Statistica Sinica **13**(2003), 1211-1235

GENERALIZED WEIGHTED CHINESE RESTAURANT PROCESSES FOR SPECIES SAMPLING MIXTURE MODELS

Hemant Ishwaran and Lancelot F. James

Cleveland Clinic Foundation and
Hong Kong University of Science and Technology

Abstract: The class of species sampling mixture models is introduced as an extension of semiparametric models based on the Dirichlet process to models based on the general class of species sampling priors, or equivalently the class of all exchangeable urn distributions. Using Fubini calculus in conjunction with Pitman (1995, 1996), we derive characterizations of the posterior distribution in terms of a posterior partition distribution that extend the results of Lo (1984) for the Dirichlet process. These results provide a better understanding of models and have both theoretical and practical applications. To facilitate the use of our models we generalize the work in Brunner, Chan, James and Lo (2001) by extending their weighted Chinese restaurant (WCR) Monte Carlo procedure, an i.i.d. sequential importance sampling (SIS) procedure for approximating posterior mean functionals based on the Dirichlet process, to the case of approximation of mean functionals and additionally their posterior laws in species sampling mixture models. We also discuss collapsed Gibbs sampling, Pólya urn Gibbs sampling and a Pólya urn SIS scheme. Our framework allows for numerous applications, including multiplicative counting process models subject to weighted gamma processes, as well as nonparametric and semiparametric hierarchical models based on the Dirichlet process, its two-parameter extension, the Pitman-Yor process and finite dimensional Dirichlet priors.

Key words and phrases: Dirichlet process, exchangeable partition, finite dimensional Dirichlet prior, two-parameter Poisson-Dirichlet process, prediction rule, random probability measure, species sampling sequence.

1. Introduction

This paper introduces the class of random mixture models, referred to as *species sampling mixture models*. These are defined as

$$\int_{\mathcal{Y}} K(x|y,\theta)\, P(dy), \tag{1}$$

where $K(x|y,\theta)$ is a known positive integrable kernel, θ is a finite dimensional parameter with parameter space Θ (often $\Theta = \Re^d$) and the measure P is drawn from a random probability measure $\mathcal{P}$ with a general exchangeable urn distribution. When $\mathcal{P}$ is the Ferguson (1973, 1974) Dirichlet process, (1) was proposed

in Lo (1984) as a method for Bayesian density estimation, or more generally as a mixture model device for inference in nonparametric problems (now often called a Dirichlet process mixture model). For example, in the context of density estimation, if $X_1, \ldots, X_n$ are observations from an unknown density f, then choosing K to be a density function in (1) yields a random kernel density

$$f(x|\theta, P) = \int_{\mathcal{Y}} K(x|y, \theta)\, P(dy),$$

where P has a Dirichlet process law. Assuming for the moment that θ is fixed, then f is estimated by the Bayesian predictive density $\widehat{f}(x) = \int\int_{\mathcal{Y}} K(x|y,\theta)P(dy)$ $\mathcal{P}(dP|X_1, \ldots, X_n)$, where $\mathcal{P}(dP|X_1, \ldots, X_n)$ is the posterior for P given the data. A typical choice for $K(\cdot|y, \theta)$ is the normal density with mean $y \in \Re$ and variance $\theta^2 > 0$. In this case, this is the Bayesian nonparametric analogue of kernel density estimation and θ is the bandwith value.

Lo (1984) provided informative characterizations for the posterior $\mathcal{P}(dP|X_1, \ldots, X_n)$ for Dirichlet process mixture models, thus providing a template for their practical as well as theoretical development. In this paper we extend this framework by looking at the larger class of species sampling mixture models which (a) allow for kernels that are integrable (and therefore not necessarily densities) and (b) which include all $\mathcal{P}$ from the class of *species sampling models* developed in Pitman (1995, 1996). This rich class corresponds to the set of all random probability measures of the form

$$\mathcal{P}(\cdot) = \sum_k p_k\, \delta_{Z_k}(\cdot) + \left(1 - \sum_k p_k\right) H(\cdot), \tag{2}$$

where $0 < p_k < 1$ are random weights such that $\sum_k p_k \leq 1$, independently of Z_k, which are i.i.d. with some *nonatomic* distribution H over a measurable Polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ (note: $\delta_{Z_k}(\cdot)$ denotes a discrete measure concentrated at Z_k). This class includes the Dirichlet process, the two parameter extension of the Dirichlet process, the two-parameter Poisson-Dirichlet process (Pitman and Yor (1997)), which we refer to as the Pitman-Yor process, and the class of finite dimensional Dirichlet priors discussed in detail in Ishwaran and Zarepour (2002a,b). Another important and rich sub-class of models consist of processes based on normalised Poisson or weighted Poisson laws with homogeneous Lévy measures. These were discussed recently in James (2002). A unifying feature of species sampling models, which form the basis of our exposition, is the intimate connection to a corresponding exchangeable partition distribution on the positive integers $\mathbb{N} = \{1, 2, \ldots\}$. These exchangeable partition distributions are related to generalizations of the Chinese restaurant process (Aldous (1995), Pitman (1996)) and are of primary interest in many applications. We show how

to use this relationship in conjunction with Fubini calculus for random measures to develop a rigorous and informative characterization of the posterior distribution in terms of a posterior partition distribution (see Theorems 1 and 2 of Section 3). The results are a generalization of the results given in Lo (1984) for the special case of the Dirichlet process. This provides a better understanding of the common structural features of the models, and has implications based on both theoretical and practical considerations.

To facilitate the use of the species sampling mixture model, we extend the work in Brunner, Chan, James and Lo (2001) (see also Lo, Brunner and Chan (1996)) on an i.i.d. sequential importance sampling (SIS) computational procedure for fitting Dirichlet process mixture models, a technique based on what they call a *weighted Chinese restaurant process* (WCR process). By exploiting the form of the posterior under a Dirichlet process, Brunner, Chan, James and Lo (2001) (hereafter abbreviated as BCJL (2001)) devised a Monte Carlo method, the iidWCR, for drawing i.i.d. values for estimating posterior mean functionals (see also MacEachern, Clyde and Liu (1999) who discuss an SIS algorithm for fitting beta-binomial Dirichlet process mixture models which is operationally equivalent to the iidWCR; also see Quintana (1998) and Quintana and Newton (2000)). Although the iidWCR algorithm was specifically developed for models involving Dirichlet process and weighted gamma process priors (Lo and Weng (1989)), in this paper we show that the method can be extended in generality to all random probability measures $\mathcal{P}$ with the property that a sample $Y_1, Y_2, \ldots$ drawn from $\mathcal{P}$ is a *species sampling sequence* (Pitman (1996)). We call this algorithm the generalized weighted Chinese restaurant algorithm (GWCR algorithm), a technique based on what we call the GWCR process. Complementary to the GWCR algorithm is an extension of the collapsed Gibbs sampler of MacEachern (1994). Both the GWCR and collapsed Gibbs methods are characterized as partition-based approaches. The use of partitions naturally leads to reduced Monte Carlo error as will be discussed in Section 4. Additionally, we develop Gibbs procedures like Escobar (1988, 1994) and SIS procedures analogous to Kong, Liu and Wong (1994) and Liu (1996) based on sampling Y_i. Section 4 presents our algorithms. Extensions are discussed in Section 5.

2. Background and Notation

A sequence of random variables $Y_1, Y_2, \ldots$ is called a species sampling sequence (Fisher, Corbet and Williams (1943), Pitman (1996)) if for each $n \geq 1$, $Y_1, \ldots, Y_{n+1}$ is an exchangeable sequence from an urn model whose joint marginal law $\mu(dY_1, \ldots, dY_{n+1})$ is determined sequentially by the following *prediction rule*:

$$\mathbb{P}\{Y_1 \in \cdot\} = H(\cdot), \qquad \mathbb{P}\{Y_{n+1} \in \cdot \,|Y_1, \ldots, Y_n\} = \ell_{0,n} H(\cdot) + \sum_{j=1}^{k(n)} \ell_{j,n}\, \delta_{Y_j^*}(\cdot), \quad (3)$$

where $Y_1^*, \ldots, Y_{k(n)}^*$ are the $k(n)$ distinct values of $Y_1, \ldots, Y_n$ in the *order as they appear*, and $\ell_{0,n}$ and $\ell_{j,n}$ are non-negative measurable functions of $Y_1, \ldots, Y_n$. As before H is some nonatomic probability measure. A random measure $\mathcal{P}$ is called a species sampling model if a sample $Y_1, Y_2, \ldots$ drawn from $\mathcal{P}$ forms a species sampling sequence. See Theorem 3 of Hansen and Pitman (2000) for necessary conditions on $\ell_{0,n}$ and $\ell_{j,n}$ to ensure that $Y_1, Y_2, \ldots$ are exchangeable. See also Kingman (1978), Hoppe (1987), Donnelly and Tavaré (1987) and Pitman (1996) for related discussion.

