The Canadian Journal of Statistics
Vol. 30, No. 2, 2002, Pages 269–283
La revue canadienne de statistique

Exact and approximate sum representations for the Dirichlet process

Hemant ISHWARAN and Mahmoud ZAREPOUR

Key words and phrases: Almost sure truncation; finite dimensional Dirichlet prior; Lévy measure; Poisson process; random probability measure; stick-breaking prior; weak convergence.

MSC 2000: Primary 60G57, 62G99.

Abstract: The Dirichlet process can be regarded as a random probability measure for which the authors examine various sum representations. They consider in particular the gamma process construction of Ferguson (1973) and the "stick-breaking" construction of Sethuraman (1994). They propose a Dirichlet finite sum representation that strongly approximates the Dirichlet process. They assess the accuracy of this approximation and characterize the posterior that this new prior leads to in the context of Bayesian nonparametric hierarchical models.

Représentations exactes et approximatives du processus de Dirichlet par des sommes

Résumé : Le processus de Dirichlet constitue une mesure de probabilité aléatoire dont les auteurs examinent différentes représentations à l'aide de sommes. Ils s'intéressent en particulier à la construction de Ferguson (1973) fondée sur la loi gamma et à la construction dite à "bâtons rompus" de Sethuraman (1994). Ils proposent une approximation forte du processus de Dirichlet par somme finie de type Dirichlet. Ils évaluent la qualité de cette approximation qui conduit à une loi a priori dont ils caractérisent la loi a posteriori dans le cadre des modèles bayésiens hiérarchiques non paramétriques.

1. INTRODUCTION

The Dirichlet process was discovered by Freedman (1963) through the notion of a tail-free measure (see also Fabius 1964), and its properties and theory were developed by Ferguson (1973, 1974) and Blackwell & MacQueen (1973). In its most general form, the Dirichlet process can be defined over an arbitrary measurable space (Ferguson 1973). However, for measure-theoretic reasons, our discussion of the process will be confined to a measurable Polish space $(\mathcal{Y}, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra for the space $\mathcal{Y}$. In his seminal paper, Ferguson (1973) provided two key definitions for the Dirichlet process. His first definition described it as a stochastic process, $\mathcal{P}$, indexed by elements B of $\mathcal{B}$, such that for each measurable partition $B_1, \ldots, B_d$ of $\mathcal{Y}$,

$$(\mathcal{P}(B_1), \ldots, \mathcal{P}(B_d)) \sim \text{Dirichlet}(\mu(B_1), \ldots, \mu(B_d)),$$

where μ is a finite non-null measure on $(\mathcal{Y}, \mathcal{B})$. Ferguson (1973) referred to this process as the Dirichlet process with parameter μ, which we will denote as $\text{DP}(\mu)$.

This characterization as a stochastic process, along with its conjugacy property to iid sampling, also established in Ferguson (1973), were enough to fuel much of the early work on the Dirichlet process. See Antoniak (1974), Berry & Christensen (1979), Lo (1984), Doss (1985), and Kuo (1986) for examples. However, for reasons to be discussed, very little of this work utilized the alternate definition of Ferguson (1973) which, by exploiting a connection to the gamma process, described the Dirichlet process as a random probability measure with an infinite sum construction. Some early exceptions were Korwar & Hollander (1973), Hannum, Hollander & Langberg (1981) and Yamato (1984). This constructive definition, and other sum constructions, especially those that can be practically used in computational algorithms in Bayesian nonparametric problems, will be the focus of this paper.

1.1. The Dirichlet process as a random measure.

The sum construction of Ferguson (1973) was based on earlier work by Ferguson & Klass (1972), who provided a representation for the gamma process based on arrival times from a homogeneous Poisson process. Let E_k be independent and identically distributed (iid) $\exp(1)$ random variables and let $\Gamma_k = E_1 + \cdots + E_k$. Let Z_k be iid elements, independent of Γ_k, with a probability distribution H over $(\mathcal{Y}, \mathcal{B})$. Then, Ferguson (1973) showed that the Dirichlet process with parameter $\mu = \alpha H$, for some real value $\alpha > 0$, could be described as the random probability measure

$$\mathcal{P}(\cdot) = \sum_{k=1}^{\infty} N^{-1}(\Gamma_k)\, \delta_{Z_k}(\cdot) \Big/ \sum_{\ell=1}^{\infty} N^{-1}(\Gamma_\ell), \tag{1}$$

where δ_Z denotes a discrete measure concentrated at Z, and

$$N(x) = \alpha \int_x^{\infty} \frac{e^{-u}}{u}\, du, \qquad \text{for } x > 0, \tag{2}$$

is the Lévy measure for a Gamma(α) random variable.

The earlier lack of interest in the Ferguson (1973) sum construction reflected perhaps the difficulty in working with the random weights

$$p_k = N^{-1}(\Gamma_k) \Big/ \sum_{\ell=1}^{\infty} N^{-1}(\Gamma_\ell)$$

used in (1), since no closed form solution exists for the inverse of the Lévy measure (2) and since each weight p_k requires computing an infinite sum. The theory for this sequence of weights was studied by Kingman (1975), who called the distribution of the infinite vector $(p_1, p_2, \ldots)$ the *Poisson–Dirichlet distribution.*

Recognizing the complexities in working with (1) (and with the formulation of the Dirichlet process as a stochastic process), Sethuraman (1994) considered an alternate "stick-breaking" sum representation using random weights constructed via a stick-breaking approach involving independent beta random variables. Using this construction, Sethuraman (1994) was able to prove directly several of the key properties of the Dirichlet process established by Ferguson (1973) and Blackwell & MacQueen (1973). See also McCloskey (1965), Patil & Taillie (1977), Sethuraman & Tiwari (1982), Hoppe (1987), Donnelly & Joyce (1989), Perman, Pitman & Yor (1992) and Pitman (1996) who have discussed this stick-breaking construction.

Although Sethuraman's stick-breaking construction was motivated more from a theoretical interest in the Dirichlet process (see his Examples section, however, where Monte Carlo schemes are suggested), the stick-breaking construction, and its approximations, have become increasingly utilized in Bayesian computational procedures. For example, Doss (1994) used the representation explicitly in a Gibbs sampling procedure for analyzing censored data, while truncation approximations have been used by Muliere & Tardella (1998) for approximating Dirichlet process functionals, and by Ishwaran & James (2001, 2002) and Gelfand & Kottas (2002) for posterior computations in Dirichlet process mixture models.

In addition to these truncation approximations, there has also been a surge of interest in other approximating sum representations for the Dirichlet process as a method for directly sampling from the random posterior measure in Bayesian nonparametric models. See for example, Liu (1996), Muliere & Secchi (1996) and Ishwaran & Zarepour (2000). More discussions of these techniques as well as those mentioned above will appear in Section 3.

