Statistical Science
2013, Vol. 28, No. 2, 209–222
DOI: 10.1214/12-STS407

Defining Predictive Probability Functions for Species Sampling Models

Jaeyong Lee, Fernando A. Quintana, Peter Müller and Lorenzo Trippa

Abstract. We review the class of species sampling models (SSM). In particular, we investigate the relation between the exchangeable partition probability function (EPPF) and the predictive probability function (PPF). It is straightforward to define a PPF from an EPPF, but the converse is not necessarily true. In this paper we introduce the notion of putative PPFs and show novel conditions for a putative PPF to define an EPPF. We show that all possible PPFs in a certain class have to define (unnormalized) probabilities for cluster membership that are linear in cluster size. We give a new necessary and sufficient condition for arbitrary putative PPFs to define an EPPF. Finally, we show posterior inference for a large class of SSMs with a PPF that is not linear in cluster size and discuss a numerical method to derive its PPF.

Key words and phrases: Species sampling prior, exchangeable partition probability functions, prediction probability functions.

1. INTRODUCTION

The status of the Dirichlet process (Ferguson, 1973) (DP) among nonparametric priors is comparable to that of the normal distribution among finite-dimensional distributions. This is in part due to the marginalization property: a random sequence sampled from a random probability measure with a Dirichlet process prior forms marginally a Polya urn sequence (Blackwell and MacQueen, 1973). Markov chain Monte Carlo simulation based on the marginalization property has been the central computational tool for the DP and facilitated a wide variety of applications. See MacEachern (1994), Escobar and West (1995) and MacEachern and Müller (1998), to name just a few. In Pitman (1995, 1996), the species sampling model (SSM) is proposed as a generalization of the DP. SSMs can be used as flexible alternatives to the popular DP model in nonparametric Bayesian inference. The SSM is defined as the directing random probability measure of an exchangeable species sampling sequence which is defined as a generalization of the Polya urn sequence. The SSM has a marginalization property similar to the DP. It therefore enjoys the same computational advantage as the DP while it defines a much wider class of random probability measures. For its theoretical properties and applications, we refer to Ishwaran and James (2003), Lijoi, Mena and Prünster (2005), Lijoi, Prünster and Walker (2005), James (2008), Navarrete, Quintana and Müller (2008), James, Lijoi and Prünster (2009) and Jang, Lee and Lee (2010).

Suppose $(X_1, X_2, \ldots)$ is a sequence of random variables. In a traditional application the sequence arises as a random sample from a large population of units, and X_i records the species of the ith individual in the sample. This explains the name SSM. Let $\tilde{X}_j$ be the jth distinct species to appear. Let n_{jn} be the number of times the jth species $\tilde{X}_j$ appears in $(X_1, \ldots, X_n)$, $j = 1, 2, \ldots$, and

$$\mathbf{n}_n = (n_{jn}, j = 1, \ldots, k_n),$$

Jaeyong Lee is Professor, Department of Statistics, Seoul National University, Seoul, 151-747, South Korea (e-mail: leejyc@gmail.com). Fernando A. Quintana is Professor, Departamento de Estadística, Pontificia Universidad Católica de Chile, Macul, Santiago 22, Chile (e-mail: quintana@mat.puc.cl). Peter Müller is Professor, Department of Mathematics, University of Texas at Austin, Austin, Texas 78712-1202, USA (e-mail: pmueller@math.utexas.edu). Lorenzo Trippa is Assistant Professor, Department of Biostatistics, Harvard University, and Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA (e-mail: ltrippa@jimmy.harvard.edu).

where $k_n = k_n(\mathbf{n}_n) = \max\{j : n_{jn} > 0\}$ is the number of different species to appear in $(X_1, \ldots, X_n)$. The sets $\{i \le n : X_i = \tilde{X}_j\}$ define clusters that partition the index set $\{1, \ldots, n\}$. When n is understood from the context we just write n_j, $\mathbf{n}$ and k or $k(\mathbf{n})$.

We now give three alternative characterizations of species sampling sequences: (i) by the predictive probability function, (ii) by the driving measure of the exchangeable sequence, and (iii) by the underlying exchangeable partition probability function.

PPF

Let ν be a diffuse (or nonatomic) probability measure on a complete separable metric space $\mathcal{X}$ equipped with Borel σ-field. An exchangeable sequence $(X_1, X_2, \ldots)$ is called a species sampling sequence (SSS) if $X_1 \sim \nu$ and

$$X_{n+1} \mid X_1, \ldots, X_n \sim \sum_{j=1}^{k_n} p_j(\mathbf{n}_n)\delta_{\tilde{X}_j} + p_{k_n+1}(\mathbf{n}_n)\nu, \tag{1}$$

where δ_x is the degenerate probability measure at x. Examples of SSS include the Pólya urn sequence $(X_1, X_2, \ldots)$ whose distribution is the same as the marginal distribution of independent observations from a Dirichlet random distribution F, that is, $X_1, X_2, \ldots \mid F \overset{\text{i.i.d.}}{\sim} F$ with $F \sim \mathrm{DP}(\alpha\nu)$, where $\alpha > 0$. The conditional distribution of the Pólya urn sequence is

$$X_{n+1} \mid X_1, \ldots, X_n \sim \sum_{j=1}^{k_n} \frac{n_j}{n+\alpha}\delta_{\tilde{X}_j} + \frac{\alpha}{n+\alpha}\nu.$$

This marginalization property has been a central tool for posterior simulation in DP mixture models, which benefit from the fact that one can integrate out F using the marginalization property. The posterior distribution becomes then free of the infinite-dimensional object F. Thus, Markov chain Monte Carlo algorithms for DP mixtures do not pose bigger difficulties than the usual parametric Bayesian models (MacEachern, 1994; MacEachern and Müller, 1998). Similarly, alternative discrete random distributions have been considered in the literature and proved computationally attractive due to analogous marginalization properties; see, for example, Lijoi, Mena and Prünster (2005, 2007).

The sequence of functions $(p_1, p_2, \ldots)$ in (1) is called a sequence of predictive probability functions (PPF). These are defined on $\mathbb{N}^* = \bigcup_{k=1}^{\infty} \mathbb{N}^k$, where $\mathbb{N}$ is the set of natural numbers, and satisfy the conditions

$$p_j(\mathbf{n}) \ge 0 \quad \text{and} \quad \sum_{j=1}^{k_n+1} p_j(\mathbf{n}) = 1 \quad \text{for all } \mathbf{n} \in \mathbb{N}^*. \tag{2}$$

Motivated by these properties of PPFs, we define a sequence of *putative PPFs* as a sequence of functions $(p_j, j = 1, 2, \ldots)$ defined on $\mathbb{N}^*$ which satisfies (2). Note that not all putative PPFs are PPFs, because (2) does not guarantee exchangeability of $(X_1, X_2, \ldots)$ in (1). Note that the weights $p_j(\cdot)$ depend on the data only indirectly through the cluster sizes $\mathbf{n}_n$. The widely used DP is a special case of a species sampling model, with $p_j(\mathbf{n}_n) \propto n_j$ and $p_{k+1}(\mathbf{n}_n) \propto \alpha$ for a DP with total mass parameter α. The use of p_j in (1) implies

$$p_j(\mathbf{n}) = \mathbb{P}(X_{n+1} = \tilde{X}_j \mid X_1, \ldots, X_n), \quad j = 1, \ldots, k_n,$$
$$p_{k_n+1}(\mathbf{n}) = \mathbb{P}\big(X_{n+1} \notin \{\tilde{X}_1, \ldots, \tilde{X}_{k_n}\} \mid X_1, \ldots, X_n\big).$$

In words, p_j is the probability of the next observation being the jth species (falling into the jth cluster) and p_{k_n+1} is the probability of a new species (starting a new cluster).

An important point in the above definition is that a sequence X_i can be a SSS only if it is exchangeable.

