Statistica Sinica **20** (2010), 581-593

# POSTERIOR CONSISTENCY OF SPECIES SAMPLING PRIORS

Gun Ho Jang[1], Jaeyong Lee[2] and Sangyeol Lee[2]

[1]*University of Toronto and* [2]*Seoul National University*

*Abstract:* Recently there has been increasing interest in species sampling priors, the nonparametric priors defined as the directing random probability measures of the species sampling sequences. In this paper, we show that not all of the species sampling priors produce consistent posteriors. In particular, in the class of Pitman-Yor process priors, the only priors rendering posterior consistency are essentially the Dirichlet process priors. Under certain conditions, we also give a set of necessary and sufficient conditions for the posterior consistency for the general species sampling prior. Considered examples include the normalized inverse-Gaussian process, the Poisson-Kingman partition, and the Gibbs partition.

*Key words and phrases:* Gibbs partition, normalized inverse-Gaussian process, Pitman-Yor process, Poisson-Kingman partition, posterior consistency, species sampling prior.

## 1. Introduction

The species sampling prior is defined as the directing random probability measure of the exchangeable species sampling sequence. As its name indicates, it has been studied as the probability model for applications to population genetics and ecology; see Pitman (1996), Aldous (1985), and references therein. Recently, there has been increasing interest in the species sampling prior as a nonparametric prior; Ishwaran and James (2003), Lijoi, Mena, and Prünster (2005), and Navarrete, Quintana and Müller (2008) discuss some theoretical properties and the mixture modeling of the species sampling prior.

Although the probabilistic properties and computational aspects of the species sampling prior (and species sampling sequence) have been studied extensively, research on its large sample properties as a nonparametric prior has been limited. Exceptions are Lijoi, Prünster, and Walker (2005) and James (2008). Lijoi, Prünster, and Walker (2005) gave a strong consistency result on the normal mixtures of discrete nonparametric prior which can be used to establish the consistency of the normal mixtures of the Pitman-Yor process. In finishing this paper, we learned that James (2008) also independently obtained a consistency result on the Pitman-Yor process prior that we consider in Section 3.

We first investigate the posterior consistency in a specific class of priors, namely Pitman-Yor processes (Pitman and Yor (1997)). The class of Pitman-Yor processes is a subclass of species sampling priors, but is large enough to contain the class of Dirichlet processes as subclass. For Pitman-Yor processes, we obtain necessary and sufficient conditions for posterior consistency; in particular, the Dirichlet process is the only prior that produces posterior consistency for all continuous and discrete distributions. Thus the general class, although a rich source, should not be used as priors for i.i.d. observations. This contrasts with the result obtained by Lijoi, Prünster, and Walker (2005) from which one can deduce strong consistency of normal mixtures of Pitman-Yor process priors. In summary, the Pitman-Yor process prior can be used in the form of the normal mixtures, but should not be used as the prior without mixtures.

Under certain conditions, we also give a set of necessary and sufficient conditions for the posterior consistency of the general species sampling prior. The necessary and sufficient conditions are given in terms of prediction probability functions. We then consider, as examples, the normalized inverse-Gaussian process, the Poisson-Kingman partition, and the Gibbs partition.

The plan for the paper is as follows. In Section 2, we present the basic theory and examples of the species sampling prior and species sampling sequence. The consistency result for the Pitman-Yor process is given in Section 3. Lastly, in Section 4, we give general results for the posterior consistency of the general species sampling prior.

## 2. Species Sampling Prior

### 2.1. Species sampling prior

Suppose $(X_1, X_2, \ldots)$ is a sequence of random variables with values in a complete separable metric space $\mathcal{X}$. Imagine this sequence is a random sample from a large population of various species, i.e. $X_i$ is the species of the $i$th sampled individual.

Let $\tilde{X}_j$ be the $j$th distinct species to appear in the sequence $(X_1, X_2, \ldots)$. Let $n_{jn}$ be the number of times the $j$th species $\tilde{X}_j$ appears in $(X_1, \ldots, X_n)$ and $\mathbf{n}_n = (n_{1n}, \ldots, n_{k_n,n})$ where $k_n$ is the number of different species in $(X_1, \ldots, X_n)$. For simplicity, the subscript $n$ in $n_{jn}, k_n$ and $\mathbf{n}_n$ is dropped if it is not confusing.