1.2. Outline and contributions of the paper.

Given such wide interest, the focus of this paper will be to present a new theory for exact and approximate sum representations of the Dirichlet process. We look at representations expressible

as random discrete measures

$$\mathcal{P}(\cdot) = \sum_{k=1}^{N} p_k \, \delta_{Z_k}(\cdot), \qquad 1 \leq N \leq \infty, \tag{3}$$

where p_k are random variables chosen to be independent of Z_k and constructed so that $0 \leq p_k \leq 1$ and $\sum_{k=1}^{\infty} p_k = 1$ with probability one. To a large extent, we will motivate such representations in the context of Bayesian nonparametric problems.

We begin in Section 2 by revisiting the Ferguson (1973) Dirichlet process construction (1). There, we prove that there is a class of alternate-sum representations expressible in terms of a homogeneous Poisson process, thus showing directly that the Dirichlet process can be represented in many ways as a discrete random measure. Moreover, the same idea can be generalized to infinitely divisible distributions and their random probability measures (see Theorem 1). In Section 3, we review the Sethuraman (1994) stick-breaking construction and describe its connection to the Poisson process. We also look at techniques for truncating this construction and discuss how they can be exploited computationally in fitting Bayesian nonparametric models.

In Section 4, we present a finite-dimensional random measure based on symmetric Dirichlet random weights. One of its nice features is its interpretation as a mixture of Dirichlet processes. We exploit this in showing that its limit is the Dirichlet process (see Theorem 2 for a precise statement). Not surprisingly, given its simple construction, this measure has been the subject of some interest in the literature, seeming to have been discovered independently by several authors in different contexts (see Section 4 for more discussion). However, little seems to be known about its theoretical properties in Bayesian nonparametric problems. We present a detailed analysis of such properties, first by discussing the adequacy in which the measure approximates the Dirichlet process, giving bounds on the $\mathcal{L}_1$ distance between marginal densities in nonparametric hierarchical models (Theorem 4) and by giving bounds on the number of distinct sampled values (Theorem 3). We then present an informative characterization of its posterior in Bayesian nonparametric hierarchical models (see Theorem 5). We show how such a characterization can be readily used with known computational algorithms, such as Pólya urn Gibbs samplers, to allow for draws directly from the posterior random measure. As will be discussed, this presents Bayesians with another method for estimating laws for arbitrary functionals of the posterior, in direct analogue to the methods discussed in Section 3.

2. FERGUSON'S SUM CONSTRUCTION

By using point process methods, we will show how a class of infinite sum representations can be constructed for random probability measures based on infinitely divisible random variables. In particular, this method directly establishes the existence of many nontrivial alternate representations of the Dirichlet process as a random discrete measure (3).

Let J be a positive infinitely divisible non-Gaussian random variable whose characteristic function can be expressed as

$$\phi(\theta) = \exp\left\{ -\int_0^{\infty} \left(e^{i\theta u} - 1\right) \, dN(u) \right\}, \quad -\infty < \theta < \infty,$$

where N (the Lévy measure) is a Borel measure defined on $(0, \infty)$ by $N(x) = \int_x^{\infty} dN(u)$ so that

$$\int_{\varepsilon}^{\infty} N^{-1}(u) \, du < \infty, \qquad \text{for each } \varepsilon > 0, \tag{4}$$

where $N^{-1}(u) = \sup\{x : N(x) \leq u\}$. If N is positive and continuous, then by Ferguson & Klass (1972) it follows that $J = \sum_{k=1}^{\infty} J_k$ almost surely, where $J_k = N^{-1}(\Gamma_k)$. Thus,

$$\sum_{k=1}^{\infty} J_k \, \delta_{Z_k}(\cdot) \Big/ \sum_{\ell=1}^{\infty} J_\ell \tag{5}$$

is a random probability measure with random weights based on an infinitely divisible distribution.

As shown by Ferguson (1973), the Dirichlet process is an example of a random measure expressible as (5). In this case, $J = \sum_{k=1}^{\infty} J_k$ is a Gamma(α) random variable and N is the Lévy measure defined by (2). Note that the integrability condition (4) is satisfied from

$$\int_0^{\infty} N^{-1}(u)\, du = \int_0^{\infty} u\, dN(u) = \alpha \int_0^{\infty} e^{-u}\, du = \alpha.$$

Our discussion will also include other infinitely divisible distributions, such as positive stable laws with index $0 < \alpha < 1$, in which case

$$N(x) = x^{-\alpha} = \int_x^{\infty} \alpha u^{-(1+\alpha)}\, du.$$

Notice that (4) is easily verified since $N^{-1}(u) = u^{-1/\alpha}$. The gamma distribution and stable laws are examples of infinitely divisible distributions which satisfy the following theorem (see the Appendix for a proof).

THEOREM 1. *Suppose that N is a positive, continuous Lévy measure satisfying* (4). *If W_k are iid positive random variables independent of Γ_k such that $E(W_1^{-1}) = 1$, then*

$$\sum_{k=1}^{\infty} \frac{N^{-1}(\Gamma_k)}{\sum_{\ell=1}^{\infty} N^{-1}(\Gamma_\ell)}\, \delta_{Z_k}(\,\cdot\,) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{\infty} \frac{N^{-1}(W_k\Gamma_k)}{\sum_{\ell=1}^{\infty} N^{-1}(W_\ell\Gamma_\ell)}\, \delta_{Z_k}(\,\cdot\,). \tag{6}$$

Observe that the random weights on the left-hand side of (6) form an ordered decreasing sequence since N is non-increasing. Thus, for the case of a gamma Lévy measure (2), Theorem 1 shows directly that there are many alternate representations for the Dirichlet process using nonordered random weights. Unfortunately, both the Ferguson sum construction and the alternate-sum construction described on the right of (6) are difficult to work with directly because of the complexity of the associated Lévy measure. However, there exists simpler representations that one can work with, also constructed from a Poisson process. One such example is given by Bondesson (1982). Let E_k^* be iid exp(1) random variables, independent of both Γ_k and Z_k. Bondesson (1982) showed that the DP(αH) process has the sum construction

$$\mathcal{P}(\,\cdot\,) = \sum_{k=1}^{\infty} e^{-\Gamma_k} E_k^*\, \delta_{Z_k}(\,\cdot\,) \Big/ \sum_{\ell=1}^{\infty} e^{-\Gamma_\ell} E_\ell^*,$$

which is somewhat easier to work with since it avoids the use of a Lévy measure. In the next section, we will see that there are even simpler representations available, also related to the Poisson process, but which are constructed using a stick-breaking method.

3. STICK-BREAKING REPRESENTATION

The unpublished thesis by McCloskey (1965) appears to be the first work that drew comparisons between the Poisson–Dirichlet random weights and random weights defined by a beta random variable stick-breaking procedure. It wasn't until Sethuraman (1994), however, that this stick-breaking representation was used as a means for directly proving key properties of the Dirichlet process. See also Sethuraman & Tiwari (1982) as well as Diaconis & Kemperman (1996) for examples of theoretical results derived using this representation. Also see Hoppe (1987), Donnelly & Joyce (1989), Perman, Pitman & Yor (1992), Pitman (1996) and Pitman & Yor (1997) for further discussion on McCloskey's result.

