i)(dx) = \frac{1}{n}\sum_{i=1}^n K(X_i)(A) = \lambda_n(A) \quad \text{a.s.}$$

Similarly, $\int K(x)(A)\,\mu(dx) = \mu(A)$ a.s. Therefore, after fixing a countable field $\mathcal{B}_0$ such that $\mathcal{B} = \sigma(\mathcal{B}_0)$, one finally obtains

$$\begin{aligned}\sup_{A\in\mathcal{B}} |\lambda_n(A) - \mu(A)| &= \sup_{A\in\mathcal{B}_0} |\lambda_n(A) - \mu(A)| \\ &\overset{a.s.}{=} \sup_{A\in\mathcal{B}_0} \left|\int K(x)(A)\,\lambda_n(dx) - \int K(x)(A)\,\mu(dx)\right| \\ &\le \sup_{A\in\sigma(K)} |\lambda_n(A) - \mu(A)| \overset{a.s.}{\longrightarrow} 0. \qquad \square\end{aligned}$$

It is worth noting that, for an arbitrary exchangeable sequence X, convergence in total variation of $P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big)$ is not guaranteed; see e.g. [6].

A further consequence of Theorem 9 is a stable CLT (stable convergence is briefly recalled in Section 2). For each $y \in \mathbb{R}^p$, let y_i denote the i-th coordinate of y.

Theorem 11. *Let $S = \mathbb{R}^p$ and $X \in \mathcal{L}$. Suppose $\int \|x\|^2\,\nu(dx) < \infty$, where $\|\cdot\|$ is the Euclidean norm, and*

$$\int y_i\,K(x)(dy) = 0 \qquad \text{for all } x \in \mathbb{R}^p \text{ and } i = 1,\dots,p.$$

Then,

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \overset{stably}{\longrightarrow} \mathcal{N}_p(0,\Sigma) \qquad \text{as } n \to \infty,$$

where Σ *is the random covariance matrix*

$$\Sigma = \left(\int y_i\, y_j\, \mu(dy) : 1 \le i, j \le p \right).$$

Proof. By standard arguments, it suffices to show that

$$\frac{\sum_{i=1}^n b' X_i}{\sqrt{n}} \overset{stably}{\longrightarrow} \mathcal{N}_1(0, b'\Sigma\, b) \qquad \text{for each } b \in \mathbb{R}^p,$$

where points of $\mathbb{R}^p$ are regarded as column vectors and b' denotes the transpose of b. Define

$$\sigma_b^2 = E\left[(b'X_1)^2 \mid \mu\right] - E(b'X_1 \mid \mu)^2.$$

For fixed $b \in \mathbb{R}^p$, one obtains

$$n^{-1/2} \sum_{i=1}^n \left\{ b'X_i - E(b'X_1 \mid \mu) \right\} \overset{stably}{\longrightarrow} \mathcal{N}_1(0, \sigma_b^2);$$

see e.g. [4, Th. 3.1] and the subsequent remark. Furthermore,

$$E(b'X_1 \mid \mu) = \int (b'y)\, \mu(dy) = \sum_{i=1}^p b_i \int y_i\, \mu(dy) \qquad \text{a.s. and}$$

$$E\left[(b'X_1)^2 \mid \mu\right] = \int (b'y)^2\, \mu(dy) = \sum_{i=1}^p \sum_{j=1}^p b_i b_j \int y_i\, y_j\, \mu(dy) = b'\Sigma\, b \quad \text{a.s.}$$

Hence, it suffices to show that $\int y_i\, \mu(dy) \overset{a.s.}{=} 0$ for all i, and this follows from Theorem 9. In fact, $\int y_i\, \mu(dy) \sim \int y_i\, \mu^*(dy)$ and

$$\int y_i\, \mu^*(dy) = \sum_j V_j \int y_i\, K(Z_j)(dy) = 0.$$

This concludes the proof. □

Theorem 11 applies to Examples 3 and 16. In fact, in Example 3, one has $p = 1$ and $K(x) = (\delta_x + \delta_{-x})/2$. Hence, $\int y\, K(x)(dy) = 0$ for all $x \in \mathbb{R}$. Example 16 is discussed below. Here, we give conditions for convergence in total variation of $n^{-1/2} \sum_{i=1}^n X_i$.