By Proposition 11 of Pitman (1996), the exchangeability of a species sampling sequence ensures that its unique values $Y_1^*, Y_2^*, \ldots$ form an i.i.d. sequence with distribution H and that the partition $\Pi = \{A_1, A_2, \ldots\}$ induced by the unique values, where $A_j = \{i : Y_i = Y_j^*\}$, is an exchangeable random partition over the set of positive integers $\mathbb{N}$. The proposition follows from Kingman's (1978) theory of partition structures as developed in Aldous (1985) and Pitman (1995). In particular, this implies that, for each n, the distribution of $Y_1, \ldots, Y_n$ is determined from its unique values and its corresponding exchangeable partition. More precisely, let $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ denote the partition of the integers $\{1, \ldots, n\}$, where $C_j = \{i : Y_i = Y_j^*\}$ are the partition sets induced by the $n(\mathbf{p})$ unique values $Y_1^*, \ldots, Y_{n(\mathbf{p})}^*$. The joint marginal law for $Y_1, \ldots, Y_n$ has the unique characterization (Pitman (1996))

$$\mathbb{P}\{\mathbf{p}; Y_j^* \in B_j, j = 1, \ldots, n(\mathbf{p})\} = \pi(\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} H(B_j), \tag{4}$$

where $\pi(\mathbf{p})$ is the *exchangeable partition probability function*, or EPPF, which can be written as $\pi(\mathbf{p}) = p(e_1, \ldots, e_{n(\mathbf{p})})$, where p is a unique symmetric function depending only upon e_j, the cardinalites for the sets C_j. The decomposition (4) contains two key facts which will form the basis of our results of Section 3. First it states implicitly that the unique values Y_j^* are i.i.d. with common distribution H. Secondly, it isolates the EPPF as the unique contribution from a species sampling model. In the species sampling mixture model (1), the EPPF $\pi(\mathbf{p})$, is the prior for $\mathbf{p}$, and thus (4) suggests that the effect on the posterior by the choice of the species sampling model must be governed by the EPPF. This is in fact the basis for a general relationship between the prior and the posterior that will be revealed in the posterior characterizations of Section 3.

2.1. Generalized Chinese restaurant processes

A key aspect of species sampling models that will be exploited in our computational algorithms, is that observations $\mathbf{p}$ from the EPPF may be generated via a generalized Chinese restaurant process. This process can be viewed as a

sequential restaurant "seating arrangement" described as follows. Suppose customers arrive sequentially at a Chinese restaurant and are randomly assigned to an unlimited number of circular tables $C_1, C_2, \ldots$, each of which have an unlimited capacity to seat customers. By default the first customer to arrive is always seated at the first table, table C_1. Subsequently, for $r \geq 1$, customer $r+1$ is seated according to the species sampling prediction rule applied to the partition $\mathbf{p}_r = \{C_{1,r}, \ldots, C_{n(\mathbf{p}_r),r}\}$ of $\{1, \ldots, r\}$ corresponding to the seating arrangement of the first r customers. That is, if $\mathbf{p}_{r+1}$ is the event that customer $r+1$ is seated at a previous table $C_{j,r}$, (denoted as $\mathbf{p}_{r+1} = \mathbf{p}_r \cup \{r+1 \in C_{j,r}\}$), then $\mathbf{p}_{r+1}$ occurs with probability

$$\pi(\mathbf{p}_{r+1}|\mathbf{p}_r) = \frac{\pi(\mathbf{p}_{r+1})}{\pi(\mathbf{p}_r)} = \ell_{j,r}, \tag{5}$$

while if $\mathbf{p}_{r+1} = \mathbf{p}_r \cup C_{n(\mathbf{p}_r)+1}$ is the event that customer $r+1$ is seated at a new table, then the event $\mathbf{p}_{r+1}$ occurs with probability

$$\pi(\mathbf{p}_{r+1}|\mathbf{p}_r) = \frac{\pi(\mathbf{p}_{r+1})}{\pi(\mathbf{p}_r)} = \ell_{0,r}. \tag{6}$$

After n steps this results in a partition $\mathbf{p} = \mathbf{p}_n$ from the EPPF, $\pi(\mathbf{p})$. The original Chinese restaurant process was devised by Lester Dubins and Jim Pitman (see Aldous (1985) and Pitman (1996)). This scheme corresponds to the choice of $\ell_{0,r} = \alpha/(\alpha+r)$ and $\ell_{j,r} = e_{j,r}/(\alpha+r)$, where $e_{j,r}$ is the size of table $C_{j,r}$. This process generates a partition based on the Ewens sampling formula (Ewens (1972)) and its EPPF is that for a Dirichlet process with shape measure $\alpha H(\cdot)$, which we will write as $\mathcal{P} = \mathrm{DP}(\alpha H)$. See also Blackwell and MacQueen (1973) for further discussion on the Dirichlet process prediction rule (often called the Blackwell-MacQueen Pólya urn). Subsequent extension to more general species sampling models may be found in Pitman (1996), Kerov (1998) and Tsilevich (1997).

Remark 1. Observe carefully that we only need to know the prediction rule for a species sampling model to draw a value for $\mathbf{p}$. This fact makes it *unnecessary to work out the explicit form* for the EPPF. This will be of practical consequence to implementing our GWCR procedures described in Section 4, because although the EPPF may be obtained uniquely via its prediction rule, we are informed by Jim Pitman (personal communication) that explicit expressions for these quantities do not have a simple form in general.

2.2. The two-parameter model

An important class of species sampling models are the two-parameter models discussed in Pitman (1996). These are defined via (3) with $\ell_{0,n} = (b + n(\mathbf{p})a)/(n+$

b) and $\ell_{j,n} = (e_{j,n} - a)/(n + b)$, where a and b are two real valued parameters. These models are well defined over two ranges of the parameters: $0 \le a < 1$ and $b > 0$, or

$$a = -\kappa < 0 \quad \text{and} \quad b = N\kappa, \text{ for } \kappa > 0 \text{ and } N = 2, 3 \ldots. \tag{7}$$

These models have EPPF

$$\pi(\mathbf{p}) = \frac{\left(\prod_{j=1}^{n(\mathbf{p})-1}(b + a\,j)\right)\left(\prod_{j=1}^{n(\mathbf{p})}\prod_{k=1}^{e_j-1}(k - a)\right)}{\prod_{j=1}^{n-1}(b + j)}. \tag{8}$$

We refer to random measures based on the specifications $0 \le a < 1$ and $b > 0$ as the Pitman-Yor process, writing this as $\mathcal{PY}(a, b)$. Note that setting $b = \alpha$ and $a = 0$ gives as a special case the Dirichlet process, $\mathrm{DP}(\alpha H)$. In this case, the EPPF formula (8) reduces to a variant of Ewens sampling formula (Ewens (1972); see also Antoniak (1974) and Ewens and Tavaré, (1997)). Another important example is the case where $a = \alpha$ and $b = 0$, which yields a measure whose random weights are based on a stable law with index $0 < \alpha < 1$. See Pitman and Yor (1997) for details.

Parameters based on (7) are the two-parameter processes $(-\kappa, N\kappa)$ and correspond to the class of symmetric finite-dimensional Dirichlet priors, which may be represented as

$$\mathcal{P}_N(\cdot) = \sum_{k=1}^{N} \frac{G_k}{\sum_{k=1}^{N} G_k}\,\delta_{Z_k}(\cdot), \tag{9}$$

where G_k are i.i.d. Gamma(κ) random variables independent of Z_k which are i.i.d. H. That is, the vector of probabilities $G_k/\sum_{k=1}^{N} G_k$ for $k = 1, \ldots, N$ has a Dirichlet distribution with N parameters equal to κ. Pitman (1996) discusses this measure in detail, referring to it as Fisher's model (Fisher, Corbet and Williams (1943)). Watterson (1976) works out the EPPF for (9). Ishwaran and Zarepour (2002a) discuss (9) in context of nonparametric statistical applications (where detailed references can be found) and show in the special case of $\kappa = \alpha/N$ that (9) approximates the Dirichlet process, $\mathrm{DP}(\alpha H)$. Limits under different choices of κ were studied by Ishwaran and Zarepour (2002b). Recently, Gyllenberg and Koski (2001) discuss (9) in applications to bacterial taxonomy, while Hoshino (2001) use it for microdata disclosure risk assessment. Ishwaran, James and Sun (2001) note that (9) has key properties for finite mixture modeling. Setting $\kappa = \alpha/N$ they utililzed the seating rule

$$\ell_{0,n} = \frac{\alpha(1 - n(\mathbf{p})/N)}{n + \alpha}, \qquad \ell_{j,n} = \frac{e_{j,n} + \alpha/N}{n + \alpha},$$

which allows for only a maximum of N occupied tables, thus making it ideally suited for finite mixture models with bounded complexity.

For additional references related to the the two-parameter models see Carlton (1999), Ewens and Tavaré (1997), Ishwaran and James (2001), Kerov (1998), Kerov and Tsilevich (1998), Mekjian and Chase (1997) and Tsilevich (1997).

3. Posterior Distributions for Species Sampling Mixture Models

In general, our methods applies to species sampling mixture models for which the posterior of $\mathcal{P}$ given the data $\mathbf{X} = (X_1, \ldots, X_n)$ can be characterized by

$$\mathcal{P}(dP|\mathbf{X}) = \int \mathcal{P}(dP|\mathbf{Y})\, \mu(d\mathbf{Y}|\mathbf{X}), \tag{10}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and

$$\mu(d\mathbf{Y}|\mathbf{X}) = \frac{\prod_{i=1}^n K(X_i|Y_i)\, \mu(d\mathbf{Y})}{\int_{\mathcal{Y}^n} \prod_{i=1}^n K(X_i|Y_i)\, \mu(d\mathbf{Y})}, \tag{11}$$

where $\mu(d\mathbf{Y})$ is the marginal (urn) distribution for $\mathbf{Y}$. Note that (10) and (11) correspond to species sampling mixture models without a parametric component θ. For convenience we suppress the use of θ here (and in subsequent sections) to simplify notation and some arguments, but the extension to models including θ are straightforward. We return to this issue in Section 5 when we discuss extensions.

3.1. Posterior characterizations

We now derive explicit characterizations for the posterior distribution of a species sampling mixture model. These new results are extensions of Lo's (1984) characterizations for Dirichlet process mixture models and rely on a development of Fubini calculus for the class of species sampling models coupled with results of Pitman (1995, 1996). For (11), the relationship between an EPPF, $\mathbf{Y}$ and its corresponding predictive rule gives the following general analogue of Lemma 1 of BCJL (2001).