In the Sethuraman (1994) stick-breaking construction, the random weights p_k are defined by

$$p_1 = V_1 \quad \text{and} \quad p_k = (1 - V_1) \cdots (1 - V_{k-1})\, V_k, \qquad \text{for } k \geq 2,$$

where V_k are iid Beta$(1, \alpha)$ random variables independent of Z_k. The distribution for this sequence of random weights $p_1, p_2, \ldots$ is often referred to as the GEM distribution with parameter α, named so by Ewens (1990) after Griffiths, Engen and McCloskey (for further details see also Johnson, Kotz & Balakrishnan 1997, ch. 41). Using such weights, Sethuraman (1994) showed that

$$\mathcal{P}(\cdot) = V_1 \delta_{Z_1}(\cdot) + \sum_{k=2}^{\infty} \{(1 - V_1) \cdots (1 - V_{k-1}) V_k\} \, \delta_{Z_k}(\cdot) \tag{7}$$

is the Dirichlet process DP(αH).

3.1. Relationship to the Poisson process.

The stick-breaking construction is related to the Poisson process in many ways. For example, the GEM weights used in the Sethuraman (1994) construction can be obtained by a size-biased random permutation of the Poisson–Dirichlet random weights (Patil & Taillie 1977; Perman, Pitman & Yor 1992; Pitman & Yor 1997). The two sets of weights are also related by

$$(p_{(1)}, p_{(2)}, \ldots,) \stackrel{\text{a.s}}{=} \left(\frac{N^{-1}(\Gamma_1)}{\sum_{\ell=1}^{\infty} N^{-1}(\Gamma_\ell)}, \frac{N^{-1}(\Gamma_2)}{\sum_{\ell=1}^{\infty} N^{-1}(\Gamma_\ell)}, \cdots \right),$$

where $p_{(1)} \geq p_{(2)} \geq \cdots$ are the ordered GEM weights and N is the gamma Lévy measure (2).

A more direct connection to the Poisson process is the following representation

$$\mathcal{P}(\cdot) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{\infty} (e^{-\Gamma_{k-1}/\alpha} - e^{-\Gamma_k/\alpha}) \, \delta_{Z_k}(\cdot), \tag{8}$$

where $\Gamma_0 = 0$. An easy way to see this is to observe that

$$p_k = e^{-\Gamma_{k-1}/\alpha} - e^{-\Gamma_k/\alpha} = e^{-E_1/\alpha} \cdots e^{-E_{k-1}/\alpha} (1 - e^{-E_k/\alpha}) \stackrel{\mathcal{D}}{=} (1 - V_1) \cdots (1 - V_{k-1}) V_k,$$

because $e^{-E_1/\alpha} \stackrel{\mathcal{D}}{=} \text{Beta}(\alpha, 1) \stackrel{\mathcal{D}}{=} 1 - \text{Beta}(1, \alpha)$.

3.2. Almost sure truncations.

The representation (8) is a useful formulation for deriving explicit bounds on truncations of the Sethuraman sum construction. In Ishwaran & James (2001, 2002), the construction (8) was used to determine the accuracy of almost sure truncations of the form

$$\begin{aligned} \mathcal{P}_N(\cdot) &= V_1 \, \delta_{Z_1}(\cdot) + \sum_{k=2}^{N} \{(1 - V_1) \cdots (1 - V_{k-1}) V_k\} \, \delta_{Z_k}(\cdot) \\ &\stackrel{\mathcal{D}}{=} (1 - e^{-\Gamma_1/\alpha}) \, \delta_{Z_1}(\cdot) + \sum_{k=2}^{N} (e^{-\Gamma_{k-1}/\alpha} - e^{-\Gamma_k/\alpha}) \, \delta_{Z_k}(\cdot), \end{aligned} \tag{9}$$

where $V_N = 1$ ($E_N = \infty$) so that $\sum_{k=1}^{N} p_k = 1$ (hereafter we use N to denote a finite positive integer). It follows automatically that $\mathcal{P}_N(g) \stackrel{\text{a.s}}{\to} \mathcal{P}(g)$ for each bounded and continuous real-valued function g. Although other methods could potentially be used to truncate the Sethuraman representation, the method of setting the Nth beta variable to one ensures that the joint distribution for the random weights has a generalized Dirichlet distribution (Connor & Mosimann 1969). This key property can be exploited to describe an efficient Gibbs sampler for Bayesian nonparametric problems in which $\mathcal{P}_N$ is used as an approximating prior to the Dirichlet process (Ishwaran & Zarepour 2000). Muliere & Tardella (1998) have also used the same truncation (9) for sampling Dirichlet process functionals. They refer to (9) as the "ε-Dirichlet random probability" and show that the truncation can be made arbitrarily accurate to within any $\varepsilon > 0$ in the total variation distance. Below in Section 3.4, we will show that the representation (8) can be used to get more explicit bounds for the prior as well as bounds for other Bayesian quantities.

3.3. Truncations applied to the posterior.

In another interesting use of an almost sure truncation, Gelfand & Kottas (2002) use (9) to approximate the posterior in Bayesian nonparametric (and semiparametric) models subject to the Dirichlet process prior. A generalization of a similar method appears in the discussion of Theorem 3 of Ishwaran & James (2001a). Note that such approaches use truncations to approximate the posterior based on a Dirichlet process prior, which is different from methods such as in Muliere & Tardella (1998) and Ishwaran & Zarepour (2000), which use (9) as the prior. It is worth elaborating more on this. In the first approach, a typical set-up involves data $\mathbf{X} = (X_1, \ldots, X_n)$, observed from the nonparametric hierarchical model (the approach readily extends to semiparametric models as well)

$$X_i \mid Y_i \overset{\text{ind}}{\sim} f(X_i \mid Y_i), \quad Y_i \mid P \overset{\text{iid}}{\sim} P, \quad P \sim \mathcal{P}, \quad i = 1, \ldots, n, \tag{10}$$

where $\mathcal{P}$ is a DP(αH) measure and $f(x \mid y)$ denotes the density for $x \in \mathcal{X}$ given $y \in \mathcal{Y}$, jointly measurable in x and y, where $f(x \mid y)$ is taken with respect to a σ-finite measure λ.

The posterior from (10) (see Antoniak 1974; Lo 1984) is characterized by

$$\mathcal{P}(\cdot \mid \mathbf{X}) = \int_{\mathcal{Y}^n} \mathcal{P}(\cdot \mid \mathbf{Y})\, d\nu(\mathbf{Y} \mid \mathbf{X}), \tag{11}$$

where $\mathcal{P}(\cdot \mid \mathbf{Y})$ is a Dirichlet process with finite measure $\mu_n = \alpha H + \sum_{i=1}^n \delta_{Y_i}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)$, and

$$d\nu(\mathbf{Y} \mid \mathbf{X}) = \frac{\prod_{i=1}^n f(X_i \mid Y_i) \prod_{i=1}^n \left\{ \alpha H(dY_i) + \sum_{j=1}^{i-1} \delta_{Y_j}(dY_i) \right\}}{\int_{\mathcal{Y}^n} \prod_{i=1}^n f(X_i \mid Y_i) \prod_{i=1}^n \left\{ \alpha H(dY_i) + \sum_{j=1}^{i-1} \delta_{Y_j}(dY_i) \right\}}.$$