Theorem 12. *In addition to the conditions of Theorem 11, suppose that* $K(x)$ *is not singular, with respect to Lebesgue measure, for* ν*-almost all* $x \in \mathbb{R}^p$*. Define*

$$Y_n = n^{-1/2} \sum_{i=1}^n X_i \quad \textit{and} \quad \lambda(A) = E\left\{ \mathcal{N}_p(0, \Sigma^*)(A) \right\}$$

$$\textit{for all } A \in \mathcal{B}, \textit{ where} \quad \Sigma^* = \left(\int y_i\, y_j\, \mu^*(dy) : 1 \le i, j \le p \right).$$

Then,

$$\lim_n \sup_{A\in\mathcal{B}} \Big|P(Y_n \in A) - \lambda(A)\Big| = 0.$$

Proof. Let D be the collection of elements of $\mathcal{P}$ which are not singular with respect to Lebesgue measure. By Theorem 9, $P(\mu \in D) = P(\mu^* \in D) = 1$. Hence, conditionally on μ, the sequence X is i.i.d. and the common distribution μ belongs to D a.s. Arguing as in Theorem 11, one also obtains $\int y_i\, \mu(dy) = 0$ and $\int \|y\|^2 \mu(dy) < \infty$ a.s. for all i. Thus, conditionally on μ, Y_n converges to $\mathcal{N}_p(0,\Sigma)$ in total variation (see e.g. [2]) that is

$$\sup_{A\in\mathcal{B}} \Big|P(Y_n \in A \mid \mu) - \mathcal{N}_p(0,\Sigma)(A)\Big| \overset{a.s.}{\longrightarrow} 0.$$

Finally, $\Sigma \sim \Sigma^*$ implies $\lambda(\cdot) = E\big\{\mathcal{N}_p(0,\Sigma)(\cdot)\big\}$. Hence,

$$\sup_{A\in\mathcal{B}} \Big|P(Y_n \in A) - \lambda(A)\Big| = \sup_{A\in\mathcal{B}} \Big|P(Y_n \in A) - E\big\{\mathcal{N}_p(0,\Sigma)(A)\big\}\Big|$$

$$\le E\left\{\sup_{A\in\mathcal{B}} \Big|P(Y_n \in A \mid \mu) - \mathcal{N}_p(0,\Sigma)(A)\Big|\right\} \longrightarrow 0 \qquad \text{as } n\to\infty. \qquad \square$$

Our last result deals with the posterior distribution of μ. We aim to find the conditional distribution of μ given $\mathcal{F}_n = \sigma(X_1,\dots,X_n)$. To this end, for each $n \ge 1$, we denote by

$$V^{(n)} = \big(V_j^{(n)} : j \ge 1\big) \quad \text{and} \quad Z^{(n)} = \big(Z_j^{(n)} : j \ge 1\big)$$

two sequences such that:

(i) $V^{(n)}$ and $Z^{(n)}$ are conditionally independent given $\mathcal{F}_n$;
(ii) $V^{(n)}$ has the stick-breaking distribution with parameter $n+\theta$ conditionally on $\mathcal{F}_n$;
(iii) $Z^{(n)}$ is i.i.d. conditionally on $\mathcal{F}_n$ with

$$P(Z_1^{(n)} \in \cdot \mid \mathcal{F}_n) = P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu(\cdot) + \sum_{i=1}^n K(X_i)(\cdot)}{n+\theta} \qquad \text{a.s.}$$

Moreover, we let

$$\mu_n^* = \sum_j V_j^{(n)} K\big(Z_j^{(n)}\big).$$

Theorem 13. *If $X \in \mathcal{L}$, then*

$$P(\mu \in C \mid \mathcal{F}_n) = P(\mu_n^* \in C \mid \mathcal{F}_n) \qquad \textit{a.s. for all } C \in \mathcal{C} \textit{ and } n \ge 1.$$