Theorem 1. *Let* $\mathbf{Y}$ *be exchangeable with distribution determined sequentially by the prediction rule* (3)*, and whose conditional distribution given* $\mathbf{X}$ *is* $\mu(d\mathbf{Y}|\mathbf{X})$ *defined by* (11)*. Then the conditional distribution of* $\mathbf{Y}$ *given* $\mathbf{X}$ *and a partition* $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$*, is such that the sequence* $Y_1, \ldots, Y_n$ *consists of* $n(\mathbf{p})$ *unique values* $Y_1^*, \ldots, Y_{n(\mathbf{p})}^*$ *which are independent, each with distribution*

$$\mu(dY_j^*|C_j) = \frac{\prod_{i\in C_j} K(X_i|Y_j^*)\, H(dY_j^*)}{\int_{\mathcal{Y}} \prod_{i\in C_j} K(X_i|Y)\, H(dY)}, \qquad \text{for } j = 1, \ldots, n(\mathbf{p}). \tag{12}$$

Moreover, the posterior distribution $\pi(\mathbf{p}|\mathbf{X})$ *is of the form*

$$\pi(\mathbf{p}|\mathbf{X}) = \frac{\pi(\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \int_{\mathcal{Y}} \prod_{i\in C_j} K(X_i|Y)\, H(dY)}{\sum_{\mathbf{p}} \pi(\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \int_{\mathcal{Y}} \prod_{i\in C_j} K(X_i|Y)\, H(dY)}, \tag{13}$$

where $\pi(\mathbf{p}) = p(e_1, \ldots, e_{n(\mathbf{p})})$*, and* p *is the unique EPPF for* $Y_1, \ldots, Y_n$*.*

Proof. The independence result follows from the arguments in the proof of Proposition 13, Pitman (1996) or by Lemma 5 of Hansen and Pitman (2000). That is from (4), given the partition, the independence result and (12) follow immediately. To obtain (13) notice that the marginal density equals

$$f(\mathbf{X}) = \int_{\mathcal{Y}^n} \prod_{i=1}^{n} K(X_i|Y_i)\, \mu(d\mathbf{Y}) = \sum_{\mathbf{p}} \pi(\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \int_{\mathcal{Y}} \prod_{i \in C_j} K(X_i|Y)\, H(dY).$$

A bit of algebraic rearrangement in (11) yields $\pi(\mathbf{p}|\mathbf{X})$. The uniqueness of $\pi(\mathbf{p})$ follows by the uniqueness of the EPPF.

Remark 2. Theorem 1, given the EPPF for a Dirichlet process, immediately yields Lemma 1 of BCJL (2001). In BCJL (2001) they note that, given $\mathbf{p}$ and $\mathbf{X}$, the distribution of Y_j^* has an interpretation as a posterior distribution of Y_j^* given data $\{X_i : i \in C_j\}$, where Y_j^* has prior distribution H. Theorem 1 now shows this interpretation holds for all species sampling models. The main difference in these methods is then summarized in the behavior of the posterior partition distribution (13) which is determined by the choice of different priors $\pi(\mathbf{p})$. This goes back to our earlier point regarding (4), that the EPPF governs the effect the choice of the species sampling prior has on the the posterior.

Remark 3. In principle, Theorem 1 and the methods developed in this paper also apply to other models which are linked either to a species sampling urn structure or, equivalently, an EPPF. In other words, (11) doesn't have to arise when $\mathcal{P}$ is a species sampling model. For instance, this is the case for mixtures of weighted Gamma process random measures with known kernels. There, in the species mixture model (1), P is replaced by a sigma-finite measure which is modeled as a weighted Gamma process. This class of models was considered by Lo and Weng (1989) and was recently discussed in terms of semiparametric Poisson/Gamma random field models by Wolpert and Ickstadt (1998) and Ishwaran and James (2002). Lo and Weng (1989) (see also BCJL(2001) and James(2003)) show that, analogous to Lo (1984) for the Dirichlet process, the posterior distributions of these models are expressible either via a *weighted* Blackwell-MacQueen urn scheme, similar to (11), or an equivalent partition based representation via the Ewens sampling formula.

Theorem 18 of Pitman (1996) coupled with the prediction rule justifies the following disintegrations. For all n, $P(dY_{n+1})\, \mathcal{P}(dP|Y_1, \ldots, Y_n) = \mathcal{P}(dP|Y_1, \ldots, Y_{n+1})\, \mu(dY_{n+1}|Y_1, \ldots, Y_n)$ a.e. H, where for all Borel-measurable B,

$$\mu(B|Y_1, \ldots, Y_n) = \int P(B)\, \mathcal{P}(dP|Y_1, \ldots, Y_n)$$

$$= \ell_{0,n}\, H(B) + \sum_{j=1}^{n(\mathbf{p})} \ell_{j,n}\, I\{Y_j^* \in B\}.$$

Now exploiting the fact that the distribution of Y_i given P and $Y_1, \ldots, Y_{i-1}$ is simply P, we obtain, for all $n \geq 1$, $\prod_{i=1}^{n} P(dY_i)\mathcal{P}(dP) = \mu(d\mathbf{Y})\mathcal{P}(dP|\mathbf{Y})$, where $\mathcal{P}(dP|\mathbf{Y})$ is defined via the updating rule in Theorem 18, Pitman (1996). Given the disintegrations above, we obtain immediately via Fubini's theorem the following analogue of Lemma 1, Lo (1984) (see also Ishwaran and James (2001) for the case of the $\mathcal{PY}(a,b)$ process). Write $\mathcal{M}$ for the space of probability measures over $\mathcal{Y}$ and $\mathcal{B}(\mathcal{M})$ for the corresponding σ-algebra induced by weak convergence.

Lemma 1. *Let g be a positive valued or quasi-integrable function (with respect to the joint probability of $\mathcal{P}$ and P) defined on $(\mathcal{Y}^n \times \mathcal{M}, \mathcal{B}(\mathcal{Y}^n) \times \mathcal{B}(\mathcal{M}))$. Then for all $n \geq 1$,*

$$\int_{\mathcal{M}} \int_{\mathcal{Y}^n} g(\mathbf{Y}, P) \prod_{i=1}^{n} P(dY_i)\, \mathcal{P}(dP) = \int_{\mathcal{Y}^n} \int_{\mathcal{M}} g(\mathbf{Y}, P)\mathcal{P}(dP|\mathbf{Y})\, \mu(d\mathbf{Y}),$$

where $\mathcal{P}(dP|\mathbf{Y})$ and $\mu(d\mathbf{Y})$ are determined by the updating rule in Theorem 18 *of Pitman* (1996). *In particular, $\mu(d\mathbf{Y})$ is the joint law of the species sampling sequence $Y_1, \ldots, Y_n$.*

Remark 4. The prediction rules correspond to the conditional moment measures for $\mathcal{P}$. See LeCam (1986), Daley and Vere-Jones (1988), Matthes, Kerstan and Mecke (1978) and Kallenberg (1986) for details about Fubini's theorem on Polish spaces.

The Fubini result in Lemma 1, in tandem with the prediction rule, facilitates calculations. In particular, one can readily compute higher order moments of products of linear functionals. We discuss this point at the end of the section. Now using Lemma 1 and Theorem 1, we obtain a characterization for the posterior distribution of P given $\mathbf{X}$ in the mixture model setting.

Theorem 2. *The posterior law for a species sampling mixture model is characterized by*

$$\begin{aligned}\int_{\mathcal{M}} g(P)\, \mathcal{P}(dP|\mathbf{X}) &= \sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \int_{\mathcal{M}} g(P)\, \mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p})\, \mu(d\mathbf{Y}^*|\mathbf{X}, \mathbf{p}) \right] \pi(\mathbf{p}|\mathbf{X}) \\ &= \sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \left[\int_{\mathcal{M}} g(P)\mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p}) \right] \prod_{j=1}^{n(\mathbf{p})} \mu(dY_j^*|C_j) \right] \pi(\mathbf{p}|\mathbf{X})\end{aligned}$$

for any nonnegative or integrable function g, where $\pi(\mathbf{p}|\mathbf{X})$ and $\mu(dY_j^*|C_j)$ are defined as in Theorem 1.

Proof. By Bayes rule

$$\int_{\mathcal{M}} g(P)\,\mathcal{P}(dP|\mathbf{X}) = \frac{\int_{\mathcal{M}} g(P)\left[\prod_{i=1}^n \int_{\mathcal{Y}} K(X_i|Y_i)\,P(dY_i)\right]\mathcal{P}(dP)}{\int_{\mathcal{M}}\left[\prod_{i=1}^n \int_{\mathcal{Y}} K(X_i|Y_i)\,P(dY_i)\right]\mathcal{P}(dP)}.$$

An application of Lemma 1 yields the following analogue of Theorem 1, Lo (1984):

$$\begin{aligned}&\int_{\mathcal{M}} g(P)\left[\prod_{i=1}^n \int_{\mathcal{Y}} K(X_i|Y_i)\,P(dY_i)\right]\mathcal{P}(dP)\\ &= \int_{\mathcal{Y}^n}\left[\int_{\mathcal{M}} g(P)\,\mathcal{P}(dP|\mathbf{Y})\right]\prod_{i=1}^n K(X_i|Y_i)\,\mu(d\mathbf{Y}).\end{aligned}$$

Hence $\int_{\mathcal{M}} g(P)\,\mathcal{P}(dP|\mathbf{X}) = \int_{\mathcal{Y}^n}\left[\int_{\mathcal{M}} g(P)\,\mathcal{P}(dP|\mathbf{Y})\right]\mu(d\mathbf{Y}|\mathbf{X})$. The result now follows by using Theorem 1 and simple algebra.

We now provide more explicit details for two-parameter models in the corollary below. The result follows from Theorem 2 in conjunction with Corollary 20 of Pitman (1996). Call a species sampling mixture model with a two-parameter process prior a two-parameter mixture model.