Let $\nu$ be a diffuse (or atomless) probability measure on $\mathcal{X}$. An exchangeable sequence $(X_1, X_2, \ldots)$ is called a species sampling sequence if $X_1 \sim \nu$ and

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

where $\delta_x$ is the degenerate probability measure at $x$ and

$$p_j(\mathbf{n}) = \mathbb{P}(X_{n+1} = \tilde{X}_j | X_1, \ldots, X_n), \quad j = 1, \ldots, k,$$
$$p_{k+1}(\mathbf{n}) = \mathbb{P}(X_{n+1} \notin \{X_1, \ldots, X_n\} | X_1, \ldots, X_n).$$

The sequence of functions $(p_1, p_2, \ldots)$ defined on $\mathbb{N}^* = \cup_{k=1}^{\infty} \mathbb{N}^k$, where $\mathbb{N}$ is the set of natural numbers, is called a sequence of prediction probability functions and satisfies the conditions $p_j(\mathbf{n}) \geq 0$ and $\sum_{j=1}^{k+1} p_j(\mathbf{n}) = 1$, for $\mathbf{n} \in \mathbb{N}^*$.

A sequence of random variables $(X_n)$ is a species sampling sequence if and only if $X_1, X_2, \ldots | F$ is random sample from $F$ where

$$F = \sum_{i=1}^{\infty} P_i \delta_{\tilde{X}_i} + R\nu \tag{2.1}$$

for some sequence of positive random variables $(P_i)$ and $R$ such that $1 - R = \sum_{i=1}^{\infty} P_i \leq 1$, $(\tilde{X}_i)$ is a random sample from $\nu$, and $(P_i)$ and $(\tilde{X}_i)$ are independent. See Pitman (1996).

The above 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 $F$ in (2.1) the *species sampling prior (or process)* of the species sampling sequence $(X_i)$. The most celebrated example of the species sampling prior is the Dirichlet process.

**Example 1 (Dirichlet Process).** Sethuraman (1994) showed that the Dirichlet process can be represented in form (2.1). Suppose $\theta > 0$ and $\nu$ is a probability measure. Let $W_1, W_2, \ldots$ be an i.i.d. sequence from $Beta(1, \theta)$. From $(W_i)$, discrete probability masses $(P_i)$ are constructed by the stick-breaking process, that is,

$$P_1 = W_1 \quad \text{and} \quad P_j = W_j \prod_{i=1}^{j-1} (1 - W_i), \quad j = 2, 3, \ldots. \tag{2.2}$$

Suppose $\tilde{X}_1, \tilde{X}_2, \ldots$ is an i.i.d. sequence from $\nu$, and $(P_i)$ and $(\tilde{X}_i)$ are independent, then, $F = \sum_{j=1}^{\infty} P_j \delta_{\tilde{X}_j} \sim DP(\theta\nu)$.

**Example 2 (Pitman-Yor Process).** Pitman and Yor (1997) introduced an interesting class of discrete random measures which includes the Dirichlet process. Let $b$ and $a$ be real numbers with either $0 \leq a < 1$ and $b > -a$, or $a < 0$ and $b = -ma$ for some $m = 1, 2, \ldots$, and let $\nu$ be a diffuse probability measure. Construct $(P_i)$ from $(W_j)$ by the stick-breaking process as in (2.2), where each $W_j$ is independently sampled from $Beta(1-a, b+ja)$. Let $(\tilde{X}_j)$ be an i.i.d. sequence from $\nu$ independent of $(P_j)$. The random probability measure $F = \sum_{j=1}^{\infty} P_j \delta_{\tilde{X}_j}$ is called the Pitman-Yor process, denoted by $PY(a, b, \nu)$ in this paper. Note $PY(0, \theta, \nu)$ is $DP(\theta\nu)$.