We recall that, if $X \in \mathcal{L}_0$ and X has parameter $\theta\nu$ (i.e., if $K = \delta$) then

$$P(\mu \in C \mid \mathcal{F}_n) = \mathcal{D}\Big(\theta\nu + \sum_{i=1}^n \delta_{X_i}\Big)(C) = P(\mu_n^* \in C \mid \mathcal{F}_n) \qquad \text{a.s.}$$

Hence, Theorem 13 extends to $\mathcal{L}$ the conjugacy property of $\mathcal{L}_0$. Such a property is clearly useful as regards Bayesian statistical inference. On one hand, the Bayesian analysis of $X \in \mathcal{L}$ is as simple as that of $X \in \mathcal{L}_0$. On the other hand, $\mathcal{L}$ is able to model much more situations than $\mathcal{L}_0$. As an obvious example, for $X \in \mathcal{L}$, it may be that $P(X_i = X_j) = 0$ if $i \neq j$. See e.g. Example 1 and Theorem 10.

Theorem 13 can be proved in various ways. We report here the simplest and most direct proof. Such a proof relies on Theorem 9 and the definition of $\mathcal{L}$ in terms of predictive distributions.

Proof of Theorem 13. Throughout this proof, if X satisfies conditions (1)-(2), we say that $X \in \mathcal{L}$ and X has *parameter* $(\theta\nu, K)$.

Fix $n \geq 1$ and define the sequence

$$X^{(n)} = \big(X_i^{(n)} : i \geq 1\big) = \big(X_{n+i} : i \geq 1\big).$$

Define also the random measure

$$J_n = \theta\nu + \sum_{i=1}^{n} K(X_i).$$

It suffices to show that, conditionally on $\mathcal{F}_n$, one obtains

$$X^{(n)} \in \mathcal{L} \quad \text{and} \quad X^{(n)} \text{ has parameter } (J_n, K) \text{ a.s.} \tag{8}$$

In fact, under (8), Theorem 9 implies that $\mu \sim \mu_n^*$ conditionally on $\mathcal{F}_n$, namely

$$P(\mu \in \cdot \mid \mathcal{F}_n) = P(\mu_n^* \in \cdot \mid \mathcal{F}_n) \qquad \text{a.s.}$$

In turn, condition (8) follows directly from the definition. Define in fact

$$P^{\mathcal{F}_n}\big(X^{(n)} \in B\big) = P\big(X^{(n)} \in B \mid \mathcal{F}_n\big) \qquad \text{for all } B \in \mathcal{B}^\infty \text{ a.s.}$$

Then,

$$P^{\mathcal{F}_n}\big(X_1^{(n)} \in \cdot\big) = P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu + \sum_{i=1}^n K(X_i)}{n+\theta} = \frac{J_n}{n+\theta} \quad \text{a.s.}$$

and

$$P^{\mathcal{F}_n}\Big(X_{m+1}^{(n)} \in \cdot \mid X_1^{(n)}, \ldots, X_m^{(n)}\Big) = P\big(X_{n+m+1} \in \cdot \mid \mathcal{F}_{n+m}\big)$$

$$= \frac{\theta\nu + \sum_{i=1}^{n+m} K(X_i)}{n+m+\theta} = \frac{J_n + \sum_{i=1}^m K\big(X_i^{(n)}\big)}{(n+\theta)+m} \quad \text{a.s. for all } m \geq 1.$$

This concludes the proof. □

4. Open problems and examples

This section is split into two parts. First, we discuss some hints for future research and then we give three further examples.