SSM

Alternatively, a SSS can be characterized by the following defining property. An exchangeable sequence of random variables $(X_1, X_2, \ldots)$ is a species sampling sequence if and only if $X_1, X_2, \ldots \mid G$ is a random sample from G where

$$G = \sum_{h=1}^{\infty} P_h \delta_{m_h} + R\nu \tag{3}$$

for some sequence of positive random variables (P_h) and R such that $1 - R = \sum_{h=1}^{\infty} P_h \le 1$ with probability 1, (m_h) is a sequence of independent variables with distribution ν, and (P_i) and (m_h) are independent. See Pitman (1996). The result is an extension of de Finetti's theorem and characterizes the directing random probability measure of the species sample sequence. We call the directing random probability measure G in equation (3) the *SSM* of the SSS (X_i).

EPPF

A third alternative definition of a SSS and corresponding SSM is in terms of the implied probability model on a sequence of random partitions.

Suppose a SSS $(X_1, X_2, \ldots)$ is given. Since the de Finetti measure (3) is partly discrete, there are ties among X_i's. The ties among $(X_1, X_2, \ldots, X_n)$ for a given n induce an equivalence relation in the set $[n] = \{1, 2, \ldots, n\}$, that is, $i \sim j$ if and only if $X_i = X_j$. This equivalence relation on $[n]$, in turn, induces the partition Π_n of $[n]$. Due to the exchangeability of $(X_1, X_2, \ldots)$, it can be easily seen that the random partition Π_n is an exchangeable random partition on $[n]$, that is, for any partition $\{A_1, A_2, \ldots, A_k\}$ of $[n]$, the probability $P(\Pi_n = \{A_1, A_2, \ldots, A_k\})$ is invariant under any permutation on $[n]$ and can be expressed as a function of $\mathbf{n} = (n_1, n_2, \ldots, n_k)$, where n_i is the cardinality of A_i for $i = 1, 2, \ldots, k$. Extending the above argument to the entire SSS, we can get an exchangeable random partition on the natural numbers $\mathbb{N}$ from the SSS. Kingman (1978, 1982) showed a remarkable result, called Kingman's representation theorem, that in fact every exchangeable random partition can be obtained by a SSS.

For any partition $\{A_1, A_2, \ldots, A_k\}$ of $[n]$, we can represent $P(\Pi_n = \{A_1, A_2, \ldots, A_k\}) = p(\mathbf{n})$ for a symmetric function $p : \mathbb{N}^* \to [0, 1]$ satisfying

$$
\begin{aligned}
&p(1) = 1,\\
&p(\mathbf{n}) = \sum_{j=1}^{k(\mathbf{n})+1} p(\mathbf{n}^{j+}) \qquad \text{for all } \mathbf{n} \in \mathbb{N}^*,
\end{aligned}
\tag{4}
$$

where $\mathbf{n}^{j+}$ is the same as $\mathbf{n}$ except that the jth element is increased by 1. This function is called an exchangeable partition probability function (EPPF) and characterizes the distribution of an exchangeable random partition on $\mathbb{N}$.

We are now ready to pose the problem for the present paper. It is straightforward to verify that any EPPF defines a PPF by

$$
p_j(\mathbf{n}) = \frac{p(\mathbf{n}^{j+})}{p(\mathbf{n})}, \qquad j = 1, 2, \ldots, k+1. \tag{5}
$$

The converse is not true. Not every putative $p_j(\mathbf{n})$ defines an EPPF and thus a SSM and a SSS. For example, it is easy to show that $p_j(\mathbf{n}) \propto n_j^2 + 1$, $j = 1, \ldots, k(\mathbf{n})$, does not. In Bayesian data analysis it is often convenient, or at least instructive, to elicit features of the PPF rather than the joint EPPF. Since the PPF is crucial for posterior computation, applied Bayesians tend to focus on it to specify the species sampling prior for a specific problem. For example, the PPF defined by a DP prior implies that the probability of joining an existing cluster is proportional to the cluster size. This is not always desirable. Can the user define an alternative PPF that allocates new observations to clusters with probabilities proportional to alternative functions $f(n_j)$ and still define a SSS? In general, the simple answer is no. We already mentioned that a PPF implies a SSS if and only if it arises as in (5) from an EPPF. But this result is only a characterization. It is of little use for data analysis and modeling since it is difficult to verify whether or not a given PPF arises from an EPPF. In this paper we develop some conditions to address this gap. We consider methods to define PPFs in two different directions. First we give an easily verifiable necessary condition for a putative PPF to arise from an EPPF (Lemma 1) and a necessary and sufficient condition for a putative PPF to arise from an EPPF. A consequence of this result is an elementary proof of the characterization of all possible PPFs with form $p_j(\mathbf{n}) \propto f(n_j)$. This result has been proved earlier by Gnedin and Pitman (2006). Although the result in Section 2 gives necessary and sufficient conditions for a putative PPF to be a PPF, the characterization is not constructive. It does not give any guidance in how to create a new PPF for a specific application. In Section 3 we propose an alternative approach to define a SSM based on directly defining a joint probability model for the P_h in (3). We develop a numerical algorithm to derive the corresponding PPF. This facilitates the use of such models for nonparametric Bayesian data analysis. This approach can naturally create PPFs with very different features than the well-known PPF under the DP.

The literature reports some PPFs with closed-form analytic expressions other than the PPF under the DP prior. There are a few directions which have been explored for constructing extensions of the DP prior and deriving PPFs. The normalization of complete random measures (CRM) has been proposed in Kingman (1975). A CRM such as the generalized gamma process (Brix, 1999), after normalization, defines a discrete random distribution and, under mild assumptions, a SSM. Developments and theoretical results on this approach have been discussed in a series of papers; see, for example, Perman, Pitman and Yor (1992), Pitman (2003) and Regazzini, Lijoi and Prünster (2003). Normalized CRM models have also been studied and applied in Lijoi, Mena and Prünster (2005), Nieto-Barajas, Prünster and Walker (2004) and more

recently in James, Lijoi and Prünster (2009). A second related line of research considered the so-called Gibbs models. In these models the analytic expressions of the PPFs share similarities with the DP model. An important example is the Pitman–Yor process. Contributions include Gnedin and Pitman (2005), Lijoi, Mena and Prünster (2007), Lijoi, Prünster and Walker (2008a, 2008b) and Gnedin, Haulk and Pitman (2010). Lijoi and Prünster (2010) provide a recent overview on major results from the literature on normalized CRM and Gibbs-type partitions.

2. WHEN DOES A PPF IMPLY AN EPPF?

Suppose we are given a putative PPF (p_j). Using equation (5), one can attempt to define a function $p:\mathbb{N}^*\to[0,1]$ inductively by the following mapping:

$$p(1)=1,$$
$$p(\mathbf{n}^{j+})=p_j(\mathbf{n})p(\mathbf{n}) \quad \text{for all } \mathbf{n}\in\mathbb{N} \text{ and } j=1,2,\ldots,k(\mathbf{n})+1. \tag{6}$$

In general, equation (6) does not lead to a unique definition of $p(\mathbf{n})$ for each $\mathbf{n}\in\mathbb{N}^*$. For example, let $\mathbf{n}=(2,1)$. Then, $p(2,1)$ could be computed in two different ways as $p_2(1)p_1(1,1)$ and $p_1(1)p_2(2)$ which correspond to partitions $\{\{1,3\},\{2\}\}$ and $\{\{1,2\},\{3\}\}$, respectively. If $p_2(1)p_1(1,1)\neq p_1(1)p_2(2)$, equation (6) does not define a function $p:\mathbb{N}^*\to[0,1]$. The following lemma shows a condition for a PPF for which equation (6) leads to a valid unique definition of $p:\mathbb{N}^*\to[0,1]$.

Suppose $\Pi=\{A_1,A_2,\ldots,A_k\}$ is a partition of $[n]$ with clusters indexed in the order of appearance. For $1\le m\le n$, let Π_m be the restriction of Π on $[m]$. Let $\mathbf{n}(\Pi)=(n_1,\ldots,n_k)$, where n_i is the cardinality of A_i, and let $\Pi(i)$ be the class index of element i in partition Π and $\Pi([n])=(\Pi(1),\ldots,\Pi(n))$.

LEMMA 1. *If and only if a putative PPF (p_j) satisfies*

$$p_i(\mathbf{n})p_j(\mathbf{n}^{i+})=p_j(\mathbf{n})p_i(\mathbf{n}^{j+}) \quad \textit{for all } \mathbf{n}\in\mathbb{N}^*, i,j=1,2,\ldots,k(\mathbf{n})+1, \tag{7}$$

then p defined by (6) is a function from $\mathbb{N}^*$ to $[0,1]$, that is, p in (6) is uniquely defined.