## 2.2. Exchangeable partition probability function

Let $[n] = \{1, \ldots, n\}$. An exchangeable sequence of random variables $(X_i)$ defines a random partition $\Pi = \{A_1, \ldots, A_k\}$ of $[n]$, where $A_i = \{j \in [n] : X_j = \tilde{X}_i\}$. Then,

$$p(\#(A_1), \ldots, \#(A_k)) = \mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\})$$

defines a function from $\mathbb{N}^*$ to $[0, 1]$. The function $p$ is called the exchangeable partition probability function (EPPF) derived from the exchangeable sequence $(X_n)$.

The following properties of the EPPF can be found in Pitman (1995). An EPPF $p$ derived from an exchangeable sequence $(X_n)$ satisfies

$$p(1) = 1, \quad \text{and} \quad p(\mathbf{n}) = \sum_{j=1}^{k(\mathbf{n})+1} p(\mathbf{n}^{j+}), \text{ for all } \mathbf{n} \in \mathbb{N}^*, \tag{2.3}$$

where $\mathbf{n}^{j+}$ is the same as $\mathbf{n}$ except that the $j$th element is increased by 1. Conversely, every symmetric function $p : \mathbb{N}^* \to [0, 1]$ satisfying (2.3) is an EPPF of some exchangeable sequence.

Note that the distribution of the species sampling prior is completely determined by the those of $(P_i, i = 1, 2, \ldots)$ and $(\tilde{X}_i, i = 1, 2, \ldots)$ as in (2.1), which in turn can be parametrized by the prediction probability function $(p_j, j = 1, 2, \ldots)$ and the diffuse probability measure $\nu$. Moreover, the $p_j$ can be specified as $p(\mathbf{n}^{j+})/p(\mathbf{n})$. The species sampling prior characterized by an EPPF $p$ and a diffuse probability measure $\nu$ is denoted by $SSP(p, \nu)$.

## 2.3. Examples of species sampling priors

In this subsection, we give three different ways extending the Dirichlet process, all of which are subclasses of species sampling priors.

**Example 3 (Normalized Inverse-Gaussian (N-IG) Process).** Lijoi, Mena, and Prünster (2005) defined the N-IG process $P$ by specifying the distribution of $(P(B_1), \ldots, P(B_k))$, for a partition $B_1, \ldots, B_k$ of $\mathcal{X}$, as the distribution of $(V_1, \ldots, V_k)/V$, where $V = V_1 + \cdots + V_k$ and $V_i \overset{\text{ind}}{\sim} IG(\theta\nu(B_i), 1)$, $i = 1, \ldots, k$. Here $IG(a, b)$ denotes the inverse-Gaussian distribution with parameters $a > 0$ and $b > 0$, whose density is $a(2\pi x^3)^{-1/2} \exp(-(a^2/x + b^2 x)/2 + ab)$ for $x > 0$.

By a calculation similar to that given in Lijoi, Mena, and Prünster (2005), one can show the N-IG process is the species sampling prior with predictive distribution $\mathbb{P}(X_{n+1} \in B | X_1, \ldots, X_n) = w_{1,n} \sum_{j=1}^{k} (n_j - 1/2) \delta_{\tilde{X}_j}(B) + w_{0,n}\nu(B)$, where

$$w_{0,n} = \frac{a \int_1^\infty (1-y^{-2})^n y^k e^{-ay} dy}{2n \int_1^\infty (1-y^{-2})^{n-1} y^{k-1} e^{-ay} dy} \quad \text{and} \quad w_{1,n} = \frac{\int_1^\infty (1-y^{-2})^n y^{k-1} e^{-ay} dy}{n \int_1^\infty (1-y^{-2})^{n-1} y^{k-1} e^{-ay} dy}.$$

**Example 4 (Poisson-Kingman Partition).** Pitman (1995, 1996, 2003) and Gnedin and Pitman (2006) developed many classes of EPPFs, that are closely related to the random partition studied by Kingman (1975) and Aldous (1985). Here we summarize two different random partitions: the Poisson-Kingman partition and the Gibbs partition. Since the species sampling prior is characterized by an EPPF $p$ and a diffuse probability measure $\nu$, these classes of EPPFs also define new classes of species sampling priors.