- **An enlargment of $\mathcal{L}$.** The class $\mathcal{L}$ could be made larger. In this case, however, some of the basic properties of $\mathcal{L}_0$ would be lost. As an example, suppose that

$$X_1 \sim \nu \quad \text{and} \quad P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = c_n\, \nu(\cdot) + (1-c_n)\, \frac{\sum_{i=1}^n K(X_i)(\cdot)}{n} \qquad \text{a.s.,}$$

where the kernel K satisfies condition (2) and $c_n \in [0,1]$ is a constant. To make X closer to $\mathcal{L}_0$, suppose also that $\lim_n c_n = 0$. Then, X is exchangeable and $X \in \mathcal{L}$ provided $c_n = \theta/(n+\theta)$. Furthermore, various properties of $\mathcal{L}_0$ are preserved, including $\mu \sim \sum_j V_j\, K(Z_j)$ where (V_j) and (Z_j) are independent sequences and (Z_j) is i.i.d. with $Z_1 \sim \nu$. Unlike Theorem 9, however, the probability distribution of (V_j) is unknown (to us). Similarly, we do not know whether some form of Theorem 13 is still valid.

- **A characterization of $\mathcal{L}$.** Denoting by $\mathcal{L}^*$ the class of kernel based Dirichlet sequences, it is tempting to conjecture that $\mathcal{L} = \mathcal{L}^*$. Since $\mathcal{L} \subset \mathcal{L}^*$, the question is whether there is an exchangeable sequence satisfying condition (1) but not condition (2). Lemma 6 and Theorem 7 may be useful to address this issue.
- **Self-similarity.** Suppose $X \in \mathcal{L}_0$ and take $A \in \mathcal{B}$ such that $0 < \nu(A) < 1$. Then, the distribution of the random probability measure $\mu(\cdot \mid A)$ is still of the Dirichlet type with ν and θ replaced by $\nu(\cdot \mid A)$ and $\theta\,\nu(A)$, respectively. In addition, $\mu(A)$, $\mu(\cdot \mid A)$ and $\mu(\cdot \mid A^c)$ are independent random elements; see [14, p. 61]. A question is whether this property of $\mathcal{L}_0$, called *self-similarity*, is still true for $\mathcal{L}$. Suppose $X \in \mathcal{L}$ and K is a r.c.d. for ν given the sub-σ-field $\mathcal{G} \subset \mathcal{B}$. If $A \in \mathcal{G}$, then $K(X_i)(A) = 1_A(X_i)$ a.s. for all i. Based on this fact, $\mu(\cdot \mid A)$ can be shown to have the same distribution as μ with ν and θ replaced by $\nu(\cdot \mid A)$ and $\theta\,\nu(A)$. Hence, $\mathcal{L}$ satisfies some form of self-similarity when $A \in \mathcal{G}$. However, we do not know whether $\mu(A)$, $\mu(\cdot \mid A)$ and $\mu(\cdot \mid A^c)$ are independent. Similarly, we do not know what happens if $A \notin \mathcal{G}$.
- **Topological support.** The topological support of a Borel probability λ on a separable metric space, denoted $\mathcal{S}(\lambda)$, is the smallest closed set A satisfying $\lambda(A) = 1$. Let $\mathcal{P}$ be equipped with the topology of weak convergence, i.e., the weakest topology on $\mathcal{P}$ which makes continuous the maps $p \mapsto \int f\,dp$ for all bounded continuous functions $f : S \to \mathbb{R}$. Moreover, let

$$\Pi(C) = P(\mu \in C), \qquad C \in \mathcal{C},$$

be the prior corresponding to μ. It is well known that

$$\mathcal{S}(\Pi) = \big\{p \in \mathcal{P} : \mathcal{S}(p) \subset \mathcal{S}(\nu)\big\}$$

whenever $X \in \mathcal{L}_0$; see [12] and [20]. As a consequence, $\mathcal{S}(\Pi) = \mathcal{P}$ if $\mathcal{S}(\nu) = S$. A (natural) question is whether, under some conditions on K, this basic property of $\mathcal{L}_0$ is preserved by $\mathcal{L}$. The next result provides a partial answer.

Proposition 14. *If $X \in \mathcal{L}$ and $\mathcal{S}(\Pi) = \mathcal{P}$, then*

$$\nu\big\{x \in S : K(x)(A) \le u\big\} < 1 \tag{9}$$

for all $u < 1$ and all non-empty open sets $A \subset S$.