Gelfand & Kottas (2002) approximate laws for functionals of the posterior $\mathcal{P}(\cdot \mid \mathbf{X})$ by first drawing $\mathbf{Y}$ from $\nu(\cdot \mid \mathbf{X})$ (they use a Pólya urn Gibbs sampler, although other Monte Carlo methods will also work) and then they compute the corresponding functional of an almost sure truncation $\mathcal{P}_N(\cdot \mid \mathbf{Y})$ to $\mathcal{P}(\cdot \mid \mathbf{Y})$. They note that because $\mathcal{P}(\cdot \mid \mathbf{Y})$ is a Dirichlet process with finite measure $\mu_n = (\alpha + n)H_n$, where $H_n = \mu_n/(\alpha + n)$ is a probability distribution, the corresponding almost sure approximation to $\mathcal{P}(\cdot \mid \mathbf{Y})$ is

$$\mathcal{P}_N(\cdot \mid \mathbf{Y}) = V_1^* \delta_{Z_1^*}(\cdot) + \sum_{k=2}^{N} \left\{ (1 - V_1^*) \cdots (1 - V_{k-1}^*)\, V_k^* \right\} \delta_{Z_k^*}(\cdot), \tag{12}$$

where V_k^* are iid Beta$(1, \alpha + n)$ random variables ($V_N^* = 1$) independent of Z_k^* which are iid with law H_n.

A generalization of this idea is discussed in Ishwaran & James (2001a, Theorem 3) which applies to the two-parameter Poisson–Dirichlet process, a generalization of the Dirichlet process, although the approximation used there for the Dirichlet process differs from (12). Ishwaran & James (2001a) first note that by Pitman (1996)

$$\mathcal{P}(\cdot \mid \mathbf{Y}) \overset{\mathcal{D}}{=} \sum_{j=1}^{m} p_j^* \delta_{Y_j^*}(\cdot) + p_{m+1}^* \mathcal{P}(\cdot),$$

where $Y_1^*, \ldots, Y_m^*$ are the unique set of Y_i values occurring each with frequencies n_j^*, and

$$(p_1^*, \ldots, p_m^*, p_{m+1}^*) \sim \text{Dirichlet}(n_1^*, \ldots, n_m^*, \alpha)$$

is independent of $\mathcal{P}$ which is a DP(αH) process. Thus, to approximate arbitrary functionals, they suggest drawing $\mathbf{Y}$ from $\nu(\cdot \mid \mathbf{X})$ and then computing the corresponding functional from

the approximation

$$\mathcal{P}_N(\cdot \mid \mathbf{Y}) = \sum_{j=1}^{m} p_j^* \, \delta_{Y_j^*}(\cdot) + p_{m+1}^* \mathcal{P}_N(\cdot),$$

where $\mathcal{P}_N$ is an independent almost sure approximation (9) to $\mathcal{P}$. Notice that the selection for N is less of an issue here, as one draws from $\mathcal{P}_N$ only with probability $\alpha/(\alpha+n)$, which becomes small for reasonable sample sizes n.

3.4. Truncations applied to the prior.

Alternately, one can apply the almost sure truncation as a prior in Bayesian nonparametric hierarchical models. Now the set-up involves data $\mathbf{X} = (X_1, \ldots, X_n)$, observed from the nonparametric hierarchical model

$$X_i \mid Y_i \overset{\text{ind}}{\sim} f(X_i \mid Y_i), \quad Y_i \mid P \overset{\text{iid}}{\sim} P, \quad P \sim \mathcal{P}_N, \quad i = 1, \ldots, n, \tag{13}$$

where $\mathcal{P}_N$ is an almost sure truncation (9) to $\mathcal{P}$.

A nice feature of using the approximation at the modeling stage is that it leads to a simple Gibbs sampling algorithm, dubbed the "blocked-Gibbs sampler" by Ishwaran & James (2001a), which allows one to directly draw values from the posterior of $\mathcal{P}_N$, thus by-passing the need to use approximations as discussed above. Selecting the truncation value N in this approach is also straightforward. One method is to choose N so that the $\mathcal{L}_1$ distance of the marginal densities under (10) and (13) are close. Let $m_N(\mathbf{X})$ be the marginal density of (13), i.e.,

$$m_N(\mathbf{X}) = \int \left\{ \prod_{i=1}^{n} \int_{\mathcal{Y}} f(X_i \mid Y_i) \, dP(Y_i) \right\} d\mathcal{P}_N(P). \tag{14}$$

If m_∞ is the marginal density for (10), then as Ishwaran & James (2001, 2002) showed,

$$\int |m_N(\mathbf{X}) - m_\infty(\mathbf{X})| \, d\lambda^n(\mathbf{X}) \approx 4n \, e^{-(N-1)/\alpha}, \tag{15}$$

by using the fact that

$$\sum_{k=1}^{N} p_k = \sum_{k=1}^{N} \left(e^{-\Gamma_{k-1}/\alpha} - e^{-\Gamma_k/\alpha} \right) = 1 - \exp(-\Gamma_N/\alpha) \approx 1 - e^{-N/\alpha}.$$

This bound provides a convenient mechanism for selecting N.

4. FINITE-DIMENSIONAL DIRICHLET PRIORS

By using Dirichlet random weights, we can construct a finite-sum random probability measure that is both computationally and theoretically easy to work with as a prior in Bayesian nonparametric problems and which has the added appeal that it approximates the Dirichlet process. Let $\mathbf{p} = (p_1, \ldots, p_N)$ have the Dirichlet$(\alpha/N, \ldots, \alpha/N)$ distribution independent of Z_k, which are iid H as before. Equivalently, let $p_k = G_k/G$, where $G = G_1 + \cdots + G_N$ and G_k are iid Gamma(α/N) random variables independent of Z_k. Define

$$\mathcal{P}_N(\cdot) = \sum_{k=1}^{N} p_k \, \delta_{Z_k}(\cdot) = \sum_{k=1}^{N} \frac{G_k}{G} \, \delta_{Z_k}(\cdot). \tag{16}$$

Then, $\mathcal{P}_N$ is called a *finite-dimensional Dirichlet prior.*

A nice feature of the finite-dimensional Dirichlet measure is the exchangeability of the random weights, and the simplicity of the Dirichlet construction. It is perhaps not too surprising,

therefore, that (16) seems to have been discovered independently by several authors in different contexts and under different names. Pitman (1996, Example 16) attributes this measure to the early work of Fisher, referring to it as "Fisher's model" (the general Fisher model, however, allows for an arbitrary shape parameter $\kappa > 0$ in place of α/N). Kingman (1975), Watterson (1976), Patil & Taillie (1977), among others, discuss the limiting form of Fisher's model. Muliere & Secchi (1995) looked at proper Bayesian bootstraps and named $\mathcal{P}_N$ a "Dirichlet-multinomial process" (as noted by Muliere & Secchi this process is not to be confused with the Dirichlet-multinomial process studied by Lo 1986, whose finite-dimensional distribution is multinomial). Liu (1996) termed (16) an "m-spike" model, and Ishwaran & Zarepour (2000) studied (16) in Gibbs sampling procedures for fitting Bayesian nonparametric models. Also, see Neal (2000) who discussed the prior in the context of Bayesian mixture models, and Green & Richardson (2001), who also studied mixture models and referred to (16) as a "Dirichlet/multinomial allocation" model. Surprisingly, however, given the wide interest in this process, there seems to be very little known about its statistical properties. Here we will study its use as a prior in Bayesian nonparametric problems, providing a characterization of its posterior, as well as quantifying how well it approximates the Dirichlet process.