Corollary 1. *The posterior distribution of a two-parameter mixture model is characterized by Theorem* 2 *with* $\pi(\mathbf{p}|\mathbf{X})$ *defined by setting* $\pi(\mathbf{p})$ *equal to* (8). *Moreover, when* $\mathcal{P} \sim \mathcal{PY}(a,b)$,

$$\mathcal{P}(dP|\mathbf{Y}^*,\mathbf{p}) = \sum_{j=1}^{n(\mathbf{p})} \frac{G_j^*}{\sum_{j=1}^{n(\mathbf{p})+1} G_j^*}\,\delta_{Y_j^*}(\cdot) + \left(1 - \sum_{j=1}^{n(\mathbf{p})} \frac{G_j^*}{\sum_{j=1}^{n(\mathbf{p})+1} G_j^*}\right)\mathcal{P}^*(\cdot), \qquad (14)$$

where $G_j^* \overset{\text{ind}}{\sim}$ *Gamma*$(e_j - a)$, $G_{n(\mathbf{p})+1}^* \sim$ *Gamma*$(b + a\,n(\mathbf{p}))$, *and all variables are mutually independent of* $\mathcal{P}^* = \mathcal{PY}(a, b + a\,n(\mathbf{p}))$. *When* $\mathcal{P} = \mathcal{P}_N$ *is a finite dimensional Dirichlet prior as in* (9), *it follows that* $\mathcal{P}(dP|\mathbf{Y}^*,\mathbf{p})$ *can be expressed as in* (14) *for* $a = -\kappa, b = N\kappa$ *and where* $\mathcal{P}^* = \mathcal{P}_{N-n(\mathbf{p})}$ *is an independent finite dimensional Dirichlet prior* (*with* $N - n(\mathbf{p})$ *terms*). *Equivalently,* $\mathcal{P}_{N-n(\mathbf{p})}$ *is a* $(-\kappa, (N - n(\mathbf{p}))\kappa)$ *two-parameter process.*

3.2. Moment calculations

Let f_l denote real-valued functions on $\mathcal{Y}$ and define $Pf_l = \int_{\mathcal{Y}} f_l(y)P(dy)$ for $l = 1,\ldots,q$. We close this section by demonstrating how the results may be used to easily obtain moment calculations for products of functionals. This result is of particular interest in the case where all moments exist since, under additional regularity conditions, this provides a characterization of the joint distribution of $(Pf_1,\ldots,Pf_q)$. This can be used to yield the Markov-Krein identity for $(Pf_1,\ldots,Pf_q)$ that was recently established by Kerov and Tsilevich (1998)

using combinatorial arguments for the case of the $\mathcal{PY}(a,b)$ processes. (An alternative (simple) proof of this identity is given in Tsilevich, Vershik and Yor (2000) using Laplace functionals). Their results extend the work of Cifarelli and Regazzini (1990) for the case of the distribution of the mean functional, $\int yP(dy)$, when P has a Dirichlet process law. The mean case is also discussed in Diaconis and Kemperman (1996) where in addition the result for the joint distribution of functionals like $(Pf_1,\ldots,Pf_q)$ was proposed as an open problem. Tsilevich (1997) establishes the case for the mean with respect to the general two-parameter processes.

The task is to calculate,

$$E\left[\prod_{l=1}^{q}(Pf_l)^{n_l}\right]=\int_{\mathcal{M}}\left[\prod_{l=1}^{q}\prod_{i=1}^{n_l}\int_{\mathcal{Y}}f_l(y_{i,l})\,P(dy_{i,l})\right]\mathcal{P}(dP),$$

where we take, without loss of generality, $n=\sum_{l=1}^{q}n_l$. The above expression can be evaluated quickly by first applying the Fubini result in Lemma 1, and using arguments similar to those used in Theorem 1 or 2 to obtain

$$E\left[\prod_{l=1}^{q}(Pf_l)^{n_l}\right]=\sum_{\mathbf{p}}\pi(\mathbf{p})\prod_{j=1}^{n(\mathbf{p})}\int_{\mathcal{Y}}\prod_{l=1}^{q}f_l^{e_{j,l}}(u)H(du), \tag{15}$$

where $e_{j,l}$ denotes the number of indices associated with f_l in C_j. For the case of the $\mathcal{PY}(a,b)$ processes, the expression in (15) coincides with that in Proposition (10.1) of Kerov and Tsilevich (1998) where the relationship to the Markov-Krein identity is established. Naturally, one may use Theorem 2 to obtain moment expressions for the posterior species sampling mixture models. That is, calculations with respect to $\mathcal{P}(\cdot|\mathbf{X})$.

4. Computational Algorithms

It is clear from the complexity of the posterior distribution $\mathcal{P}(\cdot|\mathbf{X})$ that we need efficient Monte Carlo methods to be able to broadly utilize the species sampling mixture model. This point is reflected in the case of Dirichlet process mixture models where its practical utility has increased dramatically due to computational advances. The explicit structural features of the posterior distributions developed in the previous section allow us to extend procedures based on the Dirichlet process to this more general setting. We start in Sections 4.1-4.3 by presenting the GWCR, a new generalization of the iidWCR algorithm of BCJL (2001). Later in Section 4.4 we discuss a new Gibbs counterpart to the GWCR, what we refer to as a general collapsed Gibbs sampler. Another class of algorithms we discuss are based on sampling the hidden variables $\mathbf{Y}$ without involving the partition structure. These include Pólya urn Gibbs samplers analogous to Escobar (1988, 1994) as well as SIS methods analogous to Kong, Liu and Wong (1994) and Liu (1996). See Sections 4.5 and 4.6, respectively.

Before moving on to the description of our different algorithms, it will be helpful if we first outline some general principles for selecting a suitable algorithm:

1. In general, the choice of algorithm is a tradeoff between computation and the amount of collapsing. The GWCR and collapsed Gibbs samplers do the most collapsing, or integration, as they are based on the partition structure. The computations needed for integrals can sometimes be demanding, in which case methods based on $\mathbf{Y}$ such as Pólya urn Gibbs samplers or Pólya SIS procedures might be preferable.
2. The latter algorithms are likely to be less efficient than ones based on partitions. In particular, functionals expressible in terms of partitions are Rao-Blackwell improvements over those expressed in terms of $\mathbf{Y}$, and thus can be estimated with less variability (BCJL (2001) Corollary 1). An analogous result for all species sampling models follows as an immediate corollary of Theorem 1. Empirical evidence of improved variability using partitions was observed in BCJL (2001), who compared the performance of the iidWCR to an "iidWP" procedure based on sampling $\mathbf{Y}$ in Dirichlet process mixture models. Similar results have been observed for other Monte Carlo procedures. For example, MacEachern (1994) (see also MacEachern and Müller (1998)) note that a collapsed Gibbs sampler based on partitions leads to improved mixing over Pólya urn Gibbs samplers in Dirichlet process mixture models.
3. Selecting between i.i.d. procedures and their Gibbs analogues can depend upon several things. In some cases the issue of convergence of the Markov chain may make i.i.d. procedures more attractive, although often the Gibbs procedures have simple acceleration steps that alleviate slow convergence problems. Typically the i.i.d. procedures are easy for novices to use and require less "tuning". On the other hand, Gibbs procedures can be more attractive in some semiparametric problems where θ is multi-dimensional and update rules for the GWCR or the SIS Pólya are difficult.

4.1. The GWCR process

The key to extending the iidWCR is to understand the basis for the WCR seating rule. BCJL's (2001) idea was to adjust the seating assignment of the prior, the Chinese restaurant process, using a weighted seating rule, which led to what they called the WCR process. Like the Chinese restaurant, the WCR assigns seats to customers sequentially, but now customers are assigned to previously occupied tables with probabilities proportional to the product of the number of occupants (the effect of the prior) and the *Bayesian predictive weight* of the table determined by the data (the effect of the posterior). In this way, customers (or, equivalently, Y_i values) tend to be attracted to tables that contain "similar" customers, where similarity is a function of the data and the type of kernel used in the problem. BCJL (2001) showed that the WCR seating scheme is determined

by the form of the predictive density of, say, X_{n+1} given $\mathbf{p}$ and $\mathbf{X}$ when $\mathcal{P}$ is a Dirichlet process. Using Theorem 2 we can now see what this rule looks like for any general species sampling model, which will suggest to us the extension to the GWCR process and algorithm. Theorem 2 shows that the predictive density is in general

$$E\left[\int_{\mathcal{Y}} K(X_{n+1}|Y)P(dY)|\mathbf{p},\mathbf{X}\right] = \ell_{0,n}\left[\int_{\mathcal{Y}} K(X_{n+1}|u)\,H(du)\right] + \sum_{j=1}^{n(\mathbf{p})} \ell_{j,n}\int_{\mathcal{Y}} K(X_{n+1}|Y_j^*)\,\mu(dY_j^*|C_j), \quad (16)$$

where $\ell_{0,n} = \alpha/(\alpha+n)$ and $\ell_{j,n} = e_j/(\alpha+n)$. The form of (16), in analogy to the WCR process, now identifies the GWCR seating rule. If the first r customers have been seated, the GWCR seats customer $r+1$ to an unoccupied table with probability

$$\frac{\ell_{0,r}}{\lambda(r+1)} \times \int_{\mathcal{Y}} K(X_{r+1}|Y)\,H(dY), \quad (17)$$

or to an occupied table $C_{j,r}$ with probability

$$\frac{\ell_{j,r}}{\lambda(r+1)} \times \frac{\int_{\mathcal{Y}} K(X_{r+1}|Y)\prod_{i\in C_{j,r}} K(X_i|Y)\,H(dY)}{\int_{\mathcal{Y}}\prod_{i\in C_{j,r}} K(X_i|Y)\,H(dY)}, \qquad j=1,\ldots,n(\mathbf{p}_r), \quad (18)$$

where $\lambda(r+1)$ is the appropriate normalizing constant and $\ell_{0,r}, \ell_{j,r}$ are the values from the prediction rule (3) for the specific species sampling model.