Proof. First note that $\mathcal{S}(\Pi) = \mathcal{P}$ if and only if $\Pi(U) > 0$ for each non-empty open set $U \subset \mathcal{P}$. Having noted this fact, suppose $\nu\big\{x \in S : K(x)(A) \le u\big\} = 1$, for some $u < 1$ and some non-empty open set $A \subset S$, and define

$$U = \big\{p \in \mathcal{P} : p(A) > u\big\}.$$

Then, U is open and non-empty. Moreover, if V_j, Z_j and μ^* are as in Section 3, one obtains $K(Z_j)(A) \le u$ for all j a.s. and

$$\mu^*(A) = \sum_j V_j\, K(Z_j)(A) \le u \sum_j V_j = u \qquad \text{a.s.}$$

By Theorem 9, it follows that

$$\Pi(U) = P\big(\mu(A) > u\big) = P\big(\mu^*(A) > u\big) = 0.$$

Hence, $\mathcal{S}(\Pi)$ is a proper subset of $\mathcal{P}$. □

Possibly, some version of condition (9) suffices for $\mathcal{S}(\Pi) = \mathcal{P}$. However, condition (9) alone suggests that $\mathcal{S}(\Pi)$ is usually a proper subset of $\mathcal{P}$. In Example 3, for instance, condition (9) fails (just take $A = (0,\infty)$ and note that $K(x)(A) \le 1/2$ for all x). Finally, we mention here a property of $\mathcal{L}_0$ which is preserved by $\mathcal{L}$. If $X \in \mathcal{L}$ and $\mathcal{S}(\nu) = S$, the prior Π a.s. selects probability measures with full support, i.e.

$$\Pi\big\{p \in \mathcal{P} : \mathcal{S}(p) = S\big\} = 1.$$

We next turn to examples.

Example 15 (Example 1 continued). Let $\mathcal{H} \subset \mathcal{B}$ be a countable partition of S such that $\nu(H) > 0$ for all $H \in \mathcal{H}$. Then, $K(x) = \nu\big[\cdot \mid H(x)\big]$ is a r.c.d. for ν given $\sigma(\mathcal{H})$, where $H(x)$ is the only $H \in \mathcal{H}$ such that $x \in H$. Therefore, $X \in \mathcal{L}$ provided $X_1 \sim \nu$ and

$$P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu(\cdot) + \sum_{i=1}^n \nu\big[\cdot \mid H(X_i)\big]}{n+\theta} \qquad \text{a.s.}$$

In this example, for each $A \in \mathcal{B}$, one obtains

$$\mu(A) = \lim_n P\big(X_{n+1} \in A \mid \mathcal{F}_n\big) = \sum_{H \in \mathcal{H}} \mu(H)\,\nu(A \mid H) \quad \text{a.s.}$$

where $\mu(H) \overset{a.s.}{=} \lim_n (1/n) \sum_{i=1}^n 1_H(X_i)$. To grasp further information about μ, define

$$b(H) = \sum_j V_j\, 1_H(Z_j), \qquad H \in \mathcal{H},$$

where (V_j) and (Z_j) are independent, (Z_j) is i.i.d. with $Z_1 \sim \nu$, and (V_j) has the stick breaking distribution with parameter θ. Then, Theorem 9 yields

$$\mu \sim \mu^* = \sum_{H \in \mathcal{H}} b(H)\,\nu(\cdot \mid H).$$

Therefore,

$$\big(\mu(H) : H \in \mathcal{H}\big) \sim \big(\mu^*(H) : H \in \mathcal{H}\big) = \big(b(H) : H \in \mathcal{H}\big).$$