The definition of Poisson-Kingman partition adopts an alternative definition of the Dirichlet process as given in Ferguson (1973). Let $J_1, J_2, \ldots$ be the jump sizes of the Poisson point process with the intensity (or Lévy) measure $\Lambda$. The normalized $J_i$'s, $J_i/T$, play the role of $P_i$'s in (2.1) with $R = 0$, where $T = J_1 + J_2 + \cdots$. Sufficient conditions for $T < \infty$ a.s. are

$$\int_0^1 x d\Lambda(x) < \infty \text{ and } \int_1^\infty d\Lambda(x) < \infty.$$

The EPPF of the Poisson-Kingman distribution $PK(\rho)$ with Lévy density $\rho$ is given by

$$p(n_1, \ldots, n_k) = \frac{(-1)^{n-k}}{\Gamma(n)} \int_0^\infty u^{n-1} e^{-\psi(u)} \prod_{j=1}^k \psi_{n_j}(u) du,$$

where $\psi(u) = \int_0^\infty (1 - e^{-ux})\rho(x)dx$ and $\psi_m(u) = (d^m\psi/du^m)(u) = (-1)^{m-1} \int_0^\infty x^m e^{-ux}\rho(x)dx$ for $m = 1, 2, \ldots$ (see Pitman (2003) for details). We call the species sampling prior characterized by $PK(\rho)$ and a diffuse probability measure $\nu$ the Poisson-Kingman prior (or process), and denote it by $PK(\rho, \nu)$.

The predicted probability function $p_j(\mathbf{n})$ of a $PK(\rho, \nu)$ process is given by $p_j(\mathbf{n}) = p(\mathbf{n}^{j+})/p(\mathbf{n})$ for $j = 1, \ldots, k$, and $p_{k+1}(\mathbf{n}) = 1 - (p_1(\mathbf{n}) + \cdots + p_k(\mathbf{n}))$.

**Example 5 (Gibbs Partition).** Gnedin and Pitman (2006) generalized the EPPF of the Dirichlet process. An EPPF $p$ is called the EPPF of *Gibbs form* if $p(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^k W_{n_j}$, for some nonnegative weights $W = (W_j)$ and $V = (V_{n,k})$.

Under the assumption $W_1 = V_{1,1} = 1$, every Gibbs partition is represented by $W_j$'s and $V_{n,k}$'s satisfying

$$W_j = \begin{cases} 1 & \text{if } j = 1, \\ b^{j-1} \prod_{i=0}^{j-2} (1 - a + i) & j = 2, 3, \ldots \end{cases} \quad \text{and} \quad V_{n,k} = b(n - ak)V_{n+1,k} + V_{n+1,k+1}$$

for some $b > 0$ and $a < 1$.

The predictive probability functions $p_j(\mathbf{n})$ for $j = 1, \ldots, k$ are

$$p_j(\mathbf{n}) = \frac{p(\mathbf{n}^{j+})}{p(\mathbf{n})} = \frac{V_{n+1,k}}{V_{n,k}} \frac{W_{n_j+1}}{W_{n_j}} = V_*(n,k)(n_j - a),$$

where $V_*(n,k) = bV_{n+1,k}/V_{n,k}$. The EPPF of the N-IG process has a Gibbs form. Pitman-Yor process also has the EPPF of Gibbs form. In this aspect, the EPPF of Gibbs form is an extension of the two-parameter Poisson-Dirichlet process.

## 3. Posterior Consistency of Pitman-Yor Prior

In this section, we investigate the consistency property of the Pitman-Yor process prior. Throughout this paper, we consider the following nonparametric model with various nonparametric priors:

$$\begin{aligned} X_1, \ldots, X_n | P &\sim P, \\ P &\sim \mathcal{P}, \end{aligned} \tag{3.1}$$

where $\mathcal{P}$ is a nonparametric prior on $P$. In this section, we consider the posterior consistency under (3.1) when $\mathcal{P}$ is the Pitman-Yor process prior $PY(a, b, \nu)$, where $a$ and $b$ satisfy either $0 \leq a < 1$ and $b > -a$ or $a < 0$ and $b = m|a|$ for some $m = 1, 2, \ldots$, and $\nu$ is a diffuse probability measure. The posterior of this prior family is known as

$$P|X_1, \ldots, X_n = \sum_{j=1}^{k} \tilde{P}_j \delta_{\tilde{X}_j} + \tilde{R}_k F_k, \tag{3.2}$$

where $(\tilde{P}_1, \ldots, \tilde{P}_k, \tilde{R}_k) \sim Dir(n_1 - a, \ldots, n_k - a, b + ka)$ independent of $F_k \sim PY(a, b + ka, \nu)$.