Just like the WCR, the update rules (17) and (18) for the GWCR involve a trade-off between the prior and the effect of the data. For example if K is a kernel density, then (17) is proportional to the prior probability $\ell_{0,r}$ that customer $r+1$ is seated at a new table multiplied by the predictive weight for the table under the prior H, while the probability of being assigned to a previous table $C_{j,r}$ is, by (18), proportional to the prior probability $\ell_{j,r}$ and the predictive weight under the conditional distribution for the table $\mu(\cdot|C_{j,r})$, since

$$\int_{\mathcal{Y}} K(X_{r+1}|Y)\frac{\prod_{i\in C_{j,r}} K(X_i|Y)\,H(dY)}{\int_{\mathcal{Y}}\prod_{i\in C_{j,r}} K(X_i|Y)\,H(dY)} = \int_{\mathcal{Y}} K(X_{r+1}|Y)\,\mu(dY|C_{j,r}).$$

Now seat customers $1,\ldots,n$ using the seating rules (17) and (18) to produce a partition $\mathbf{p} = \{C_1,\ldots,C_{n(\mathbf{p})}\}$ of $\{1,\ldots,n\}$. The resulting density for $\mathbf{p}$ is the GWCR density. Analogous to Lemma 3 in BCJL (2001) we have the following result characterizing its relationship to the posterior. The proof is similar to one in BCJL (2001) and is omitted for brevity.

Lemma 2. *The n-step GWCR seating algorithm results in the density $q(\mathbf{p})$ given by $\Lambda(\mathbf{p})q(\mathbf{p}) = \pi(\mathbf{p})f(\mathbf{X}|\mathbf{p})$, where $f(\mathbf{X}|\mathbf{p}) = \prod_{j=1}^{n(\mathbf{p})}\int_{\mathcal{Y}}\prod_{i\in C_j} K(X_i|Y)\,H(dY)$,*

$\pi(\mathbf{p})$ is the EPPF for the species sampling model and $\Lambda(\mathbf{p}) = \lambda(1) \times \cdots \times \lambda(n)$, where $\lambda(1) = \int_{\mathcal{Y}} K(X_1|Y)\, H(dY)$.

4.2. The GWCR algorithm (for approximating laws of functionals)

Theorem 2 combined with the GWCR identity of Lemma 2 shows that the posterior $\mathcal{P}(dP|\mathbf{X})$ can be expressed as

$$\sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \mu(dY_j^*|C_j) \right] \pi(\mathbf{p}|\mathbf{X}) = \frac{\sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \mu(dY_j^*|C_j) \right] \Lambda(\mathbf{p}) q(\mathbf{p})}{\sum_{\mathbf{p}} \Lambda(\mathbf{p}) q(\mathbf{p})}. \quad (19)$$

Now one can use (19) to suggest the GWCR algorithm. The following procedure can be used to obtain approximate draws from $\mathcal{P}(\cdot|\mathbf{X})$ in the case where it is possible to obtain an exact or approximate draw from $\mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p})$ for a given value of $(\mathbf{Y}^*, \mathbf{p})$.

1. Draw a random partition $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ by seating customers $1, \ldots, n$ sequentially, using the seating assignments defined by (17) and (18). Then $\mathbf{p}$ is a draw from the GWCR process with density $q(\mathbf{p})$ having importance weight $\Lambda(\mathbf{p})$.
2. Use the value of $\mathbf{p}$ to draw Y_j^* independently from (12) for $j = 1, \ldots, n(\mathbf{p})$. This yields $\mathbf{Y}^* = (Y_1^*, \ldots, Y_{n(\mathbf{p})}^*)$.
3. Draw a random measure P from $\mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p})$ using the current value of $(\mathbf{Y}^*, \mathbf{p})$ obtained from steps 1 and 2.
4. To approximate the posterior law of a functional $g(P)$, run the previous steps B times independently, obtaining values $P^{(b)}$ with importance weights $\Lambda(\mathbf{p}^{(b)})$ for $b = 1 \ldots, B$. Approximate the law, $\mathcal{P}\{g(P) \in \cdot|\mathbf{X}\}$, with $\sum_{b=1}^{B} I\{g(P^{(b)}) \in \cdot\}\, \Lambda(\mathbf{p}^{(b)}) / \sum_{b=1}^{B} \Lambda(\mathbf{p}^{(b)})$.

Remark 5. Note that by Corollary 1, the algorithm applies readily to the two parameter processes. In particular P is drawn from (14), where the draw from $\mathcal{P}^*$ is exact for the $(-\kappa, (N - n(\mathbf{p}))\kappa)$ two-parameter process, while for a Pitman-Yor process, $\mathcal{PY}(a, b)$, the draw from $\mathcal{P}^*$ will need to be approximated. Several approximation methods are available. In case $\mathcal{P}$ is a $\mathrm{DP}(\alpha H)$ process, one can use the methods discussed in Gelfand and Kottas (2002), Ishwaran and Zarepour (2002a) and Muliere and Tardella (1998). While these techniques should perform well for the Dirichlet process, Ishwaran and James (2001, Theorem 1) suggest that similar types of constructions may not work well for all classes of $\mathcal{PY}(a, b)$. A method that can be applied in general is to approximate

$\mathcal{P}^* = \mathcal{PY}(a, b + a\,n(\mathbf{p}))$ by the random measure based on simulated data from the associated species sampling sequence (3).

4.3. The GWCR algorithm (for approximating mean functionals)

Theorem 2 can also be used to approximate posterior expectations of functionals $h(P)$. However, it is clear by the arguments in Theorem 2 that these calculations inevitably reduce to integrals of functions $g(\mathbf{Y})$ with respect to $\mu(d\mathbf{Y}|\mathbf{X})$, provided $\int_{\mathcal{M}} h(P)\,\mathcal{P}(dP|\mathbf{Y}) = g(\mathbf{Y})$ can be computed explicitly (naturally, if an explicit expression for this quantity is difficult to obtain and it is possible to draw $\mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p})$ then the algorithm in Section 4.2 is more appropriate). It follows that the GWCR approximation schemes discussed here can be used for a wider range of functions than those arising as functionals of P. That is, they apply for the approximation of integrals $E[g(\mathbf{Y})|\mathbf{X}] = \int g(\mathbf{Y})\mu(d\mathbf{Y}|\mathbf{X})$ for arbitrary functions g.

The approximation schemes can be divided into two scenarios depending upon the complexity of the integration. In the first case suppose it is possible to explicitly compute

$$E[g(\mathbf{Y})|\mathbf{X}, \mathbf{p}] = t(\mathbf{p}). \tag{20}$$

For example for a linear functional $h(P) = \int \psi(u)P(du)$, by Theorem 2 and the prediction rule (3), its posterior mean $E[h(P)|\mathbf{X}]$ equals

$$\left[\int_{\mathcal{Y}} \psi(u)\,H(du)\right] \sum_{\mathbf{p}} \ell_{0,n}\,\pi(\mathbf{p}|\mathbf{X}) + \sum_{\mathbf{p}} \left[\sum_{j=1}^{n(\mathbf{p})} \ell_{j,n} \int_{\mathcal{Y}} \psi(Y_j^*)\,\mu(dY_j^*|C_j)\right] \pi(\mathbf{p}|\mathbf{X}),$$

which can be written as $\sum_{\mathbf{p}} t(\mathbf{p})\pi(\mathbf{p}|\mathbf{X})$ provided we can compute $E[\psi(Y_j^*)|C_j]$ explicitly. In general, for arbitrary functions $g(\mathbf{Y})$ such that (20) holds,

$$\int g(\mathbf{Y})\mu(d\mathbf{Y}|\mathbf{X}) = \sum_{\mathbf{p}} t(\mathbf{p})\pi(\mathbf{p}|\mathbf{X}) = \sum_{\mathbf{p}} \frac{t(\mathbf{p})\Lambda(\mathbf{p})q(\mathbf{p})}{\sum_{\mathbf{p}} \Lambda(\mathbf{p})q(\mathbf{p})}.$$

In this case the algorithm is simply a one-step procedure:

1. Draw $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ from the GWCR process. Repeat B times independently and approximate $E[g(\mathbf{Y})|\mathbf{X}]$ with $\sum_{b=1}^{B} t(\mathbf{p}^{(b)})\Lambda(\mathbf{p}^{(b)})/\sum_{b=1}^{B} \Lambda(\mathbf{p}^{(b)})$.

If $t(\mathbf{p})$ cannot be computed explicitly, we use the following algorithm.

1. Draw $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ from the GWCR process. Given $\mathbf{p}$, draw $(\mathbf{Y}^*|\mathbf{X}, \mathbf{p})$ using (12). Repeat B times independently and approximate $E[g(\mathbf{Y})|\mathbf{X}]$ with $\sum_{b=1}^{B} g(\mathbf{Y}^{*(b)}, \mathbf{p}^{(b)})\Lambda(\mathbf{p}^{(b)})/\sum_{b=1}^{B} \Lambda(\mathbf{p}^{(b)})$.

Thus, for example, for approximating $h(P) = \int \psi(u)P(du)$ when $E[\psi(Y_j^*)|C_j]$ is not easily computed, use

$$g(\mathbf{Y}^*, \mathbf{p}) = \ell_{0,n} \int_{\mathcal{Y}} \psi(u)\,H(du) + \sum_{j=1}^{n(\mathbf{p})} \ell_{j,n}\psi(Y_j^*). \tag{21}$$

4.4. Collapsed Gibbs samplers

Theorem 2, in combination with work developed in this section, points to several other algorithms that could be used to fit species sampling mixture models. These can be either based on the partition structure $\mathbf{p}$, such as the collapsed Gibbs sampler we discuss here, or in those cases where the necessary integral computations are difficult, by basing algorithms on $\mathbf{Y}$ as done in Sections 4.5 and 4.6.