To evaluate the posterior distribution of μ, fix $n \ge 1$ and take two sequences $V^{(n)} = \big(V_j^{(n)} : j \ge 1\big)$ and $Z^{(n)} = \big(Z_j^{(n)} : j \ge 1\big)$ satisfying conditions (i)-(ii)-(iii). Recall that, by (iii), $Z^{(n)}$ is i.i.d. conditionally on $\mathcal{F}_n$ with

$$P(Z_1^{(n)} \in \cdot \mid \mathcal{F}_n) = P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) \qquad \text{a.s.}$$

Define

$$b_n(H) = \sum_j V_j^{(n)}\, 1_H\big(Z_j^{(n)}\big) \quad \text{and} \quad \mu_n^* = \sum_{H\in\mathcal{H}} b_n(H)\, \nu(\cdot \mid H).$$

Then, Theorem 13 implies $\mu \sim \mu_n^*$ conditionally on $\mathcal{F}_n$.

Example 16. Let $\|\cdot\|$ be the Euclidean norm on $S = \mathbb{R}^p$. For $t \geq 0$, let $\mathcal{U}_t \in \mathcal{P}$ be uniform on the spherical surface $\{x : \|x\| = t\}$ (with $\mathcal{U}_0 = \delta_0$) and

$$\nu(A) = \int_0^\infty \mathcal{U}_t(A)\, e^{-t}\, dt \qquad \text{for all } A \in \mathcal{B}.$$

Then, $K(x) = \mathcal{U}_{\|x\|}$ is a r.c.d. for ν given $\sigma(\|\cdot\|)$. Hence, $X \in \mathcal{L}$ whenever $X_1 \sim \nu$ and

$$P\big(X_{n+1} \in \cdot \mid \mathcal{F}_n\big) = \frac{\theta\nu(\cdot) + \sum_{i=1}^n \mathcal{U}_{\|X_i\|}(\cdot)}{n+\theta} \qquad \text{a.s.}$$

Theorem 11 applies to this example. To see this, first note that

$$\int \|x\|^2\, \nu(dx) = \int_0^\infty \int \|x\|^2\, \mathcal{U}_t(dx)\, e^{-t} dt = \int_0^\infty t^2\, e^{-t} dt < \infty.$$

Moreover, since $\mathcal{U}_t$ is invariant under rotations,

$$\int y_i\, \mathcal{U}_t(dy) = \int y_i\, y_j\, \mathcal{U}_t(dy) = 0 \quad \text{and} \quad \int y_i^2\, \mathcal{U}_t(dy) = t^2/p \tag{10}$$

for all t, all i and all $j \neq i$. (Recall that y_i denotes the i-th coordinate of a point $y \in \mathbb{R}^p$). Because of (10),

$$\int y_i\, K(x)(dy) = \int y_i\, \mathcal{U}_{\|x\|}(dy) = 0 \qquad \text{for all } x \in \mathbb{R}^p \text{ and } i = 1, \ldots, p.$$

Therefore, Theorem 11 yields

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \xrightarrow{stably} \mathcal{N}_p(0, \Sigma)$$

where Σ is the random covariance matrix with entries

$$\sigma_{ij} = \int y_i\, y_j\, \mu(dy) = \lim_n \frac{1}{n} \sum_{r=1}^n \int y_i\, y_j\, \mathcal{U}_{\|X_r\|}(dy) \qquad \text{a.s.}$$

It is even possible be more precise about Σ. In fact, using (10) again, one obtains $\sigma_{ij} = 0$ for $i \neq j$ and

$$\sigma_{ii} = \lim_n \frac{1}{n} \sum_{r=1}^n \int y_i^2\, \mathcal{U}_{\|X_r\|}(dy) = \frac{1}{p} \lim_n \frac{1}{n} \sum_{r=1}^n \|X_r\|^2 = \frac{1}{p} \int \|x\|^2\, \mu(dx) \quad \text{a.s.}$$

Hence, if I denotes the $p \times p$ identity matrix,

$$\Sigma = \sigma_{11}\, I \quad \text{where} \quad \sigma_{11} = (1/p) \int \|x\|^2\, \mu(dx).$$