PROOF. Let $\mathbf{n}=(n_1,\ldots,n_k)$ with $\sum_{i=1}^k n_i=n$ and Π and Ω be two partitions of $[n]$ with $\mathbf{n}(\Pi)=\mathbf{n}(\Omega)=\mathbf{n}$. Let $p^{\Pi}(\mathbf{n})=\prod_{i=1}^{n-1}p_{\Pi(i+1)}(\mathbf{n}(\Pi_i))$ and $p^{\Omega}(\mathbf{n})=\prod_{i=1}^{n-1}p_{\Omega(i+1)}(\mathbf{n}(\Omega_i))$. We need to show that $p^{\Pi}(\mathbf{n})=p^{\Omega}(\mathbf{n})$. Without loss of generality, we can assume $\Pi([n])=(1,\ldots,1,2,\ldots,2,\ldots,k,\ldots,k)$, where i is repeated n_i times for $i=1,\ldots,k$. Note that $\Omega([n])$ is just a certain permutation of $\Pi([n])$ and by a finite times of swapping two consecutive elements in $\Omega([n])$, one can change $\Omega([n])$ to $\Pi([n])$. Thus, it suffices to show when $\Omega([n])$ is different from $\Pi([n])$ in only two consecutive positions. But, this is guaranteed by condition (7).

The opposite is easy to show. Assume p_j defines a unique $p(\mathbf{n})$. Consider (7) and multiply on both sides with $p(\mathbf{n})$. By assumption, we get on either side $p(\mathbf{n}^{i+j+})$. This completes the proof. □

Note that the conclusion of Lemma 1 is not (yet) that p is an EPPF. The missing property is exchangeability, that is, invariance of p with respect to permutations of the group indices $j=1,\ldots,k(\mathbf{n})$. When the function p, recursively defined by expression (6), satisfies the balance imposed by equation (7) it is called the *partially exchangeable probability function* (Pitman, 1995, 2006) and the resulting random partition of $\mathbb{N}$ is termed partially exchangeable. In Pitman (1995), it is proved that a $p:\mathbb{N}^*\to[0,1]$ is a *partially exchangeable probability function* if and only if it exists as a sequence of nonnegative random variables P_i, $i=1,\ldots,$ with $\sum_i P_i\le 1$ such that

$$p(n_1,\ldots,n_k)=E\left[\prod_{i=1}^{k}P_i^{n_i-1}\prod_{i=1}^{k-1}\left(1-\sum_{j=1}^{i}P_i\right)\right], \tag{8}$$

where the expectation is with respect to the distribution of the sequence (P_i). We refer to Pitman (1995) for an extensive study of partially exchangeable random partitions.

It is easily checked whether or not a given PPF satisfies the condition of Lemma 1. Corollary 1 describes all possible PPFs that have the probability of cluster memberships depend on a function of the cluster size only. This result is part of a theorem in Gnedin and Pitman (2006), but we give here a more straightforward proof.

COROLLARY 1. *Suppose a putative PPF (p_j) satisfies (7) and*

$$p_j(n_1,\ldots,n_k)\propto\begin{cases} f(n_j), & j=1,\ldots,k,\\ \theta, & j=k+1,\end{cases} \tag{9}$$

where $f(k)$ is a function from $\mathbb{N}$ to $(0,\infty)$ and $\theta>0$. Then, $f(k)=ak$ for all $k\in\mathbb{N}$ for some $a>0$.

PROOF. Note that for any $\mathbf{n} = (n_1, \ldots, n_k)$ and $i = 1, \ldots, k+1$,

$$p_i(n_1, \ldots, n_k) = \begin{cases} \dfrac{f(n_i)}{\sum_{u=1}^k f(n_u) + \theta}, & i = 1, \ldots, k, \\ \dfrac{\theta}{\sum_{u=1}^k f(n_u) + \theta}, & i = k+1. \end{cases}$$

Equation (7) with $1 \le i \ne j \le k$ implies

$$\frac{f(n_i)}{\sum_{u=1}^k f(n_u) + \theta} \frac{f(n_j)}{\sum_{u \ne i}^k f(n_u) + f(n_i + 1) + \theta} = \frac{f(n_j)}{\sum_{u=1}^k f(n_u) + \theta} \frac{f(n_i)}{\sum_{u \ne j}^k f(n_u) + f(n_j + 1) + \theta},$$

which in turn implies

$$f(n_i) + f(n_j + 1) = f(n_j) + f(n_i + 1)$$

or

$$f(n_j + 1) - f(n_j) = f(n_i + 1) - f(n_i).$$

Since this holds for all n_i and n_j, we have for all $k \in \mathbb{N}$

$$f(m) = am + b \tag{10}$$

for some $a, b \in \mathbb{R}$.

Now consider $i = k+1$ and $1 \le j \le k$. Then,

$$\frac{\theta}{\sum_{u=1}^k f(n_u) + \theta} \frac{f(n_j)}{\sum_{u=1}^k f(n_u) + f(1) + \theta} = \frac{f(n_j)}{\sum_{u=1}^k f(n_u) + \theta} \frac{\theta}{\sum_{u \ne j}^k f(n_u) + f(n_j + 1) + \theta},$$

which implies $f(n_j) + f(1) = f(n_j + 1)$ for all n_j. This together with (10) implies $b = 0$. Thus, we have $f(k) = ak$ for some $a > 0$. □

For any $a > 0$, the putative PPF

$$p_i(n_1, \ldots, n_k) \propto \begin{cases} an_i, & i = 1, \ldots, k, \\ \theta, & i = k+1, \end{cases}$$

defines a function $p : \mathbb{N} \to [0, 1]$,

$$p(n_1, \ldots, n_k) = \frac{\theta^{k-1} a^{n-k}}{[\theta + 1]_{n-1;a}} \prod_{i=1}^k (n_i - 1)!,$$

where $[\theta]_{k;a} = \theta(\theta + a) \cdots (\theta + (k-1)a)$. Since this function is symmetric in its arguments, it is an EPPF. This is the EPPF for a DP with total mass θ/a. Thus, Corollary 1 implies that the EPPF under the DP is the only EPPF that satisfies (9). The corollary shows that it is not an entirely trivial matter to come up with a putative PPF that leads to a valid EPPF. A version of Corollary 1 is also well known as Johnson's Sufficientness postulate (Good, 1965). See also the discussion in Zabell (1982).

We now give a necessary and sufficient condition for the function p defined by (5) to be an EPPF, without any constraint on the form of p_j (as were present in the earlier results). Suppose σ is a permutation of $[k]$ and $\mathbf{n} = (n_1, \ldots, n_k) \in \mathbb{N}^*$. Define $\sigma(\mathbf{n}) = \sigma(n_1, \ldots, n_k) = (n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)})$. In words, σ is a permutation of group labels and $\sigma(\mathbf{n})$ is the corresponding permutation of the group sizes $\mathbf{n}$.

THEOREM 1. *Suppose a putative PPF (p_j) satisfies (7) as well as the following condition: for all $\mathbf{n} = (n_1, \ldots, n_k) \in \mathbb{N}^*$, and permutations σ on $[k]$ and $i = 1, \ldots, k$,*

$$p_i(n_1, \ldots, n_k) = p_{\sigma^{-1}(i)}(n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)}). \tag{11}$$

Then, p defined by (6) is an EPPF. The condition is also necessary; if p is an EPPF, then (11) holds.

PROOF. Fix $\mathbf{n} = (n_1, \ldots, n_k) \in \mathbb{N}^*$ and a permutation on $[k]$, σ. We wish to show that for the function p defined by (6)

$$p(n_1, \ldots, n_k) = p(n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)}). \tag{12}$$

Let Π be the partition of $[n]$ with $\mathbf{n}(\Pi) = (n_1, \ldots, n_k)$ such that

$$\Pi([n]) = (1, 2, \ldots, k, 1, \ldots, 1, 2, \ldots, 2, \ldots, k, \ldots, k),$$

where after the first k elements $1, 2, \ldots, k$, i is repeated $n_i - 1$ times for all $i = 1, \ldots, k$. Then,

$$p(\mathbf{n}) = \prod_{i=2}^k p_i(\mathbf{1}_{(i-1)}) \times \prod_{i=k}^{n-1} p_{\Pi(i+1)}(\mathbf{n}(\Pi_i)),$$

where $\mathbf{1}_{(j)}$ is the vector of length j whose elements are all 1's.