Intuitively, it is not surprising that $\mathcal{P}_N$ is a good approximation of the DP(αH) measure. Notice that for a fixed value of $\mathbf{Z} = (Z_1, \ldots, Z_N)$, it is already a Dirichlet process:

$$(\mathcal{P}_N \mid \mathbf{Z}) \sim \mathrm{DP}\{\alpha \xi_N(\mathbf{Z}, \cdot)\},$$

where

$$\xi_N(\mathbf{Z}, \cdot) = \frac{1}{N} \sum_{k=1}^{N} \delta_{Z_k}(\cdot)$$

is the empirical measure of $Z_1, \ldots, Z_N$. Thus, since $\xi_N \approx H$, we expect $\mathcal{P}_N \approx \mathrm{DP}(\alpha H)$. Somewhat more formally, notice that $\mathcal{P}_N$ can be expressed as a mixture of Dirichlet processes because $\xi_N(\mathbf{Z}, \cdot)$ is a random measure. Conditioning on $\mathbf{Z}$, and then integrating, we can write

$$\mathcal{P}_N \stackrel{\mathcal{D}}{=} \int \mathrm{DP}\{\alpha \xi_N(\mathbf{Z}, \cdot)\}\, dH^N(\mathbf{Z}).$$

A conditioning argument will also show that $\mathcal{P}_N$ can be used to approximate integrable functionals of the Dirichlet process. See the Appendix for a proof of the following theorem.

THEOREM 2. *For each real-valued measurable function g, integrable with respect to H, $\mathcal{P}_N(g) \xrightarrow{\mathcal{D}} \mathcal{P}(g)$, where $\mathcal{P} = \mathrm{DP}(\alpha H)$.*

Note that the theorem gives a stronger form of convergence than weak convergence, which only applies to bounded and continuous functions. See Muliere & Secchi (1995) for a proof of weak convergence for $\mathcal{P}_N$. See also Ishwaran & Zarepour (2000) where Theorem 2 is discussed (without proof).

4.1. Clustering and marginal density approximations.

Another method for comparing $\mathcal{P}_N$ to the DP(αH) process is to compare their clustering behaviour under sampling. Call $\mathbf{Y} = (Y_1, \ldots, Y_n)$ a sample from $\mathcal{P}_N$ (or $\mathcal{P}$) if Y_i given P are iid P and P has the law $\mathcal{P}_N$ (or $\mathcal{P}$). With a nonatomic distribution H, clustering events for $\mathbf{Y}$ can be reinterpreted in terms of partitions of the set $\{1, \ldots, n\}$. Towards this end, we introduce the following notation. Let $\mathbf{P} = \{C_j : j = 1, \ldots, n(\mathbf{P})\}$ be a partition of the set $\{1, \ldots, n\}$, where C_j is the jth cell of the partition, e_j is the number of elements in a cell C_j, and $n(\mathbf{P})$ is the number of cells in the partition. Also, for each $\gamma > 0$, let $\gamma^{[0]} = 1$ and $\gamma^{[j]} = \gamma(\gamma+1)\cdots(\gamma+j-1)$ for $j \geq 1$. For a proof of the following theorems, see the Appendix.

THEOREM 3. *Let D_N and D_∞ equal the number of distinct values in $\mathbf{Y}$ when sampled under $\mathcal{P}_N$ and $\mathcal{P} = \mathrm{DP}(\alpha H)$, respectively. If H is nonatomic, then*

$$\frac{N!}{N^k(N-k)!} \leq \frac{\Pr\{D_N = k\}}{\Pr\{D_\infty = k\}} \leq n^{\alpha k/N}, \qquad \text{for } k = 1, \ldots, \min(n, N).$$

Notice that the two distributions agree in the limit as $N \to \infty$ because both left and right-hand sides converge to one for each k. See Ishwaran & Zarepour (2000), where Theorem 3 was stated and its implication discussed in more detail. The same type of calculations used in the proof of the theorem can also be used to compute the $\mathcal{L}_1$-distance between the marginal densities under $\mathcal{P}_N$ and the Dirichlet process, similar to the calculation (15).

THEOREM 4. *Let m_∞ and m_N be the marginal densities for* (10) *and* (13) *under the* $\mathrm{DP}(\alpha H)$ *prior and the prior* (16), *respectively. Suppose that H is nonatomic and $N \geq n$. Then,*

$$\begin{aligned} \int |m_N(\mathbf{X}) - m_\infty(\mathbf{X})|\, d\lambda^n(\mathbf{X}) &\leq \sum_{\mathbf{P}} \pi(\mathbf{P}) \left| 1 - \frac{N!}{N^k(N-k)!} \prod_{j=1}^{k=n(\mathbf{P})} \frac{(1+\alpha/N)^{[e_j-1]}}{(e_j-1)!} \right| \\ &\leq \sum_{\mathbf{P}} \pi(\mathbf{P}) \left\{ \left| 1 - \frac{N!}{N^k(N-k)!} \right| \vee \left| 1 - n^{\alpha k/N} \right| \right\}, \quad (17) \end{aligned}$$

where we write $k = n(\mathbf{P})$ for clarity, the above the sum is over all partitions $\mathbf{P}$ of $\{1, \ldots, n\}$, and

$$\pi(\mathbf{P}) = \frac{\alpha^k}{\alpha^{[n]}} \prod_{j=1}^{k} (e_j - 1)!$$

is the probability for a partition $\mathbf{P}$ under the $\mathrm{DP}(\alpha H)$ *prior.*

The theorem is easily adjusted for the case when $N < n$. In this case, replace the expression following $\pi(\mathbf{P})$ with the value one for each partition $\mathbf{P}$ with $n(\mathbf{P}) > N$. Although the bound given in (17) is more difficult to work with than the bound (15) for the almost sure truncation, it is more typical for theoretical results for finite-dimensional Dirichlet priors to be simpler because of their connection to the Dirichlet process. For example, in the forthcoming section, we are able to work out an explicit and informative characterization of its posterior.

4.2. Posterior for $\mathcal{P}_N$.

The following theorem characterizes the posterior of $\mathcal{P}_N$ in the nonparametric hierarchical model (13). It extends Theorem 3 of Ishwaran & James (2001a) who considered priors $\mathcal{P}$ with two-parameter Poisson–Dirichlet laws. The proof relies on Theorem 1 of Lo (1984) and Corollary 20 of Pitman (1996), while the representation for $\mathcal{P}_N(\cdot \mid \mathbf{Y})$ appearing in the theorem was suggested to us by Lancelot James (personal communication).