Two last remarks are in order. First, in the notation of Theorem 9,

$$\int \|x\|^2\,\mu(dx) \sim \int \|x\|^2\,\mu^*(dx) = \sum_j V_j\,\|Z_j\|^2.$$

Second, exploiting stable convergence and $\sigma_{11} > 0$ a.s., one also obtains

$$\sqrt{p}\,\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n \|X_i\|^2}} = \sqrt{p}\,\frac{n^{-1/2}\sum_{i=1}^n X_i}{\sqrt{n^{-1}\sum_{i=1}^n \|X_i\|^2}} \overset{stably}{\longrightarrow} \mathcal{N}_p(0, I).$$

Example 17. Let F be a countable class of measurable maps $f : S \to S$ and

$$\mathcal{I} = \{\lambda \in \mathcal{P} : \lambda = \lambda \circ f^{-1} \text{ for each } f \in F\}$$

the set of F-invariant probability measures. Let

$$\mathcal{G} = \{A \in \mathcal{B} : f^{-1}(A) = A \text{ for all } f \in F\}$$

be the sub-σ-field of F-invariant measurable sets. In this example, we assume that $\nu \in \mathcal{I}$ and conditions (1)-(2) hold with $\mathcal{G}$ as above.

Under these conditions, it is not hard to see that $K(x) \in \mathcal{I}$ for ν-almost all $x \in S$; see e.g. [19]. Hence, $P(X_{n+1} \in \cdot \mid \mathcal{F}_n) \in \mathcal{I}$ a.s. which in turn implies

$$\mu(f^{-1}A) \overset{a.s.}{=} \lim_n P(f(X_{n+1}) \in A \mid \mathcal{F}_n) \overset{a.s.}{=} \lim_n P(X_{n+1} \in A \mid \mathcal{F}_n) \overset{a.s.}{=} \mu(A)$$

for fixed $A \in \mathcal{B}$ and $f \in F$. Since F is countable and $\mathcal{B}$ countably generated, one finally obtains

$$P(\mu \in \mathcal{I}) = 1.$$

This fact is meaningful from the Bayesian point of view. It means that the prior corresponding to μ (namely, $\Pi(C) = P(\mu \in C)$ for all $C \in \mathcal{C}$) selects F-invariant laws a.s. Such priors are actually useful in some practical problems; see e.g. [9] and [17].

Example 3 is a special case of the previous choice of $\mathcal{G}$. Another example, borrowed from [3, Ex. 12], is $S = \mathbb{R}^d$ and F the class of all permutations of $\mathbb{R}^d$. In this case, $\mathcal{I}$ is the set of exchangeable probabilities on the Borel sets of $\mathbb{R}^d$. Moreover, if ν is exchangeable, K can be written as

$$K(x) = \frac{\sum_{\pi \in F} \delta_{\pi(x)}}{d!} \qquad \text{for all } x \in \mathbb{R}^d.$$

A last remark is in order.

Claim: If $A_1, \ldots, A_k$ is a partition of S such that $A_i \in \mathcal{G}$ for all i, then the k-dimensional vector $(\mu(A_1), \ldots, \mu(A_k))$ has Dirichlet distribution with parameters $\theta\,\nu(A_1), \ldots, \theta\,\nu(A_k)$.

To prove the Claim, because of Theorem 9, it suffices to show that $(\mu^*(A_1), \ldots, \mu^*(A_k))$ has the desired distribution. In addition, $K(x)(A_i) = 1_{A_i}(x) = \delta_x(A_i)$, for ν-almost all $x \in S$, since $A_i \in \mathcal{G}$ and K is a r.c.d. for ν given $\mathcal{G}$. Therefore,

$$\mu^*(A_i) = \sum_j V_j\,K(Z_j)(A_i) \overset{a.s.}{=} \sum_j V_j\,\delta_{Z_j}(A_i) \qquad \text{for all } i,$$

and this implies that $\big(\mu^*(A_1),\dots,\mu^*(A_k)\big)$ has Dirichlet distribution with parameters $\theta\nu(A_1),\dots,\theta\nu(A_k)$.