Now consider a partition Ω of $[n]$ with $\mathbf{n}(\Omega) = (n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)})$ such that

$$\Omega([n]) = (1, 2, \ldots, k, \sigma^{-1}(1), \ldots, \sigma^{-1}(1), \sigma^{-1}(2), \ldots, \sigma^{-1}(2), \ldots, \sigma^{-1}(k), \ldots, \sigma^{-1}(k)),$$

where after the first k elements $1, 2, \ldots, k$, $\sigma^{-1}(i)$ is repeated $n_i - 1$ times for all $i = 1, \ldots, k$. Then,

$$\begin{aligned}
&p(n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)}) \\
&\quad= \prod_{i=2}^{k} p_i(\mathbf{1}_{(i-1)}) \times \prod_{i=k}^{n-1} p_{\Omega(i+1)}(\mathbf{n}(\Omega_i)) \\
&\quad= \prod_{i=2}^{k} p_i(\mathbf{1}_{(i-1)}) \times \prod_{i=k}^{n-1} p_{\sigma^{-1}(\Omega(i+1))}(\sigma(\mathbf{n}(\Omega_i))) \\
&\quad= \prod_{i=2}^{k} p_i(\mathbf{1}_{(i-1)}) \times \prod_{i=k}^{n-1} p_{\Pi(i+1)}(\mathbf{n}(\Pi_i)) \\
&\quad= p(n_1, \ldots, n_k),
\end{aligned}$$

where the second equality follows from (11). This completes the proof of the sufficient direction.

Finally, we show that every EPPF p satisfies (6) and (11). By Lemma 1, every EPPF satisfies (6). Condition (12) is true by the definition of an EPPF, which includes the condition of symmetry in its arguments. And (12) implies (11). □

Fortini, Ladelli and Regazzini (2000) prove results related to Theorem 1. They provide sufficient conditions for a system of predictive distributions $p(X_n \mid X_1, \ldots, X_{n-1})$, $n = 1, \ldots,$ of a sequence of random variables (X_i) that imply exchangeability. The relation between these conditions and Theorem 1 becomes apparent by constructing a sequence (X_i) that induces a p-distributed random partition of $\mathbb{N}$. Here, it is implicitly assumed the mapping of (X_i) to the only partition such that $i, j \in \mathbb{N}$ belongs to the same subset if and only if $X_i = X_j$.

A second more general example, which extends the predictive structure considered in Corollary 1, includes the so-called Gibbs random partitions. Within this class of models

$$p(n_1, n_2, \ldots, n_k) = V_{n,k} \prod_{i=1}^{k} W_{n_i}, \tag{13}$$

where $(V_{n,k})$ and (W_{n_i}) are sequences of positive real numbers. In this case the predictive probability of a novel species is a function of the sample size n and of the number of observed species k. See Lijoi, Mena and Prünster (2007) for related distributional results on Gibbs type models. Gnedin and Pitman (2006) obtained sufficient conditions for the sequences $(V_{n,k})$ and (W_{n_i}), which imply that p is an EPPF.

3. SSMS BEYOND THE DP

3.1 The SSM(p, v)

We know that an SSM with a nonlinear PPF, that is, p_j different from the PPF of a DP, cannot be described as a function $p_j \propto f(n_j)$ of n_j only. It must be a more complicated function $f(\mathbf{n})$. Alternatively, one could try to define an EPPF and deduce the implied PPF. But directly specifying a symmetric function $p(\mathbf{n})$ such that it complies with (4) is difficult. As a third alternative we propose to consider the weights $\mathbf{P} = \{P_h, h = 1, 2, \ldots\}$ in (3). Figure 1(a) illustrates $p(\mathbf{P})$ for a DP model. The sharp decline is typical. A few large weights account for most of the probability mass. The stick breaking construction for a DP prior with total mass θ implies $E(P_h) = \theta^{h-1}(1+\theta)^{-h}$. Such geometrically decreasing mean weights are inappropriate to describe prior information in many applications. The weights can be interpreted as asymptotic relative cluster sizes. A typical application of the DP prior is, for example, a partition of patients in a clinical study into clusters. However, if clusters correspond to disease subtypes defined by variations of some biological process, then one would rather expect a number of clusters with a priori comparable size. Many small clusters with very few patients are implausible and would also be of little clinical use. This leads us to propose the use of alternative SSMs.

Figure 1(b) shows an alternative probability model $p(\mathbf{P})$. There are many ways to define $p(\mathbf{P})$; we consider, for $h = 1, 2, \ldots,$

$$P_h \propto u_h \quad \text{or} \quad P_h = \frac{u_h}{\sum_{i=1}^{\infty} u_i},$$

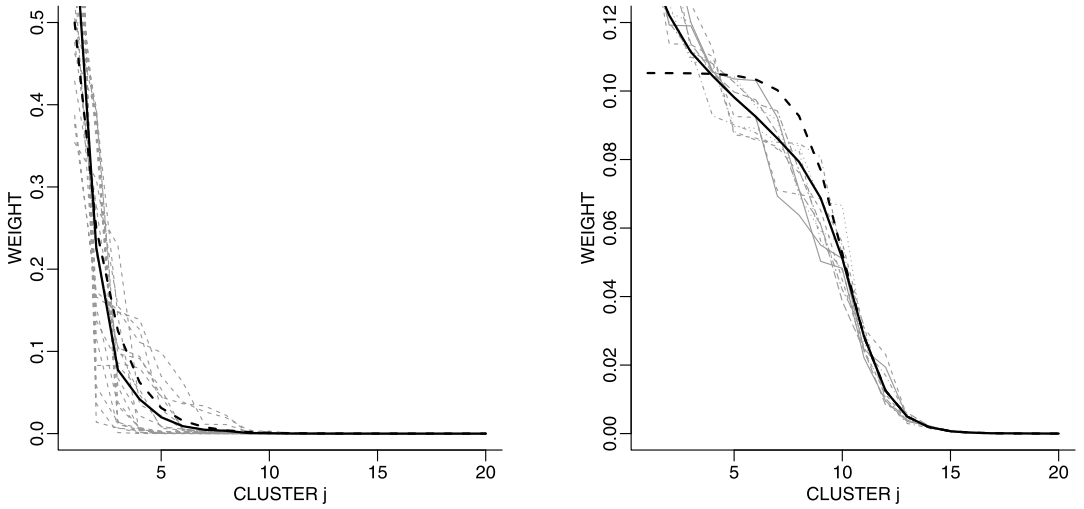
where u_h are independent and nonnegative random variables with

$$\sum_{i=1}^{\infty} u_i < \infty \quad \text{a.s.} \tag{14}$$

A sufficient condition for (14) is

$$\sum_{i=1}^{\infty} E(u_i) < \infty \tag{15}$$

by the monotone convergence theorem. Note that when the unnormalized random variables u_h are defined as the sorted atoms of a nonhomogeneous Poisson process on the positive real line, under mild assumptions, the above (P_h) construction coincides with the Poisson–Kingman models. Ferguson and Klass (1972) provide a detailed discussion on the outlined mapping



(a) DP $(M = 1, \nu)$ (b) SSM(p, ν) (note the shorter y-scale).

FIG. 1. *The lines in each panel show* 10 *draws* $\mathbf{P} \sim p(\mathbf{P})$ *for the DP (left) and for the SSM defined in* (16) *below (right). The* P_h *are defined for integers* h *only. We connect them to a line for presentation only. Also, for better presentation we plot the* sorted *weights. The thick line shows the prior mean. For comparison, a dashed thick line plots the prior mean of the* unsorted *weights. Under the DP the sorted and unsorted prior means are almost indistinguishable.*

of a Poisson process into a sequence of unnormalized positive weights. In this particular case the mean of the Poisson process has to satisfy minimal requirements (see, e.g., Pitman, 2003) to ensure that the sequence (P_i) is well defined.