THEOREM 5. *Write $\mathcal{P}_N(\cdot \mid \mathbf{X})$ for the posterior from* (13) *under the prior* (16). *Then,*

$$\mathcal{P}_N(\cdot \mid \mathbf{X}) = \int_{\mathcal{Y}^n} \mathcal{P}_N(\cdot \mid \mathbf{Y})\, d\nu_N(\mathbf{Y} \mid \mathbf{X}), \quad (18)$$

where

$$d\nu_N(\mathbf{Y} \mid \mathbf{X}) \propto \prod_{i=1}^{n} f(X_i \mid Y_i) \prod_{i=1}^{n} \left\{ \frac{\alpha(1 - m_i/N)}{\alpha + i - 1} H(dY_i) + \sum_{j=1}^{m_i} \frac{n^*_{j,i} + \alpha/N}{\alpha + i - 1} \delta_{Y^*_{j,i}}(dY_i) \right\},$$

and $Y^*_{1,i}, \ldots, Y^*_{m_i,i}$ are the m_i unique values in the sequence $Y_1, \ldots, Y_{i-1}$, each occurring with frequency $n^*_{j,i}$ (note: $m_1 = 0$). Moreover,

$$\mathcal{P}_N(\cdot \mid \mathbf{Y}) = \sum_{j=1}^{m} p^*_j \, \delta_{Y^*_j}(\cdot) + p^*_{m+1} \mathcal{P}^*_N(\cdot),$$

where $Y^*_1, \ldots, Y^*_m$ are the unique values in the full sequence $Y_1, \ldots, Y_n$ occurring each with frequencies n^*_j, and

$$(p^*_1, \ldots, p^*_m, p^*_{m+1}) \sim \text{Dirichlet}(n^*_1 + \alpha/N, \ldots, n^*_m + \alpha/N, \alpha(1 - m/N))$$

is independent of

$$\mathcal{P}^*_N(\cdot) = \sum_{j=1}^{N-m} \left\{ V_j \prod_{i=1}^{j-1} (1 - V_i) \right\} \delta_{Z_j}(\cdot), \tag{19}$$

where V_j are independent Beta$(1 + \alpha/N, \alpha(1 - (m+j)/N))$ random variables for $j < N - m$ ($V_{N-m} = 1$), and all variables are independent of Z_j which are iid H.

Note that the characterization (18) is in direct analogue to (11). Hence one can draw values from the laws for functionals of the posterior using the same ideas discussed earlier. Thus, to draw from the law of some arbitrary functional, draw $\mathbf{Y}$ from $\nu_N(\cdot \mid \mathbf{X})$ and then use this value to compute the functional of the corresponding random measure $\mathcal{P}_N(\cdot \mid \mathbf{Y})$. Notice here that no form of approximation is needed for $\mathcal{P}_N(\cdot \mid \mathbf{Y})$, thanks to its simple finite-sum expression. We also note that Monte Carlo techniques are readily available for drawing $\mathbf{Y}$ from $\nu_N(\cdot \mid \mathbf{X})$. For example, see Ishwaran & James (2001a) who describe a Pólya urn Gibbs sampler for drawing from such measures.

Remark. Corollary 20 of Pitman (1996) also explicitly identifies (19) as an example of Fisher's model. That is, (19) is a finite-dimensional Dirichlet prior with $N - m$ parameters equal to α/N. This in fact implies the following equivalent representation for $\mathcal{P}_N$:

$$\mathcal{P}_N(\cdot) = \sum_{k=1}^{N} \frac{G_k}{G} \delta_{Z_k}(\cdot) \stackrel{\mathcal{D}}{=} \sum_{j=1}^{N} \left\{ V_j \prod_{i=1}^{j-1} (1 - V_i) \right\} \delta_{Z_j}(\cdot), \tag{20}$$

where V_j are independent Beta$(1 + \alpha/N, \alpha(1 - j/N))$ random variables for $j < N$ and $V_N = 1$. This relationship can also be deduced from Patil & Taillie (1977, Example 2.7), who note that the random weights on the right-hand side of (20) are a size-biased random permutation of the ranked weights on the left-hand side.

APPENDIX: PROOFS

Proof of Theorem 1: Let T be a measurable map $T: [0, \infty] \times [0, \infty] \to [0, \infty]$. Then,

$$\zeta(\cdot) = \sum_{k=1}^{\infty} \delta_{T(W_k, \Gamma_k)}(\cdot)$$

is a Poisson point process with mean measure $\nu \circ T^{-1}$, where $d\nu = dF \times d\lambda$, F is the distribution for W_k, and λ is Lebesgue measure on $[0, \infty]$. See Resnick (1987, Chapter 3).

The Laplace functional for ζ is

$$\mathrm{E}\left\{e^{-\zeta(g)}\right\} = \exp\left[-\int_0^{\infty} \left\{1 - e^{-g(z)}\right\} \nu \circ T^{-1}(dz) \right],$$

where g is a nonnegative bounded measurable function on $[0,\infty] \times [0,\infty]$. This applies to $T(x,y) = N^{-1}(xy)$, which is measurable by the continuity of N. For this choice of T,

$$\nu \circ T^{-1}[t,\infty] = \nu\{(x,y) : N^{-1}(xy) \geq t\}.$$

Remembering that N is a decreasing function, this equals

$$\int_0^\infty \int_0^{N(t)/y} dF(x)\, dy = \int_0^\infty F\{N(t)/y\}\, dy = \int_0^\infty \mathrm{P}\{N(t)\, W_1^{-1} \geq u\}\, du$$

where the last equality can be seen to equal $\mathrm{E}\{N(t)\, W_1^{-1}\}$. Therefore,

$$\mathrm{E}\{e^{-\zeta(g)}\} = \exp\left[-\mathrm{E}(W_1^{-1}) \int_0^\infty \{1 - e^{-g(z)}\}\, dN(z)\right],$$

which is also the Laplace functional for the Poisson point process $\sum_{k=1}^\infty \delta_{N^{-1}(h\Gamma_k)}(\,\cdot\,)$, where $h = 1/\mathrm{E}(W_1^{-1})$. Therefore, by the uniqueness of the Laplace functional

$$\sum_{k=1}^\infty \delta_{N^{-1}(h\Gamma_k)}(\,\cdot\,) \stackrel{\mathcal{D}}{=} \sum_{k=1}^\infty \delta_{N^{-1}(W_k\Gamma_k)}(\,\cdot\,).$$

The identity (6) follows when $h = 1$. □

Proof of Theorem 2: As noted by Sethuraman (1994), the representation (7) can be used to decompose the Dirichlet process random probability measure into two conditionally independent components. Sethuraman (1994) observed that

$$\mathcal{P}(\,\cdot\,) \stackrel{\mathcal{D}}{=} V_1 \delta_{Z_1}(\,\cdot\,) + (1 - V_1)\mathcal{P}(\,\cdot\,), \tag{21}$$

where on the right-hand side $\mathcal{P}$ is independent of V_1 and Z_1.