In view of the Claim, μ is a Dirichlet invariant process in the sense of Definition 2 of [9]. Thus, arguing as above, a large class of such processes can be easily obtained. Note also that, unlike [9], F is not necessarily a group.

Acknowledgements

This paper has been improved by the remarks and suggestions of the AE and four anonymous referees.

References

[1] Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969

[2] Bally, V. and Caramellino, L. (2016). Asymptotic development for the CLT in total variation distance. *Bernoulli* **22** 2442–2485. MR3498034 https://doi.org/10.3150/15-BEJ734

[3] Berti, P., Dreassi, E., Pratelli, L. and Rigo, P. (2021). A class of models for Bayesian predictive inference. *Bernoulli* **27** 702–726. MR4177386 https://doi.org/10.3150/20-BEJ1255

[4] Berti, P., Pratelli, L. and Rigo, P. (2004). Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **32** 2029–2052. MR2073184 https://doi.org/10.1214/009117904000000676

[5] Berti, P., Pratelli, L. and Rigo, P. (2012). Limit theorems for empirical processes based on dependent data. *Electron. J. Probab.* **17** no. 9, 18. MR2878788 https://doi.org/10.1214/EJP.v17-1765

[6] Berti, P., Pratelli, L. and Rigo, P. (2013). Exchangeable sequences driven by an absolutely continuous random measure. *Ann. Probab.* **41** 2090–2102. MR3098068 https://doi.org/10.1214/12-AOP786

[7] Berti, P. and Rigo, P. (2007). 0-1 laws for regular conditional distributions. *Ann. Probab.* **35** 649–662. MR2308591 https://doi.org/10.1214/009117906000000845

[8] Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614

[9] Dalal, S.R. (1979). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Process. Appl.* **9** 99–107. MR0544719 https://doi.org/10.1016/0304-4149(79)90043-7

[10] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**. MR0325177 https://doi.org/10.1016/0040-5809(72)90035-4

[11] Favaro, S., Lijoi, A. and Prünster, I. (2012). On the stick-breaking representation of normalized inverse Gaussian priors. *Biometrika* **99** 663–674. MR2966776 https://doi.org/10.1093/biomet/ass023

[12] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

[13] Fortini, S., Ladelli, L. and Regazzini, E. (2000). Exchangeability, predictive distributions and parametric models. *Sankhyā Ser. A* **62** 86–109. MR1769738

[14] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge: Cambridge Univ. Press. MR3587782 https://doi.org/10.1017/9781139029834

[15] Hahn, P.R., Martin, R. and Walker, S.G. (2018). On recursive Bayesian predictive distributions. *J. Amer. Statist. Assoc.* **113** 1085–1093. MR3862341 https://doi.org/10.1080/01621459.2017.1304219

[16] Hansen, B. and Pitman, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statist. Probab. Lett.* **46** 251–256. MR1745692 https://doi.org/10.1016/S0167-7152(99)00109-1

[17] Hosseini, R. and Zarepour, M. (2021). Bayesian bootstrapping for symmetric distributions. *Statistics* **55** 711–732. MR4313446 https://doi.org/10.1080/02331888.2021.1961141

[18] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. MR0733519 https://doi.org/10.1214/aos/1176346412

[19] Maitra, A. (1977). Integral representations of invariant measures. *Trans. Amer. Math. Soc.* **229** 209–225. MR0442197 https://doi.org/10.2307/1998506
[20] Majumdar, S. (1992). On topological support of Dirichlet prior. *Statist. Probab. Lett.* **15** 385–388. MR1193898 https://doi.org/10.1016/0167-7152(92)90171-Z
[21] Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. Hayward, CA: IMS. MR1481784 https://doi.org/10.1214/lnms/1215453576
[22] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129 https://doi.org/10.1214/aop/1024404422
[23] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433
[24] Hjort, N.L., Holmes, C., Muller, P. and Walker, S.G. (2010). *Bayesian Nonparametric*. Cambridge: Cambridge Univ. Press.

Received September 2021 and revised January 2022