To investigate the behavior of the posterior under the sampling distribution, we need to postulate the form of the true probability measure, $P_0$, from which the sample, $X_1, X_2, \ldots$ is drawn. We assume that $P_0$ is decomposed into the discrete and atomless parts,

$$P_0 = \sum_j q_j \delta_{z_j} + \lambda \mu, \tag{3.3}$$

where $z_j \in \mathcal{X}, q_1 \geq q_2 \geq \cdots \geq 0, \lambda = 1 - \sum_j q_j \leq 1$, and $\mu$ is a diffuse probability measure. Let $\mathcal{Z} = \{z_1, z_2, \ldots\}$.

**Theorem 1.** *Suppose $X_1, X_2, \ldots$ is an i.i.d. sequence from $P_0$ of form (3.3). Under the model (3.1) with prior $PY(a, b, \nu)$, the posterior given $X_1, \ldots, X_n$ is weakly consistent at $P_0$ if and only if one of the following holds: $a = 0$; when $a > 0$, $P_0$ is discrete or $\mu = \nu$; $a < 0$ and $P_0$ is a mixture of at most $m = |b/a|$ degenerated measures.*

**Remark 1.** If $P_0$ is discrete, all Pitman-Yor process priors with $0 \leq a < 1$ entail consistent posteriors. However, if $P_0$ is continuous, the Dirichlet process is the only prior among the Pitman-Yor process priors which renders posterior consistency.

**Remark 2.** The story is completely different in the mixture setting. Consider the normal mixture model

$$\begin{aligned} X_i|\theta_i, h &\sim \text{ ind } N(\theta_i, h^2), \quad i = 1, \ldots, n, \\ \theta_i|P &\sim \text{i.i.d.} P, \qquad\qquad i = 1, \ldots, n, \\ P &\sim PY(a, b, \nu), \\ h^2 &\sim \mu, \end{aligned} \tag{3.4}$$

where $P$ and $h$ are independent *a priori*. Then, $\mu \times \mathcal{P}$ is a prior on $\mathcal{F}$, the class of all densities with respect to Lebesgue measure. Suppose the supports of $\mu$ and the diffuse probability measure $\nu$ are $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}$, respectively.

By Theorem 3 in Ghosal, Ghosh and Ramamoorthi (1999), every element $f_{h,P}(x) = \int \phi_h(x - y) dP(y)$ of $\mathcal{G} = \{f_{h,P} : h > 0, P \text{ is compactly supported}\}$ is in the Kullback-Leibler support of $\mathcal{P}$, where $\phi_h$ is the density of the normal distribution with mean 0 and variance $h^2$. Schwartz (1965) implies that $\mu \times \mathcal{P}$ is weakly consistent at all $f_{h,P} \in \mathcal{G}$. Moreover, strong consistency can be obtained from Theorem 1 in Lijoi, Prünster, and Walker (2005).

**Remark 3.** The second condition, in part, has the diffuse probability measure $\nu$ proportional to the continuous part $\mu$ of the true probability measure $P_0$. This is impractical.

**Remark 4.** The same result was obtained in James (2008). He used a slightly different approach to get the result using the notion of the seminorm and did not consider the case $a < 0$.

Before we present the proof of Theorem 1, we need the following two lemmas which will be used subsequently. Most technical details of the paper, including the proofs of the two lemmas, are given in the Appendix of the paper that can be found at http://www.stat.sinica.edu.tw/statistica.