By the distributional equation (21), it follows that the characteristic function ψ of $\mathcal{P}(g)$ equals

$$\psi(t) = \mathrm{E}\exp\left[it\{V_1 g(Z_1) + (1 - V_1)\mathcal{P}(g)\}\right] = \mathrm{E}\left[\phi(tV_1)\psi\{t(1 - V_1)\}\right], \tag{22}$$

where we have used the independence of V_1, Z_1 and $\mathcal{P}(g)$ and we write ϕ for the characteristic function of $g(Z_1)$. Any characteristic function which satisfies (22) must be the characteristic function for $\mathcal{P}(g)$. See Lemma 3.3 of Sethuraman (1994).

Let ψ_N represent the characteristic function for $\mathcal{P}_N(g)$. Recall that for a fixed value of $\mathbf{Z}$, the measure $\mathcal{P}_N$ is a Dirichlet process. Therefore, using a similar argument as above, while conditioning on the value for $\mathbf{Z}$, we have

$$\psi_N(t) = \mathrm{E}\left(\left[\frac{1}{N}\sum_{k=1}^N \exp\{itV_1 g(Z_k)\}\right]\psi_{N\,|\,\mathbf{Z}}\{t(1 - V_1)\}\right), \tag{23}$$

where

$$\psi_{N\,|\,\mathbf{Z}}(t) = \mathrm{E}\left[\exp\{it\mathcal{P}_N(g)\}\,|\,\mathbf{Z}\right].$$

Rewrite (23) as two terms, where the first term replaces the average on the right-hand side of (23) by its almost sure limit $\phi(tV_1)$, while the second term is the remaining expression. Using the strong law of large numbers and Lebesgue's dominated convergence theorem, we can deduce that the remainder term goes to zero (note that $\psi_{N\,|\,\mathbf{Z}}(t)$ is bounded by one), while integrating over $\mathbf{Z}$ and taking limits for the first term we have

$$\psi_\infty(t) = \mathrm{E}\left[\phi(tV_1)\psi_\infty\{t(1 - V_1)\}\right],$$

where ψ_∞ is the limit of ψ_N. Furthermore, because $\mathcal{P}_N(g)$ is tight (by the integrability of g), it follows that ψ_∞ must be the characteristic function for the limiting law of $\mathcal{P}_N(g)$. Therefore, since ψ_∞ satisfies the same equation as (22), we deduce that $\mathcal{P}_N(g)$ converges weakly to $\mathcal{P}(g)$. □

Proof of Theorem 3: The probabilities in question are obtained by summing over all partitions $\mathbf{P}$ with $n(\mathbf{P}) = k$. Let $\pi(\mathbf{P})$ denote the prior probability for a partition $\mathbf{P}$ under the DP(αH) process. Following the work of Antoniak (1974), if H is nonatomic and $n(\mathbf{P}) = k$,

$$\pi(\mathbf{P}) = \frac{\alpha^k}{\alpha^{[n]}} \prod_{j=1}^{k} (e_j - 1)!. \tag{24}$$

Under $\mathcal{P}_N$, this same set has the probability

$$\begin{aligned} \pi_N(\mathbf{P}) &= \sum_{\{i_1 \neq \cdots \neq i_k\}} \mathrm{E}\,(p_{i_1}^{e_1} \cdots p_{i_k}^{e_k}) = \frac{N!}{(N-k)!}\,\mathrm{E}\,(p_1^{e_1} \cdots p_k^{e_k}) \\ &= \frac{\alpha^k N!}{N^k \alpha^{[n]} (N-k)!} \prod_{j=1}^{k} (1 + \alpha/N)^{[e_j - 1]}. \end{aligned} \tag{25}$$

Dividing (25) by (24), we get

$$\Delta(\mathbf{P}) = \frac{\pi_N(\mathbf{P})}{\pi(\mathbf{P})} = \frac{N!}{N^k (N-k)!} \prod_{j=1}^{k} \frac{(1 + \alpha/N)^{[e_j - 1]}}{(e_j - 1)!}. \tag{26}$$

The first term on the extreme right is bounded above by 1, while the term to the right of this is bounded below by 1. Hence,

$$\frac{N!}{N^k (N-k)!} \leq \Delta(\mathbf{P}) \leq \prod_{j=1}^{k} \frac{(1 + \alpha/N)^{[e_j - 1]}}{(e_j - 1)!} = \exp\left\{ \sum_{j=1}^{k} \sum_{r=1}^{e_j - 1} \log\left(1 + \frac{\alpha}{rN}\right) \right\}. \tag{27}$$

The right-hand side can be bounded by $\exp(\alpha k \log n/N)$. Summing over all partitions $\mathbf{P}$ with $n(\mathbf{P}) = k$ gives the desired bound. □

Proof of Theorem 4: Let $\pi_{\mathbf{p}}$ denote the distribution for $\mathbf{p}$, the vector of random weights. Integrating over the random measure P in the integrand of $m_N(\mathbf{X})$ (see for example (14)), and keeping $\mathbf{Z}$ fixed until the end, we have

$$\begin{aligned} m_N(\mathbf{X}) &= \iiint \prod_{i=1}^{n} f(X_i \mid Y_i) \prod_{i=1}^{n} \left\{ \sum_{k=1}^{N} p_k\, \delta_{Z_k}(dY_i) \right\} d\pi_{\mathbf{p}}(\mathbf{p})\, dH^N(\mathbf{Z}) \\ &= \int \left\{ \sum_{\mathbf{P}} \sum_{\{i_1 \neq \cdots \neq i_k\}} \mathrm{E}\,(p_{i_1}^{e_1} \cdots p_{i_k}^{e_k}) \prod_{j=1}^{k=n(\mathbf{P})} \prod_{i \in C_j} f(X_i \mid Z_{i_j}) \right\} dH^N(\mathbf{Z}) \\ &= \sum_{\mathbf{P}} \pi_N(\mathbf{P}) \prod_{j=1}^{n(\mathbf{P})} \int_{\mathcal{Y}} \prod_{i \in C_j} f(X_i \mid Y)\, dH(Y). \end{aligned}$$

Lo (1984, Lemma 2) established a parallel identity for m_∞ with $\pi_N(\mathbf{P})$ replaced with $\pi(\mathbf{P})$ in the above expansion. The bounds in (17) now follow automatically from (26) and (27). □

Proof of Theorem 5: The characterization (18) follows by mimicking the proof of Theorem 1 of Lo (1984), replacing the marginal distribution for $\mathbf{Y}$ under the Dirichlet process with that under $\mathcal{P}_N$:

$$m(dY_1, \ldots, dY_n) = \prod_{i=1}^{n} \left\{ \frac{\alpha(1 - m_i/N)}{\alpha + i - 1} H(dY_i) + \sum_{j=1}^{m_i} \frac{n_{j,i}^* + \alpha/N}{\alpha + i - 1} \delta_{Y_{j,i}^*}(dY_i) \right\}.$$