As an illustrative example in the following discussion, we define, for $h = 1, 2, \ldots,$

$$P_h \propto e^{X_h} \quad \text{with } X_h \sim N\bigl(\log\bigl(1 - \{1 + e^{b-ah}\}^{-1}\bigr), \sigma^2\bigr), \tag{16}$$

where a, b, σ^2 are positive constants. The existence of such random probabilities is guaranteed by (15), which is easy to check.

The S-shaped nature of the random distribution (16), when plotted against h, distinguishes it from the DP model. The first few weights are a priori of equal size (before sorting). This is in contrast to the stochastic ordering of the DP and the Pitman–Yor process in general. In panel (a) of Figure 1 the prior mean of the sorted and unsorted weights is almost indistinguishable, because the prior already implies strong stochastic ordering of the weights.

The prior in Figure 1(b) reflects prior information of an investigator who believes that there should be around 5 to 10 clusters of comparable size in the population. This is in sharp contrast to the (often implausible) assumption of one large dominant cluster and geometrically smaller clusters that is reflected in panel (a). Prior elicitation can exploit such readily interpretable implications of the prior choice to propose models like (16).

We use SSM(p, ν) to denote a SSM defined by $p(\mathbf{P})$ for the weights P_h and $m_h \overset{\text{i.i.d.}}{\sim} \nu$. The attraction of defining the SSM through $\mathbf{P}$ is that by (3) any joint probability model $p(\mathbf{P})$ such that $P(\sum_h P_h = 1)$ defines a proper SSM. There are no additional constraints as for the PPF $p_j(\mathbf{n})$ or the EPPF $p(\mathbf{n})$. However, we still need the implied PPF to implement posterior inference and also to understand the implications of the defined process. Thus, a practical use of this second approach requires an algorithm to derive the PPF starting from an arbitrarily defined $p(\mathbf{P})$.

3.2 An Algorithm to Determine the PPF

Recall definition (3) for an SSM random probability measure. Assuming a proper SSM, we have

$$G = \sum_{h=1}^{\infty} P_h \delta_{m_h}. \tag{17}$$

Let $\mathbf{P} = (P_h, h \in \mathbb{N})$ denote the sequence of weights. Recall the notation $\tilde{X}_j$ for the jth unique value in the SSS $\{X_i, i = 1, \ldots, n\}$. The algorithm requires indicators that match the $\tilde{X}_j$ with the m_h, that is, that match the clusters in the partition with the point masses of the SSM. Let $\pi_j = h$ if $\tilde{X}_j = m_h$, $j = 1, \ldots, k_n$. In the

following discussion it is important that the latent indicators π_j are only introduced up to $j = k$. Conditional on m_h, $h \in \mathbb{N}$ and $\tilde{X}_j$, $j \in \mathbb{N}$, the indicators π_j are deterministic. After marginalizing with respect to the m_h or with respect to the $\tilde{X}_j$, the indicators become latent variables. Also, we use cluster membership indicators $s_i = j$ for $X_i = \tilde{X}_j$ to simplify notation. We use the convention of labeling clusters in the order of appearance, that is, $s_1 = 1$ and $s_{i+1} \in \{1, \ldots, k_i, k_i + 1\}$.

In words, the algorithm proceeds as follows. We write the desired PPF $p_j(\mathbf{n})$ as an expectation of the conditional probabilities $p(X_{n+1} = \tilde{X}_j \mid \mathbf{n}, \pi, \mathbf{P})$. The expectation is with respect to $p(\mathbf{P}, \pi \mid \mathbf{n})$. Next we approximate the integral with respect to $p(\mathbf{P}, \pi \mid \mathbf{n})$ by a weighted Monte Carlo average over samples $(\mathbf{P}^{(\ell)}, \pi^{(\ell)}) \sim p(\mathbf{P}^{(\ell)}) p(\pi^{(\ell)} \mid \mathbf{P}^{(\ell)})$ from the prior. Note π and $\mathbf{P}$ together define the size-biased permutation of (P_j),

$$\tilde{P}_j = P_{\pi_j}, \quad j = 1, 2, \ldots.$$

The size-biased permutation $(\tilde{P}_j)$ of (P_j) is a resampled version of (P_j) where sampling is done with probability proportional to P_j and without replacement. Once the sequence (P_j) is simulated, it is computationally straightforward to get $(\tilde{P}_j)$. Note also that the properties of the random partition can be characterized by the distribution on $\mathbf{P}$ only. The point masses m_h are not required.

Using the cluster membership indicators s_i and the size-biased probabilities $\tilde{P}_j$, we write the desired PPF as

$$\begin{aligned} p_j(\mathbf{n}) &= p(s_{n+1} = j \mid \mathbf{n}) \\ &= \int p(s_{n+1} = j \mid \mathbf{n}, \tilde{\mathbf{P}}) p(\tilde{\mathbf{P}} \mid \mathbf{n}) \, \mathrm{d}\tilde{\mathbf{P}} \\ &\propto \int p(s_{n+1} = j \mid \mathbf{n}, \tilde{\mathbf{P}}) p(\mathbf{n} \mid \tilde{\mathbf{P}}) p(\tilde{\mathbf{P}}) \, \mathrm{d}\tilde{\mathbf{P}} \\ &\approx \frac{1}{L} \sum p\big(s_{n+1} = j \mid \mathbf{n}, \tilde{\mathbf{P}}^{(\ell)}\big) p\big(\mathbf{n} \mid \tilde{\mathbf{P}}^{(\ell)}\big). \end{aligned} \tag{18}$$

The Monte Carlo sample $\tilde{\mathbf{P}}^{(\ell)}$ or, equivalently, $(\mathbf{P}^{(\ell)}, \pi^{(\ell)})$, is obtained by first generating $\mathbf{P}^{(\ell)} \sim p(\mathbf{P})$ and then $p(\pi_j^{(\ell)} = h \mid \mathbf{P}^{(\ell)}, \pi_1^{(\ell)}, \ldots, \pi_{j-1}^{(\ell)}) \propto P_h^{(\ell)}$, $h \notin \{\pi_1^{(\ell)}, \ldots, \pi_{j-1}^{(\ell)}\}$. In actual implementation the elements of $\mathbf{P}^{(\ell)}$ and $\pi^{(\ell)}$ are only generated as and when needed.

The terms in the last line of (18) are easily evaluated. The first factor is given as predictive cluster membership probabilities

$$\begin{aligned} &p(s_{n+1} = j \mid \mathbf{n}, \tilde{\mathbf{P}}) \\ &\quad = \begin{cases} \tilde{P}_j, & j = 1, \ldots, k_n, \\ \left(1 - \sum_{j=1}^{k_n} \tilde{P}_j\right), & j = k_n + 1. \end{cases} \end{aligned} \tag{19}$$

The second factor is evaluated as

$$p(\mathbf{n} \mid \tilde{\mathbf{P}}) = \prod_{j=1}^{k} \tilde{P}_j^{n_j - 1} \prod_{j=1}^{k-1} \left(1 - \sum_{i=1}^{j} \tilde{P}_i\right).$$

Note that the second factor coincides with the previously mentioned [cf. expression (8)] Pitman's representation result for partially exchangeable partitions.

Figure 2 shows an example. The figure plots $p(s_{i+1} = j \mid \mathbf{s})$ against cluster size n_j. In contrast, the DP Polya urn would imply a straight line. The plotted probabilities are averaged with respect to all other features of $\mathbf{s}$, in particular, the multiplicity of cluster sizes, etc. The figure also shows probabilities (19) for specific simulations.

3.3 A Simulation Example

Many data analysis applications of the DP prior are based on DP mixtures of normals as models for a random probability measure F. Applications include density estimation, random effects distributions, generalizations of a probit link, etc. We consider a stylized example that is chosen to mimic typical features of such models.