**Lemma 1.** *Let $\mathcal{M}$ be the class of all Borel probability measures on a complete separable metric space $\mathcal{X}$, and $\mathcal{P}$ be a prior on $\mathcal{M}$. Suppose we postulate the model* (3.1) *and that $X_1, X_2, \ldots$ is an i.i.d. sequence from the true probability measure $P_0$. The posterior is weakly consistent at $P_0$ if and only if for any $P_0$-continuity set $U$ of $\mathcal{X}$,*

(i) $\lim_{n\to\infty} \mathbb{E}(P(U)|X_1, \ldots, X_n) = P_0(U), \quad P_0^\infty - a.s.,$

(ii) $\lim_{n\to\infty} \mathbb{V}ar(P(U)|X_1,\ldots,X_n) = 0, \quad P_0^\infty - a.s.,$

*where $P_0^\infty$ is the infinite product of the true probability measure $P_0$.*

**Lemma 2.** *Suppose $X_1, X_2, \ldots, X_n$ are sampled from $P_0$ of the form (3.3). Let $\tilde{X}_1,\ldots,\tilde{X}_k$ be the distinct values among $X_1,\ldots,X_n$, $k^* = \sum_{j=1}^k I(\tilde{X}_j \notin \mathcal{Z})$, and $G_k$ be the empirical distribution of $\tilde{X}_1,\ldots,\tilde{X}_k$. Then,*
(i) $k_n/n \to \lambda$ *and* $k_n^*/n \to \lambda$, $P_0^\infty - a.s.$,
(ii) $G_{k_n} \to \mu$, $P_0^\infty - a.s.$ *if* $\lambda > 0$.

**Proof of Theorem 1.** First, consider the case $a < 0$. Then, $b = m|a|$ for some integer $m$ and the Pitman-Yor process is a finite random mixture of point random measures. By Theorem 4.3.1 of Ghosh and Ramamoorthi (2003) we get that the posterior is consistent at $P_0$ if and only if $P_0$ is a discrete probability measure having at most $m$ point masses.

For $a \geq 0$. Lemma 3 in the Appendix shows that

$$\begin{aligned}\mathbb{E}[P(B)|X_1,\ldots,X_n] &= [\frac{n}{n+b}]\tilde{F}_n(B),\\ \mathbb{E}[P(B)^2|X_1,\ldots,X_n] &= [(n+b)(n+b+1)]^{-1}(n^2\tilde{F}_n(B)^2 + n\tilde{F}_n(B)\\ &\quad -an(1-\nu(B))),\end{aligned}$$

where $\tilde{F}_n(B) = n^{-1}\sum_{i=1}^n \delta_{X_i}(B) - ak_nG_{k_n}(B)/n + (b+ak_n)\nu(B)/n$. The Strong Law of Large Numbers and Lemma 2 yield $\tilde{F}_n(B) \to P_0(B) - a\lambda(\mu(B) - \nu(B))$ and $\mathbb{V}ar(P(B)|X_1,\ldots,X_n) \to 0$, $P_0^\infty - a.s.$ Then, the conclusion of the theorem follows immediately by Lemma 1.

**Remark 5.** *In the proof, we have identified the weak limit of the posterior, that is, $P_0 - a\lambda I(a \geq 0)(\mu - \nu)$.*

## 4. Posterior Consistency of Species Sampling Prior

We consider the same true probability measure (3.3) and the nonparametric model (3.1) with

$$\mathcal{P} = SSP(p,\nu), \tag{4.1}$$

where $p$ is an EPPF and $\nu$ is a diffuse probability measure. For the results in this section, we need two assumptions. We need *the smoothness condition* for the prediction probability function: as $n \to \infty$,

$$S_n = S_n(\mathbf{n}) = \max_{1\leq i\leq k}\sum_{j=1}^k \left|p_j(\mathbf{n}) - p_j(\mathbf{n}^{i+})\right| \to 0, \;\; P_0^\infty - a.s., \tag{4.2}$$

and *the separability condition* on the support of the discrete part of $P_0$: there exists $\epsilon > 0$ such that for all $i \neq j$

$$d(z_i, z_j) > \epsilon, \tag{4.3}$$

where $d$ is the metric of $\mathcal{X}$.