To arrive at the representation for $\mathcal{P}_N(\cdot \mid \mathbf{Y})$, deduce by equations (51) and (52) prior to Corollary 20 of Pitman (1996) that

$$\mathcal{P}_N(\cdot \mid \mathbf{Y}) = \sum_{j=1}^{m} \left\{ V_j \prod_{i=1}^{j-1}(1-V_i) \right\} \delta_{Y_j^*}(\cdot) + \sum_{j=m+1}^{N} \left\{ V_j \prod_{i=1}^{j-1}(1-V_i) \right\} \delta_{Z_j}(\cdot),$$

where V_j are independent Beta$(n_j^* + \alpha/N, \alpha(1-j/N) + \sum_{i=j+1}^{m} n_i^*)$ random variables for $1 \le j \le m$, while for $m+1 \le j < N-1$, they are independent Beta$(1+\alpha/N, \alpha(1-j/N))$ random variables, and $V_N = 1$. Let p_j^* denote the random weights in the first sum. Using the fact that

$$1 - \sum_{j=1}^{m} p_j^* = \prod_{i=1}^{m}(1-V_i),$$

a little bit of rearrangement gives

$$\mathcal{P}_N(\cdot \mid \mathbf{Y}) \stackrel{\mathcal{D}}{=} \sum_{j=1}^{m} p_j^* \, \delta_{Y_j^*}(\cdot) + \left(1 - \sum_{j=1}^{m} p_j^*\right) \sum_{j=1}^{N-m} \left\{ V_j^* \prod_{i=1}^{j-1}(1-V_i^*) \right\} \delta_{Z_j}(\cdot)$$

where $V_j^* = V_{m+j}$ for $j = 1, \ldots, N-m$. It is straightforward to verify that the joint distribution for the p_j^* has the stated Dirichlet law. □

ACKNOWLEDGEMENTS

The authors thank Lancelot James for his help with Theorem 5 and for pointing out the size-biased stick-breaking representation for Fisher's model. They also thank the Editor, Richard Lockhart, for his many useful suggestions which greatly improved the final manuscript. Dr. Zarepour's research was supported by grant funds from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

C. E. Antoniak (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2, 1152–1174.

D. A. Berry & R. Christensen (1979). Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes. *The Annals of Statistics*, 7, 558–568.

D. Blackwell & J. B. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1, 353–355.

L. Bondesson (1982). On simulation from infinitely divisible distributions. *Advances in Applied Probability*, 14, 885–869.

R. J. Connor & J. E. Mosimann (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, 64, 194–206.

P. Diaconis & J. Kemperman (1996). Some new tools for Dirichlet priors. In *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds.), Oxford University Press, pp. 97–106.

P. Donnelly & P. Joyce (1989). Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Processes and Their Applications*, 31, 89–103.

H. Doss (1985). Bayesian nonparametric estimation of the median; Part I: computation of the estimates. *The Annals of Statistics*, 13, 1432–1444.

H. Doss (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *The Annals of Statistics*, 22, 1763–1786.

W. Ewens (1990). Population genetics theory—the past and the future. In *Mathematical and Statistical Developments of Evolutionary Theory: Proceedings of the NATO Advanced Study Institute and Séminaire de mathématiques supérieures on Mathematical and Statistical Developments of Evolutionary Theory, Montreal (August 3–21, 1987)*, (S. Lessard, ed.), Kluwer Academic, pp. 177–227.

J. Fabius (1964). Asymptotic behavior of Bayes estimates. *The Annals of Mathematical Statistics*, 35, 846–856.

T. S. Ferguson (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209–230.

T. S. Ferguson (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2, 615–629.

T. S. Ferguson & M. J. Klass (1972). A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics*, 43, 1634–1643.

D. A. Freedman (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Mathematical Statistics*, 34, 1386–1403.

A. Gelfand & A. Kottas (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11, 289–305.

P. Green & S. Richardson (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, 28, 355–377.

R. C. Hannum, M. Hollander & N. Langberg (1981). Distributional results for random functionals of a Dirichlet process. *The Annals of Probability*, 9, 665–670.

F. M. Hoppe (1987). The sampling theory of neutral alleles and an urn model in population genetics. *Journal of Mathematical Biology*, 25, 123–159.

H. Ishwaran & L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96, 161–173.

H. Ishwaran & L. F. James (2002). Approximate Dirichlet process computing for finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical Statistics*, in press.

H. Ishwaran & M. Zarepour (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87, 371–390.

N. L. Johnson, S. Kotz & N. Balakrishnan (1997). Multivariate Ewens distribution, Ch. 41 of their *Discrete Multivariate Distributions*. Wiley, New York, pp. 232–246.

J. F. C. Kingman (1975). Random discrete distributions. *Journal of the Royal Statistical Society Series B*, 37, 1–22.

R. Korwar & M. Hollander (1973). Contributions to the theory of Dirichlet processes. *The Annals of Probability*, 1, 705–711.

L. Kuo (1986). Computations of mixtures of Dirichlet processes. *SIAM Journal of Scientific and Statistical Computing*, 7, 60–71.

J. S. Liu (1996). Nonparametric hierarchical Bayes via sequential imputations. *The Annals of Statistics*, 24, 911–930.

A. Y. Lo (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12, 351–357.

A. Y. Lo (1986). Bayesian statistical inference for sampling a finite population. *The Annals of Statistics*, 14, 1226–1233.

J. W. McCloskey (1965). A model for the distribution of individuals by species in an environment. Unpublished doctoral dissertation, Michigan State University, East Lansing.

P. Muliere & P. Secchi (1995). A note on a proper Bayesian bootstrap. Technical report No 18, Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, Italy.

P. Muliere & P. Secchi (1996). Bayesian nonparametric predictive inference and bootstrap techniques. *Annals of the Institute of Statistical Mathematics*, 48, 663–673.

P. Muliere & L. Tardella (1998). Approximating distributions of random functionals of Ferguson–Dirichlet priors. *The Canadian Journal of Statistics*, 26, 283–297.

R. M. Neal (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249–265.

G. P. Patil & C. Taillie (1977). Diversity as a concept and its implications for random communities. *Bulletin of the International Statistical Institute*, 47, 497–515.

M. Perman, J. Pitman & M. Yor (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92, 21–39.

J. Pitman (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T. S. Ferguson, L. S. Shapley & J. B. MacQueen, eds.), IMS Lecture Notes–Monograph Series, vol. 30, pp. 245–267.

J. Pitman & M. Yor (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25, 855–900.

S. I. Resnick (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

J. Sethuraman (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

J. Sethuraman & R. C. Tiwari (1982). Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical Decision Theory and Related Topics III* (S. S. Gupta & J. O. Berger, eds.), Academic Press, vol. 2, pp. 305–315.

G. A. Watterson (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *Journal of Applied Probability*, 13, 639–651.

H. Yamato (1984). Characteristic functions of means of distributions chosen from a Dirichlet process. *The Annals of Probability*, 12, 262–267.

Received 5 January 2001
Accepted 28 January 2002

Hemant ISHWARAN: ishwaran@bio.ri.ccf.org
Department of Biostatistics and Epidemiology
Cleveland Clinic Foundation, Cleveland, OH 44195, USA

Mahmoud ZAREPOUR: zarepour@mathstat.uottawa.ca
Department of Mathematics and Statistics
University of Ottawa, Ottawa, Ontario
Canada K1N 6N5