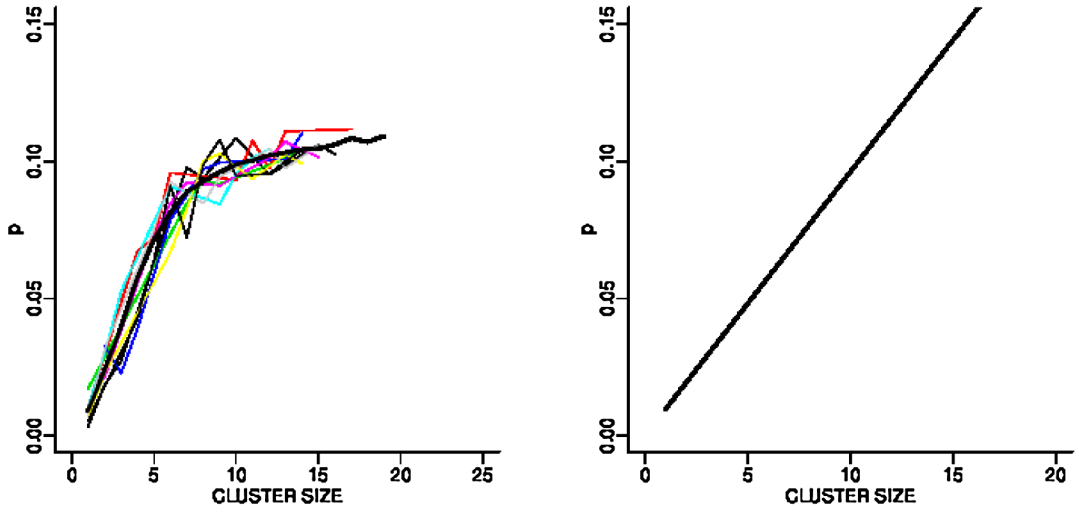
In this section we show posterior inference conditional on the data set $(y_1, y_2, \ldots, y_9) = (-4, -3, -2, \ldots, 4)$. The use of these data highlights the differences in posterior inference between the SSM and DP priors. Assume $y_i \overset{\text{i.i.d.}}{\sim} F$, with a semi-parametric mixture of normal prior on F,

$$y_i \overset{\text{i.i.d.}}{\sim} F \quad \text{with } F(y_i) = \int N\big(y_i; \mu, \sigma^2\big) \, \mathrm{d}G\big(\mu, \sigma^2\big).$$

Here $N(x; m, s^2)$ denotes a normal distribution with moments (m, s^2) for the random variable x. We estimate F under two alternative priors,

$$G \sim \mathrm{SSM}(p, \nu) \quad \text{or} \quad G \sim \mathrm{DP}(M, \nu).$$

The distribution p of the weights for the $\mathrm{SSM}(p, \cdot)$ prior is defined as in (16). The total mass parameter M



(a) SSM$(p,\cdot)$ (b) DP$(M,\cdot)$

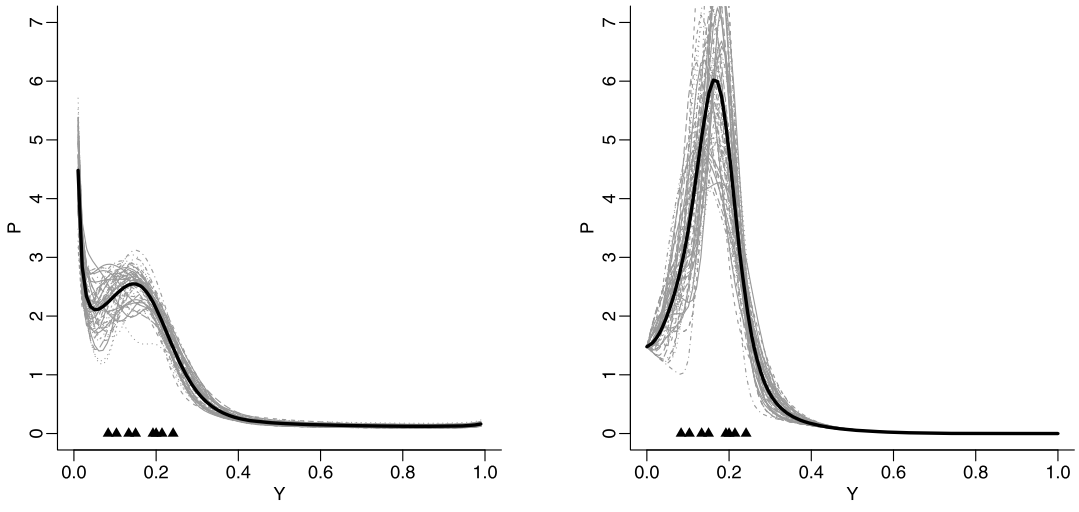
FIG. 2. *Panel* (a) *shows the PPF* (19) *for a random probability measure* $G \sim \mathrm{SSM}(p,v)$, *with* P_h *as in* (16). *The thick line plots* $p(s_{n+1}=j \mid \mathbf{s})$ *against* n_j, *averaging over multiple simulations. In each simulation we used the same simulation truth to generate* $\mathbf{s}$ *and stop simulation at* $n=100$. *The* 10 *thin lines show* $p_j(\mathbf{n})$ *for* 10 *simulations with different* $\mathbf{n}$. *In contrast, under the DP Polya urn the curve is a straight line and there is no variation across simulations* [*panel* (b)].

in the DP prior is fixed to match the prior mean number of clusters, $E(k_n)$, implied by (16). We find $M=2.83$. Let $\mathrm{Ga}(x;a,b)$ indicate that the random variable x has a Gamma distribution with shape parameter a and inverse scale parameter b. For both prior models we use

$$v(\mu,1/\sigma^2)=N(x;\mu_0,c\sigma^2)\,\mathrm{Ga}(1/\sigma^2;a/2,b/2).$$

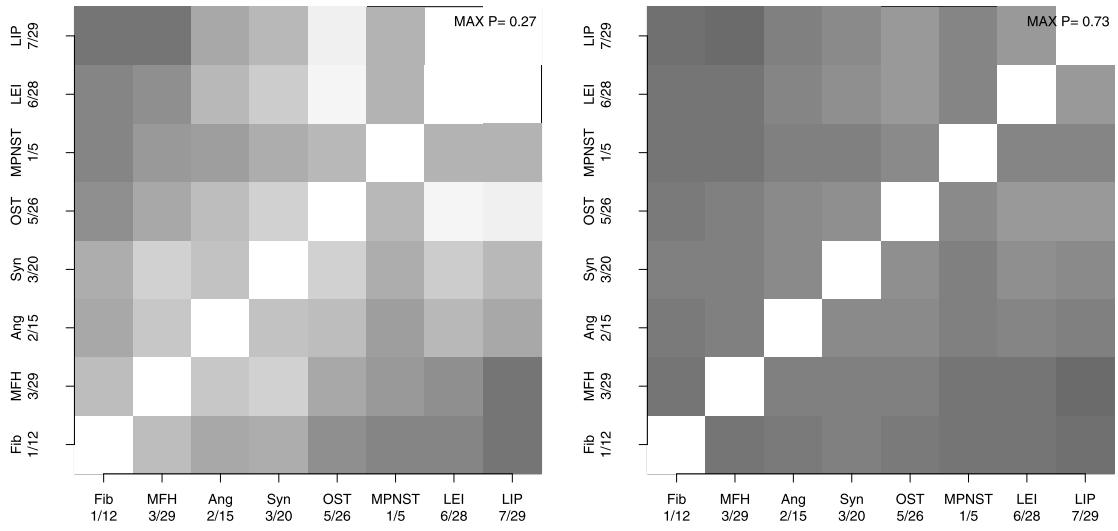
We fix $\mu_0=0$, $c=10$ and $a=b=4$. The model can alternatively be written as $y_i \sim N(\mu_i,\sigma_i^2)$ and $X_i=(\mu_i,1/\sigma_i^2)\sim G$.

Figures 3 and 4 show some inference summaries. Inference is based on Markov chain Monte Carlo (MCMC) posterior simulation with 1000 iterations. Posterior simulation is for $(s_1,\ldots,s_n)$ only. The



(a) $G \sim \mathrm{SSM}(p,\nu)$ (b) $G \sim \mathrm{DP}(M,\nu)$

FIG. 3. *Posterior estimated sampling model* $\overline{F}=E(F \mid \text{data})=p(y_{n+1} \mid \text{data})$ *under the* SSM(p,v) *prior and a comparable DP prior. The triangles along the x-axis show the data.*



(a) $G \sim \text{SSM}(p, \nu)$ (b) $G \sim \text{DP}(M, \nu)$

FIG. 4. *Co-clustering probabilities* $p(s_i = s_j \mid \text{data})$ *under the two prior models.*

cluster-specific parameters $(\tilde{\mu}_j, \tilde{\sigma}_j^2)$, $j = 1, \ldots, k_n$, are analytically marginalized. One of the transition probabilities (Gibbs sampler) in the MCMC requires the PPF under SSM(p, ν). It is evaluated using (18).