MacEachern (1994) discusses a collapsed Gibbs sampler for fitting Dirichlet process mixture models. This method operates on the partition structure and works by creating a Markov chain with stationary distribution $\pi(\mathbf{p}|\mathbf{X})$. The collapsed Gibbs sampler cycles through draws $\pi(\mathbf{p}_j|\mathbf{p}_{-j},\mathbf{X})$ for $j=1,\ldots,n$, where $\mathbf{p}_j$ is the draw for the jth partition and $\mathbf{p}_{-j}$ is the partition associated with $\mathbf{p}_{j-1}$ formed by removing customer j (when $j=1$, $\mathbf{p}_{-1}$ is formed by removing customer 1 from $\mathbf{p}_0$, the current state of the Markov chain). An important point noted by BCJL (2001) is that the Markov transitions are equivalent to the n-step weighted Chinese restaurant seating rule. Thus, for example, $\pi(\mathbf{p}_n|\mathbf{p}_{-n},\mathbf{X})$, the nth draw in one Gibbs cycle, is the weighted Chinese seating rule for seating customer n based on the seating arrangement $\mathbf{p}_{-n}=\{C_{1,n-1},\ldots,C_{n(\mathbf{p}_{-n}),n-1}\}$ for the first $n-1$ customers. This of course readily extends to species sampling models by the use of the more general seating rules (17) and (18), thus suggesting a collapsed Gibbs sampler for species sampling mixture models. By Theorem 2, the draws for $\mathbf{p}$ from such a sampler can be used to approximate laws for functionals or to approximate posterior mean functionals, similar to Sections 4.2 and 4.3, by drawing conditional values for $\mathbf{Y}^*$ or P and then averaging over these draws.

4.5. Pólya urn Gibbs samplers

Ishwaran and James (2001) showed how to extend the Escobar (1988, 1994) Pólya urn Gibbs sampler to models based on the general class of stick-breaking measures. This same idea extends to species sampling models. The method operates on $\mathbf{Y}$ and creates a Markov chain with stationary distribution $\mu(d\mathbf{Y}|\mathbf{X})$ for the species sampling mixture model. Here the Gibbs sampler cycles through draws $\mu(dY_i|\mathbf{Y}_{-i},\mathbf{X})$ for $i=1,\ldots,n$, where $\mathbf{Y}_{-i}$ is the subvector of $\mathbf{Y}$ formed by removing Y_i. For example, by (11) and the prediction rule (3), the conditional draw for Y_n is defined by (a similar result holds for $\mu(dY_i|\mathbf{Y}_{-i},\mathbf{X})$ for $i<n$):

$$\mathbb{P}\{Y_n\in\cdot\,|\,\mathbf{Y}_{-n},\mathbf{X}\}=\ell^*_{0,n-1}\,\mathbb{P}\{Y_n\in\cdot\,|\,X_n\}+\sum_{j=1}^{n(\mathbf{p}_{n-1})}\ell^*_{j,n-1}\,\delta_{Y^*_{j,n-1}}(\cdot),\qquad(22)$$

where $\ell^*_{0,n-1}\propto\ell_{0,n-1}\times\int_{\mathcal{Y}}K(X_n|Y)\,H(dY)$, $\ell^*_{j,n-1}\propto\ell_{j,n-1}\times K(X_n|Y^*_{j,n-1})$, subject to the constraint $\sum_{j=0}^{n(\mathbf{p}_{n-1})}\ell^*_{j,n-1}=1$. Here the $Y^*_{j,n-1}$ are the $n(\mathbf{p}_{n-1})$

unique values for $\mathbf{Y}_{-n} = \{Y_1, \ldots, Y_{n-1}\}$. Note that (22) can be thought of as a generalized Pólya urn distribution. Now using draws from $\mathbf{Y}$, it is possible to approximate functionals of the posterior using Theorem 2. For example, to approximate $h(P) = \int \psi(u)P(du)$, average (21) over the draws for $\mathbf{Y}$.

4.6. SIS Pólya urn samplers

BCJL (2001) extend the SIS techniques discussed in Kong, Liu and Wong (1994) and Liu (1996) to the class of weighted Blackwell-MacQueen Pólya processes. Here we show that the method also applies to species sampling models. Now, analogously to the GWCR draw for $\mathbf{p}$, the idea is to produce a sequential draw for $\mathbf{Y}$ satisfying

$$q(d\mathbf{Y}) = \frac{\mu(d\mathbf{Y})\, f(\mathbf{X}|\mathbf{Y})}{\Lambda(\mathbf{Y})} = \frac{\mu(d\mathbf{Y}) \prod_{i=1}^{n} K(X_i|Y_i)}{\Lambda(\mathbf{Y})}, \tag{23}$$

where $q(d\mathbf{Y})$ is the generalized weighted Pólya density for the species sampling model and $\Lambda(\mathbf{Y})$ is its importance weight.

The density for $q(d\mathbf{Y})$ is derived as follows. Draw Y_1 from $\mu(dY_1|X_1)$. Now, given values $Y_1, \ldots, Y_r$, draw Y_{r+1} from the distribution defined by

$$\mathbb{P}\{Y_{r+1} \in \cdot \,|\, Y_1, \ldots, Y_r, \mathbf{X}\} = \ell^*_{0,r}\, \mathbb{P}\{Y_{r+1} \in \cdot \,|\, X_{r+1}\} + \sum_{j=1}^{n(\mathbf{p}_r)} \ell^*_{j,r}\, \delta_{Y^*_{j,r}}(\cdot),$$

where $Y^*_{j,r}$ are the $n(\mathbf{p}_r)$ unique values for $Y_1, \ldots, Y_r$, $\ell^*_{0,r} = (\ell_{0,r}/\lambda(r)) \int_{\mathcal{Y}} K(X_{r+1}|Y)\, H(dY)$, $\ell^*_{j,r} = (\ell_{j,r}/\lambda(r)) K(X_{r+1}|Y^*_{j,r})$, and $\lambda(r)$ is the normalizing constant. Running this procedure for $r = 2, \ldots, n$ generates a value $\mathbf{Y}$ with the density (23), where $\Lambda(\mathbf{Y}) = \lambda(1) \times \cdots \times \lambda(n)$ and $\lambda(1) = \int_{\mathcal{Y}} K(X_1|Y)\, H(dY)$. These draws can be used to approximate functionals of the posterior. For example, by (11) and Theorem 2, to approximate $h(P) = \int \psi(u)P(du)$, draw B i.i.d. draws for $\mathbf{Y}$ from $q(d\mathbf{Y})$ and take a weighted average of (21) with respect to the importance weights $\Lambda(\mathbf{Y})$.

5. Extensions to the GWCR Algorithm

Here we discuss various extensions to the GWCR algorithm, including a "shuffling" step to reduce order dependence (Section 5.1), a method for dealing with non-conjugacy (Section 5.2), and extensions to semiparamepric models which include parametric values θ (Section 5.3).

5.1. Shuffling the data

For small sample sizes, the seating arrangement produced by the WCR and GWCR process can sometimes depend heavily on the order in which the data

$\mathbf{X}$ are sorted (recall that seat assignments are directly related to the order of the data). See Lee (1999) for some empirical evidence of this problem in small sample sizes for the WCR.

One way to resolve this problem is to randomize the order of the data. Let σ be a permutation of the set $\{1,\ldots,n\}$. Then $\mathcal{P}(dP|\mathbf{X}) = (1/n!)\sum_\sigma \sum_{\mathbf{p}} \mathcal{P}(dP|\mathbf{X}_\sigma, \mathbf{p})\pi(\mathbf{p}|\mathbf{X}_\sigma)$, where the inner sum is applied to the data $\mathbf{X}_\sigma$ under a specific permutation σ. Decomposing the posterior as before, we find that the GWCR algorithm is now extended to include an extra step (say step 0) involving a random permutation of the data, but otherwise remains the same. Note that shuffling the data only directly effects the step where we draw $\mathbf{p}$, which now gives us a draw from an *exchangeable version of the GWCR process*.