Figure 3 shows the posterior estimated sampling distributions F. The figure highlights a limitation of the DP prior. The single total mass parameter M controls both, the number of clusters and the prior precision. A small value for M favors a small number of clusters and implies low prior uncertainty. Large M implies the opposite. Also, we already illustrated in Figure 1 that the DP prior implies stochastically ordered cluster sizes, whereas the chosen SSM prior allows for many approximately equal size clusters. The equally spaced grid data $(y_1, \ldots, y_n)$ implies a likelihood that favors a moderate number of approximately equal size clusters. The posterior distribution on the random partition is shown in Figure 4. Under the SSM prior the posterior supports a moderate number of similar size clusters. In contrast, the DP prior shrinks the posterior toward a few dominant clusters. Let $n_{(1)} \equiv \max_{j=1,\ldots,k_n} n_j$ denote the leading cluster size. Related evidence can be seen in the marginal posterior distribution (not shown) of k_n and $n_{(1)}$. We find $E(k_n \mid \text{data}) = 6.4$ under the SSM model versus $E(k_n \mid \text{data}) = 5.1$ under the DP prior. The marginal posterior modes are $k_n = 6$ under the SSM prior and $k_n = 5$ under the DP prior. The marginal posterior modes for $n_{(1)}$ is $n_{(1)} = 2$ under the SSM prior and $n_{(1)} = 3$ under the DP prior.

3.4 Analysis of Sarcoma Data

We analyze data from of a small phase II clinical trial for sarcoma patients that was carried out in the M. D. Anderson Cancer Center. The study was designed to assess efficacy of a treatment for sarcoma patients across different subtypes. We consider the data accrued for 8 disease subtypes that were classified as having overall intermediate prognosis, as presented in Table 1. Each table entry indicates the total number of patients for each sarcoma subtype and the number of patients who reported a treatment success. See further discussion in León-Novelo et al. (2012).

One limitation of these data is the small sample size, which prevents separate analysis for each disease subtype. On the other hand, it is not clear that we should simply treat the subtypes as exchangeable. We deal with these issues by modeling each table entry as a binomial response and adopt a hierarchical framework for the success probabilities. The hierarchical model includes a random partition of the subtypes. Condi-

TABLE 1

Sarcoma data. For each disease subtype (top row) we report the total number of patients and the number of treatment successes. See León-Novelo et al. (2012) for a discussion of disease subtypes

Sarcoma	LEI	LIP	MFH	OST	Syn	Ang	MPNST	Fib
	6/28	7/29	3/29	5/26	3/20	2/15	1/5	1/12

tional on a given partition, data across all subtypes in the same cluster are pooled, thus allowing more precise inference on the common success probabilities for all subtypes in this cluster. We consider two alternative models for the random partition, based on a DP(M, ν) prior versus a SSM(p, ν) prior. Specifically, we consider the following models:

$$y_i|\pi_i \sim \text{Bin}(n_i, \pi_i),$$
$$\pi_i|G \sim G,$$
$$G \sim \text{DP}(M, \nu) \quad \text{or} \quad \text{SSM}(p, \nu),$$

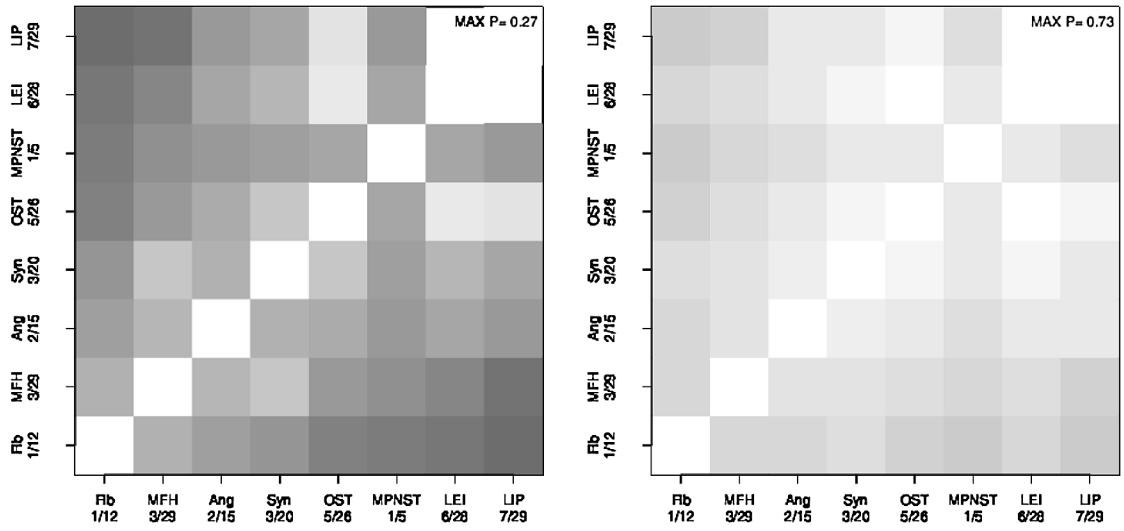
where ν is a diffuse probability measure on $[0, 1]$ and p is again defined as in (16).

The hierarchical structure of the data and the aim of clustering subpopulations in order to achieve borrowing of strength is in continuity with a number of applied contributions. Several of these, for instance, are meta analyses of medical studies (Berry and Christensen, 1979), with subpopulations defined by medical institutions or by clinical trials. In most cases the application of the DP is chosen for computational advantages and (in some cases) due to the easy implementation of strategies for prior specification (Liu, 1996). With a small number of studies, as in our example, ad hoc construction of alternative SSM combines hierarchical modeling with advantageous posterior clustering. The main advantage is the possibility of avoiding the exponential decrease typical of the ordered DP atoms.

In this particular analysis, we used $M = 2.83$ and chose ν to be the Beta$(0.15, 0.85)$ distribution, which was designed to match the prior mean of the observed data and has prior equivalent sample size of 1. The total mass $M = 2.83$ for the DP prior was selected to achieve matching prior expected number of clusters under the two models. The DP prior on G favors the formation of large clusters (with matched prior mean number of clusters) which leads to less posterior shrinkage of cluster-specific means. In contrast, under the SSM prior the posterior puts more weight on several smaller clusters.

Figure 5 shows the estimated posterior probabilities of pairwise co-clustering for model (16) in the left panel and for the DP case (right panel). Clearly, compared to the DP model, the chosen SSM induces a posterior distribution with more clusters, as reflected in the lower posterior probabilities $p(s_i = s_j \mid y)$ for all i, j.

Figure 6 shows the posterior distribution of the number of clusters under the SSM and DP mixture models. Under the DP (right panel) includes high probability for a single cluster, $k = 1$, with $n_1 = 8$. The high posterior probability for few large clusters also implies high posterior probabilities $\widehat{p}_{ij}$ of co-clustering. Under the SSM (left panel) the posterior distribution on ρ retains substantial uncertainty. Finally, the same pattern is confirmed in the posterior distribution of sizes of the largest cluster, $p(n_1 \mid y)$, shown in Figure 7. The high posterior probability for a single large cluster of



$p(s_i = s_j \mid y)$ SSM DP

FIG. 5. *Posterior probabilities of pairwise co-clustering, $p_{ij} = p(s_i = s_j \mid y)$. The grey scales in the two panels are scaled as black for $p_{ij} = 0$ to white for $p_{ij} = \max_{r,s} p_{rs}$. The maxima are indicated in the right top of the plots.*

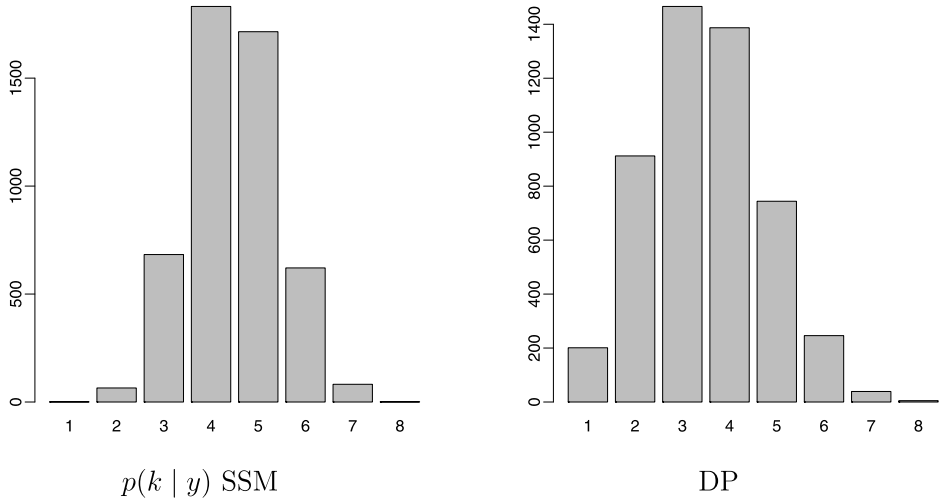


FIG. 6. *Posterior distribution on the number of clusters.*

all $n = 8$ sarcoma subtypes seems unreasonable for the given data.