5.2. Non-conjugacy

Without the benefits of conjugacy, the GWCR algorithm will require some form of approximation to deal with integrals of the form (18) and to draw from distributions like (12). A simple method for dealing with this problem is to replace H with the approximate random discrete measure $H_N(\cdot) = \sum_{k=1}^N \delta_{Z_k}(\cdot)/N$, where Z_k are i.i.d. H. Notice the simplification in the seating assignment (18), since now for a table $C_{j,r}$,

$$\frac{\int_{\mathcal{Y}} K(X_{r+1}|Y)\prod_{i\in C_{j,r}} K(X_i|Y)\,H_N(dY)}{\int_{\mathcal{Y}}\prod_{i\in C_{j,r}} K(X_i|Y)\,H_N(dY)} = \frac{\sum_{k=1}^N K(X_{r+1}|Z_k)\prod_{i\in C_{j,r}} K(X_i|Z_k)}{\sum_{k=1}^N \prod_{i\in C_{j,r}} K(X_i|Z_k)}.$$

This will also allow us to draw $\mu(dY_j^*|C_j)$ in (12) approximately using

$$\mu_N(dY_j^*|C_j) = \frac{\prod_{i\in C_j} K(X_i|Y_j^*)\,H_N(dY_j^*)}{\int_{\mathcal{Y}}\prod_{i\in C_j} K(X_i|Y)\,H_N(dY)} = \frac{\sum_{k=1}^N \prod_{i\in C_j} K(X_i|Z_k)\,\delta_{Z_k}(dY_j^*)}{\sum_{k=1}^N \prod_{i\in C_j} K(X_i|Z_k)}. \tag{24}$$

To illustrate, consider how this method works for $\mathcal{PY}(a,b)$ processes. Approximate the posterior decomposition (19) by

$$\mathcal{P}(dP|\mathbf{X}) \approx \frac{\displaystyle\int \sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \mathcal{P}_N(dP|\mathbf{Y}^*,\mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \mu_N(dY_j^*|C_j)\right] \Lambda_N(\mathbf{p}) q_N(\mathbf{p}) H^N(d\mathbf{Z})}{\displaystyle\int \sum_{\mathbf{p}} \Lambda_N(\mathbf{p}) q_N(\mathbf{p}) H^N(d\mathbf{Z})}, \tag{25}$$

where $\mathcal{P}_N(dP|\mathbf{Y}^*,\mathbf{p})$, $\mu_N(dY_j^*|C_j)$, $\Lambda_N(\mathbf{p})$ and $q_N(\mathbf{p})$ are the usual expressions, but with H replaced by H_N conditioned on the current value of $\mathbf{Z} = (Z_1,\ldots,Z_N)$. In particular, similar to (14),

$$\mathcal{P}_N(dP|\mathbf{Y}^*,\mathbf{p}) = \sum_{j=1}^{n(\mathbf{p})} \frac{G_j^*}{\sum_{j=1}^{n(\mathbf{p})+1} G_j^*}\,\delta_{Y_j^*}(\cdot) + \left(1 - \sum_{j=1}^{n(\mathbf{p})} \frac{G_j^*}{\sum_{j=1}^{n(\mathbf{p})+1} G_j^*}\right) \mathcal{PY}_N(\cdot), \tag{26}$$

where the G_j^* are as in (14) and are independent of $\mathcal{PY}_N(\cdot)$, which is a modified Pitman-Yor process with parameters $(a, b+a\,n(\mathbf{p}))$ and measure $H_N(\cdot)$ (for fixed $\mathbf{Z}$). That is, $\mathcal{PY}_N(\cdot) = \mathcal{PY}(a, b + a\,n(\mathbf{p}), H_N)$.

5.2.1. The N-GWCR algorithm

The approximation (25), coupled with (24) and (26), suggests how to implement the N-GWCR algorithm.

1. Draw N i.i.d. values $Z_1, \ldots, Z_N$ from H.
2. Given $\mathbf{Z}$, draw $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ sequentially from the N-GWCR process with density $q_N(\mathbf{p})$ and importance weight $\Lambda_N(\mathbf{p})$.
3. Draw $(\mathbf{Y}^*|\mathbf{X}, \mathbf{p}, \mathbf{Z})$ using $\mu_N(dY_j^*|C_j)$ defined by (24).
4. Draw a random measure P_N from (26) conditioned on $(\mathbf{Y}^*, \mathbf{p}, \mathbf{Z})$.

Given B i.i.d. draws from the above algorithm approximate the posterior law of $g(P)$ using

$$\frac{\sum_{b=1}^{B} I\{g(P_N^{(b)}) \in \cdot\}\, \Lambda_N(\mathbf{p}^{(b)})}{\sum_{b=1}^{B} \Lambda_N(\mathbf{p}^{(b)})}.$$

5.3. Models with θ

We now discuss an extension for fitting semiparametric species sampling mixture models (1) involving a parametric value θ. If $m(d\theta)$ is the prior for θ, it follows similar to the previous work that the posterior for such models can be characterized by

$$\mathcal{P}(dP|\mathbf{X}) = \frac{\displaystyle\int_\Theta \sum_{\mathbf{p}} \left[\int_{\mathcal{Y}^{n(\mathbf{p})}} \mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p}) \prod_{j=1}^{n(\mathbf{p})} \mu(dY_j^*|C_j, \theta) \right] \Lambda_\theta(\mathbf{p}) q(\mathbf{p}|\theta)\, m(d\theta)}{\displaystyle\int_\Theta \sum_{\mathbf{p}} \Lambda_\theta(\mathbf{p}) q(\mathbf{p}|\theta)\, m(d\theta)},$$

$$\mu(d\mathbf{Y}^*|\mathbf{X}, \mathbf{p}, \theta) = \prod_{j=1}^{n(\mathbf{p})} \mu(dY_j^*|C_j, \theta) = \prod_{j=1}^{n(\mathbf{p})} \frac{\prod_{i\in C_j} K(X_i|Y_j^*, \theta)\, H(dY_j^*)}{\int_{\mathcal{Y}} \prod_{i\in C_j} K(X_i|Y_j, \theta)\, H(dY)},$$

and $q(\mathbf{p}|\theta)$, the GWCR density for a fixed value of θ with importance weight $\Lambda_\theta(\mathbf{p})$. The decomposition now suggests the following GWCR method.

1. Draw θ from $m(d\theta)$.
2. Given θ, draw $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ sequentially from the associated GWCR density $q(\mathbf{p}|\theta)$ with importance weight $\Lambda_\theta(\mathbf{p})$.
3. Repeat the two steps B times independently, using the draws for $\mathbf{p}^{(b)}$ (and draws from $\mathbf{Y}^*$ or P as needed) with associated importance weights $\Lambda_{\theta^{(b)}}(\mathbf{p}^{(b)})$ to estimate posterior quantities as before.

6. Finite Normal Mixtures

To illustrate the methods we apply the GWCR and N-GWCR algorithm to semiparametric finite normal mixtures subject to the Dirichlet process. These are species sampling mixture models $(X_i|Y_i,\theta) \overset{\text{ind}}{\sim} N(Y_i,\theta)$, $(Y_i|P) \overset{\text{i.i.d.}}{\sim} P$, $P \sim \mathrm{DP}(\alpha H)$, $\theta \sim m(d\theta)$, $i = 1,\ldots,n$, where we will take H to have a $\mathrm{N}(0,A)$ distribution. Note here that the kernel K is

$$K(X_i|Y_i,\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{1}{2\theta}(X_i - Y_i)^2\right).$$

We look at two different scenarios. We first apply the GWCR algorithm by exploiting the conjugacy of H, then we apply the N-GWCR algorithm to the same model, but without taking advantage of conjugacy.

6.1. The GWCR algorithm

From straightforward calculations, the GWCR seating rule assigns customer $r+1$ to a new table with probability

$$\frac{\alpha}{\lambda_\theta(r+1)} \times \frac{1}{\sqrt{2\pi(\theta+A)}} \exp\left(-\frac{X_{r+1}^2}{2(\theta+A)}\right),$$

or to a previous table $C_{j,r}$ with probability

$$\frac{e_{j,r}}{\lambda_\theta(r+1)} \times \sqrt{\frac{\theta + Ae_{j,r}}{2\pi\theta\,[\theta + A(e_{j,r}+1)]}} \times \exp\Big[-\frac{1}{2\theta}\Big(X_{r+1}^2 - \frac{A\left(\sum_{i\in C_{j,r}} X_i + X_{r+1}\right)^2}{\theta + A(e_{j,r}+1)} + \frac{A\left(\sum_{i\in C_{j,r}} X_i\right)^2}{\theta + Ae_{j,r}}\Big)\Big].$$

Moreover, the unique Y_j^* values are drawn using $(Y_j^*|\mathbf{X},\mathbf{p},\theta) \overset{\text{ind}}{\sim} \mathrm{N}(\mu_j,\sigma_j)$, where $\sigma_j^{-1} = e_j/\theta + 1/A$ and $\mu_j = \sigma_j \sum_{i\in C_j} X_i/\theta$. The GWCR algorithm is completed by drawing from $\mathcal{P}(dP|\mathbf{Y}^*,\mathbf{p})$, which we draw from a Dirichlet process approximation as discussed in Ishwaran and Zarepour (2002a).

6.2. The N-GWCR algorithm

The N-GWCR algorithm is run similarly, but with the assignment to a new table replaced with its empirical approximation

$$\frac{\alpha}{\lambda_\theta(r+1)} \times \frac{1}{N\sqrt{2\pi\theta}} \sum_{k=1}^{N} \exp\left(-\frac{1}{2\theta}(X_{r+1} - Z_k)^2\right)$$

and the assignment to a previous table $C_{j,r}$ replaced with

$$\frac{e_{j,r}}{\lambda_\theta(r+1)} \times \frac{\sum_{k=1}^{N} \exp\left[-\frac{1}{2\theta}\left(\sum_{i\in C_{j,r}}(X_i - Z_k)^2 + (X_{r+1} - Z_k)^2\right)\right]}{\sqrt{2\pi\theta}\,\sum_{k=1}^{N} \exp\left[-\frac{1}{2\theta}\sum_{i\in C_{j,r}}(X_i - Z_k)^2\right]},$$

where Z_k are i.i.d. H. The unique Y_j^* are then drawn from (24) and the random measure $\mathcal{P}(dP|\mathbf{Y}^*, \mathbf{p})$ is drawn similarly as above.

6.3. Galaxy data

To illustrate, we re-analyze the galaxy data in Roeder (1990) consisting of the relative velocities in thousands of kilometers per second of $n = 82$ galaxies from six well-separated conic sections of space. This data has been analyzed by several authors using models similar to the one proposed here, including Escobar and West (1995) who used a Pólya urn Gibbs sampler for model fitting.

Figure 1 contains the results for the GWCR algorithm, while Figure 2 is for the N-GWCR ($N = 50$). In both cases, the shuffling mechanism discussed in Section 5.1 was used and both models drew θ from a uniform distribution on $[0, 10]$ using the externalization method outlined in Section 5.3. We took $A = 1,000$ for the variance for H and set $\alpha = 1$ for the Dirichlet process mass parameter. The analyses presented in the figures were based on 20,000 sampled values. As can be seen, the two analyses are virtually identical, with both indicating the presence of anywhere from three to five modes. This suggests how the N-GWCR may perform in situations where the GWCR cannot be used without modification.

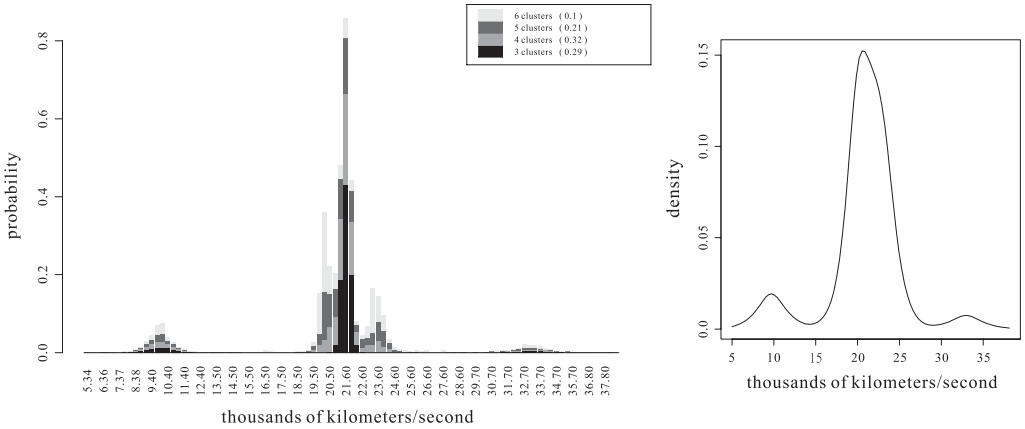

Figure 1. The GWCR algorithm (using conjugacy) with θ drawn from a uniform distribution on $[0, 10]$ (analysis based on 20,000 i.i.d. sampled values). Left-hand side is barplot of atoms and probabilities for the averaged posterior cdf when indexed by the number of partitions (note: (a) barplots are stacked probabilities (b) legend describes frequency of a partition size). Right-hand side is averaged predictive density.

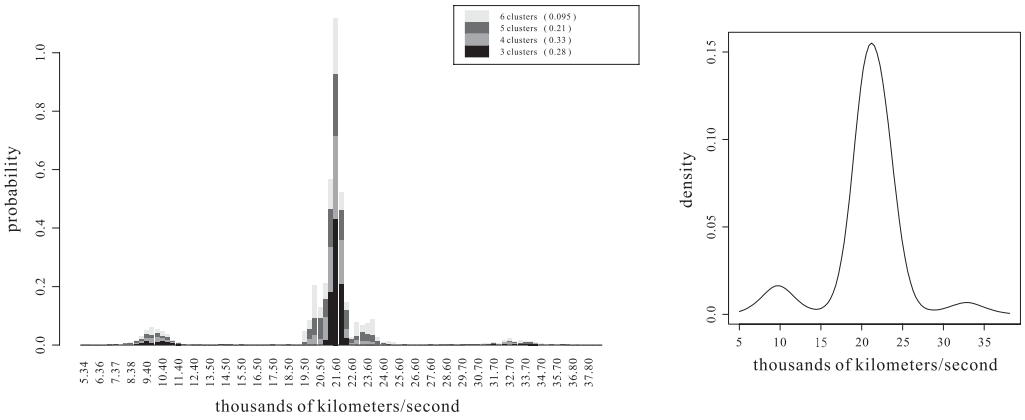

Figure 2. Results of the (non-conjugate) N-GWCR algorithm ($N = 50$).

Acknowledgement

This work partially supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics, Award #00-1, Department of Mathematical Sciences, Johns Hopkins University.

References

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XII* (Edited by P. L. Hennequin). Springer Lecture Notes in Mathematics, Vol. 1117.

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

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353-355.

Brunner, L. J., Chan, A. T., James, L. F. and Lo, A. Y. (2001). Weighted Chinese restaurant processes and Bayesian mixture models. Unpublished Manuscript.

Carlton, M. A. (1999). Applications of the two-parameter Poisson-Dirichlet distribution. Unpublished Ph.D. thesis, Dept. of Statistics, University of California, Los Angeles.

Cifarelli, D. M. and Regazzini, E. (1990). Some remarks on the distribution of the means of a Dirichlet process. *Ann. Statist.* **18**, 429-442.

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.

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

Donnelly, P. and Tavaré, S. (1987). The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.* **25**, 381-391.

Escobar, M. D. (1988). Estimating the means of several normal populations by nonparametric estimation of the distribution of the means. Unpublished Ph.D. thesis, Department of Statistics, Yale University.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *J. Amer. Statist. Assoc.* **89**, 268-277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577-588.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87-112.

Ewens, W. and Tavaré, S. (1997). Multivariate Ewens distribution. In *Discrete Multivariate Distributions* (Edited by S. Kotz and N. Balakrishnan). Wiley, New York.

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

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures *Ann. Statist.* **2**, 615-629.

Fisher, R. A., Corbet, A. S. and Williams C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population *J. Animal Ecology* **12**, 42-58.

Gelfand, A. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.* **11**, 289-305.

Gyllenberg, M. and Koski, T. (2001). Probabilistic models for bacterial taxonomy *Internat. Statist. Rev.* **69**, 249-276.

Hansen, B. and Pitman, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statist. Probab. Lett.* **46**, 251-256.

Hoshino, N. (2001). Applying Pitman's sampling formula to microdata disclosure risk assesment *J. Official Statist.* **17**, 499-520.

Hoppe, F. M. (1987). The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* **25**, 123-159.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161-173.

Ishwaran, H. and James, L. F. (2002). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes and panel count data. Manuscript

Ishwaran H., James, L. F. and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc.* **96**, 1316-1332.

Ishwaran, H. and Zarepour, M. (2002a). Exact and approximate sum-representations for the Dirichlet process. *Canad. J. Statist.* **30**, 269-283.

Ishwaran, H. and Zarepour, M. (2002b). Dirichlet prior sieves in finite normal mixtures *Statist. Sinica* **12**, 941-963.

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. The arXiv, arXiv:math.PR/0205093

James, L. F. (2003). Bayesian calculus for gamma processes with applications to semiparametric models. *Sankhyā Ser. A* **65**, 196-223.

Kallenberg, O. (1986). Random Measures. 4th edition. Akademie-Verlag and Academic Press, Berlin and London.

Kerov, S. (1998). Interlacing measures, *Amer. Math. Soc. Transl.* **181**, 35-83.

Kerov, S. and Tsilevich, N. V. (1998). The Markov-Krein correspondence in several dimensions. PDMI Preprint **1**.

Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89**, 278-288.

Kingman, J. F. C. (1978). The representation of partition structures. *J. London Math. Soc.* **18**, 374-380.

Kuo, L. (1986). Computations of mixtures of Dirichlet processes *SIAM J. Sci. Statist. Comput.* **7**, 60-71.

LeCam, L. (1986), *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

Lee, Y. F. (1999). Some studies oj the weighted Chinese restaurant processes. Unpublished Masters thesis, Hong Kong University of Science and Technology.

Liu, J. S. (1996). Nonparametric hierarchical bayes via sequential imputations. *Ann. Statist.* **24**, 911-930.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351-357.

Lo, A. Y., Brunner, L. J. and Chan, A. T. (1996). Weighted Chinese restaurant processes and Bayesian mixture models. Research Report 1, Hong Kong University of Science and Technology.

Lo, A. Y. and Weng, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rates estimates. *Ann. Inst. Statist. Math.* **41**, 227-245.

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

MacEachern, S. N., Clyde, M. and Liu, J. S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation. *Canad. J. Statist.* **27**, 251-267.

MacEachern, S. N and Müller, P. (1998). Estimating mixture of Dirichlet process models. *J. Comput. Graph. Statist.* **7**, 223-238.

Matthes, K. Kerstan, J., and Mecke, J. (1978). *Infinitely Divisible Point Processes.* English Edition Wiley, Chichester.

Mekjian A. Z. and Chase, K. C. (1997). Disordered systems, power laws and random processes. *Phys. Lett. A* **229**, 340-346.

Muliere P. and Tardella, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *Canad. J. Statist.* **26**, 283-297.

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

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (Edited by T. S. Ferguson, L. S. Shapley and J. B. MacQueen), 245-267. IMS Lecture Notes-Monograph series, Vol. **30**.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855-900.

Quintana, F. A. (1998). Nonparametric Bayesian analysis for assessing homogeneity in $k \times l$ contingency tables with fixed right margin totals. *J. Amer. Statist. Assoc.* **93**, 1140-1149.

Quintana, F. A. and Newton, M. A. (2000). Computationals aspects of nonparametric Bayesian analysis with applications to the modeling of multiple binary sequences. *J. Comput. Graph. Statist.* **9**, 711-737.

Roeder, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *J. Amer. Statist. Assoc.* **85**, 617-624.

Tsilevich, N. V. (1997). Distribution of the mean value for certain random measures. POMI preprint No. 240, Steklov Institute of Mathematics, St. Petersburg.

Tsilevich, N. V., Vershik A. M. and Yor, M. (2000). Distinguished properties of the gamma process and related topics. Prépublications du Laboratoire de Probabilités et Modèles aléatoires, no. 575.

Watterson, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13**, 639-651.

Wolpert R. L. and Ickstadt K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251-267.

Cleveland Clinic Foundation, Department of Biostatistics / Wb4, 9500 Euclid Avenue, Cleveland, U.S.A.

E-mail: ishwaran@bio.ri.ccf.org

Hong Kong University of Science and Technology, Department of Information Systems and Management, Clear Water Bay, Kowloon, Hong Kong.

E-mail: lancelot@ust.hk

(Received October 2001; accepted February 2003)