4. DISCUSSION

We have reviewed alternative definitions of SSMs. We also reviewed the fact that all SSMs with a PPF of the form $p_j(\mathbf{n}) = f(n_j)$ must necessarily be a linear function of n_j and provided a new elementary proof. In other words, the PPF $p_j(\mathbf{n})$ depends on the current data only through the cluster sizes. The number of clusters and any other aspect of the partition Π_n do not change the prediction. This is an excessively simplifying assumption for most data analysis problems.

We provide an alternative class of models that allows for more general PPFs. These models are obtained by directly specifying the distribution of unnormalized weights u_h. The proposed approach for defining SSMs allows the incorporation of the desired qualitative properties concerning the decrease of the ordered clusters cardinalities. This flexibility comes at the cost of additional computation required to implement the algorithm described in Section 3.2, compared to the standard approaches under DP-based models. Nevertheless, the benefits obtained in the case of data sets that require more flexible models compensate the increase in computational effort. A different strategy for constructing discrete random distributions has been dis-

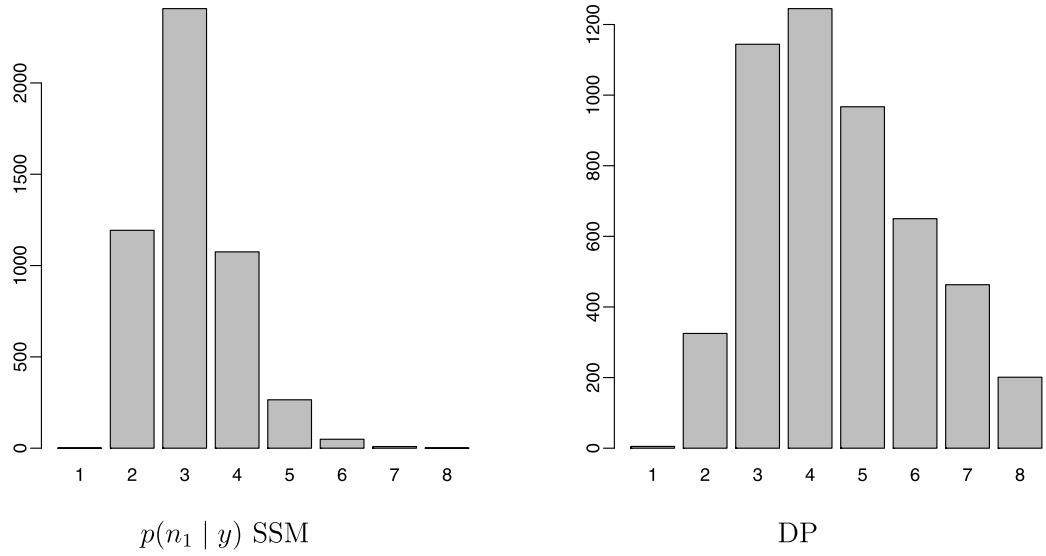


FIG. 7. *Posterior distribution on the size of the largest cluster.*

cussed in Trippa and Favaro (2012). In several applications, the scope for which SSMs are to be used suggests these *desired qualitative properties*. Nonetheless, we see the definition of a theoretical framework supporting the selection of a SSM as an open problem.

R code for an implementation of posterior inference under the proposed new model is available at http://math.utexas.edu/users/pmueller/.

ACKNOWLEDGMENTS

Jaeyong Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030811). Fernando Quintana was supported by Grant FONDECYT 1100010. Peter Müller was partially funded by Grant NIH/NCI CA075981.

REFERENCES

BERRY, D. A. and CHRISTENSEN, R. (1979). Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes. *Ann. Statist.* **7** 558–568. MR0527491

BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614

BRIX, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.* **31** 929–953. MR1747450

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

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

FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43** 1634–1643. MR0373022

FORTINI, S., LADELLI, L. and REGAZZINI, E. (2000). Exchangeability, predictive distributions and parametric models. *Sankhyā Ser. A* **62** 86–109. MR1769738

GNEDIN, A., HAULK, C. and PITMAN, J. (2010). Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In *Probability and Mathematical Genetics. London Mathematical Society Lecture Note Series* **378** 264–298. Cambridge Univ. Press, Cambridge. MR2744243

GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **325** 83–102, 244–245. MR2160320

GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* **138** 5674–5685.

GOOD, I. J. (1965). *The Estimation of Probabilities. An Essay on Modern Bayesian Methods. Research Monograph* **30**. MIT Press, Cambridge, MA. MR0185724

ISHWARAN, H. and JAMES, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica* **13** 1211–1235. MR2026070

JAMES, L. F. (2008). Large sample asymptotics for the two-parameter Poisson–Dirichlet process. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. Collect.* **3** 187–199. IMS, Beachwood, OH. MR2459225

JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. MR2508332

JANG, G. H., LEE, J. and LEE, S. (2010). Posterior consistency of species sampling priors. *Statist. Sinica* **20** 581–593. MR2682631

KINGMAN, J. F. C. (1975). Random discrete distribution (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **37** 1–22. MR0368264

KINGMAN, J. F. C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* (2) **18** 374–380. MR0509954

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034

LEÓN-NOVELO, L., BEKELE, B. N., MÜLLER, P. and QUINTANA, F. A. (2012). Borrowing strength with non-exchangeable priors over subpopulations. *Biometrics* **68** 550–558. MR2959622

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291. MR2236441

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. MR2416792

LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *J. Amer. Statist. Assoc.* **100** 1292–1296. MR2236442

LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008a). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica* **18** 1653–1668. MR2469329

LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008b). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18** 1519–1547. MR2434179

LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds.) 80–136. Cambridge Univ. Press, Cambridge. MR2730661

LIU, J. S. (1996). Nonparametric hierarchical Bayes via sequential imputations. *Ann. Statist.* **24** 911–930. MR1401830

MACEACHERN, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.* **23** 727–741. MR1293996

MACEACHERN, S. N. and MÜLLER, P. (1998). Estimating mixtures of Dirichlet process models. *J. Comput. Graph. Statist.* **7** 223–239.

NAVARRETE, C., QUINTANA, F. A. and MÜLLER, P. (2008). Some issues in nonparametric Bayesian modelling using species sampling models. *Stat. Model.* **8** 3–21. MR2750628

NIETO-BARAJAS, L. E., PRÜNSTER, I. and WALKER, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.* **32** 2343–2360. MR2153987

PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249

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. IMS, Hayward, CA. MR1481784

PITMAN, J. (2003). Poisson–Kingman partitions. In *Statistics and Science*: *A Festschrift for Terry Speed*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. MR2004330

PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math*. **1875**. Springer, Berlin.

REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann*. *Statist*. **31** 560–585. MR1983542

TRIPPA, L. and FAVARO, S. (2012). A class of normalized random measures with an exact predictive sampling scheme. *Scand*. *J*. *Stat*. **39** 444–460.

ZABELL, S. L. (1982). W. E. Johnson's "sufficientness" postulate. *Ann*. *Statist*. **10** 1090–1099 (1 plate). MR0673645