A sufficient condition for posterior consistency is given below. The proof is given in the Appendix.

**Proposition 2.** *Suppose $X_1, \ldots, X_n$ is an i.i.d. sequence from $P_0$ of form* (3.3)*. Under the model* (3.1) *with prior* (4.1) *and the smoothness condition* (4.2)*, the posterior $P$ given $X_1, \ldots, X_n$ is weakly consistent at $P_0$ if*

$$C_n = C_n(\mathbf{n}) = \sum_{j=1}^{k} \left| p_j(\mathbf{n}) - \frac{n_j}{n} \right| \to 0, \ \ P_0^\infty - a.s. \quad as \quad n \to \infty. \tag{4.4}$$

**Remark 6.** Condition (4.4) gives an intuitive sufficient condition for posterior consistency: the posterior predictive distribution, the conditional distribution of $X_{n+1}$ given $X_1, \ldots, X_n$, should behave like the empirical distribution of $X_1, \ldots, X_n$. Proposition 3 and Theorem 4 show that this is almost a necessary and sufficient condition.

**Remark 7.** The smoothness condition (4.2) for the predictive probability function $p_j(\mathbf{n})$ ensures a small change in $\mathbf{n}$ does not change $p_j(\mathbf{n})$ much. For instance, note that

$$C_{n+1}(\mathbf{n}^{i+}) \le S_n + C_n + \frac{2n}{n(n+1)},$$

which together with (4.4) and (4.2) implies $C_{n+1}(\mathbf{n}^{i+}) \to 0$, $P_0^\infty - a.s.$.

If the true probability measure $P_0$ is atomless, (4.4) is equivalent to

$$\sum_{j=1}^{k} \left| p_j(\mathbf{n}) - \frac{n_j}{n} \right| = n \left| \frac{1}{n} - p_k(\mathbf{n}) \right| = |1 - np_k(\mathbf{n})| = p_{n+1}(\mathbf{n}) \to 0, P_0^\infty - a.s.,$$

which is also equivalent to $np_k(\mathbf{n}) \to 1$, $P_0^\infty - a.s.$ This argument is generalized in the next proposition, whose proof is given in the Appendix.

**Proposition 3.** *Suppose that $X_1, X_2, \ldots$ is an i.i.d. sequence from $P_0$ of form* (3.3)*. Set $k^* = \sum_{j=1}^{k} I(\tilde{X}_j \notin \mathcal{Z})$ and*

$$p_+(\mathbf{n}) = \begin{cases} p_j(\mathbf{n}) & \text{if } n_j = 1 \text{ for some } j \le k, \\ 0 & \text{otherwise.} \end{cases}$$

*If the predictive probability function $p_j s$ satisfies*

$$\lim_{n\to\infty}\sum_{j=1}^{k}|p_j(\mathbf{n})-\frac{n_j}{n}|I(\tilde{X}_j\in\mathcal{Z})=0,\quad P_0^\infty-a.s., \tag{4.5}$$

*then the following hold.*

(i) *For all Borel set $B$, $\sum_{i=1}^{k}p_i(\mathbf{n})I(\tilde{X}_i\in\mathcal{Z}\cap B)\to\sum_{j=1}^{\infty}q_jI(z_j\in B)$, $P_0^\infty-a.s.$;*

(ii) *$\sum_{i=1}^{k}p_i(\mathbf{n})I(\tilde{X}_i\in\mathcal{Z})\to 1-\lambda$, and $k^*p_+(\mathbf{n})+p_{k+1}(\mathbf{n})\to\lambda$, $P_0^\infty-a.s.$;*

(iii) *(4.4) is equivalent to $p_{k+1}(\mathbf{n})\to 0$ as $n\to\infty$, $P_0^\infty$-a.s. Furthermore, if $\lambda>0$, (4.4) is equivalent to $np_+(\mathbf{n})\to 1$, as $n\to\infty$, $P_0^\infty-a.s.$.*

**Remark 8.** Note that even if there are multiple $j$ with $n_j=1$, $p_+(\mathbf{n})$ is well defined by the exchangeability of the EPPF $p$.

**Theorem 4.** *Suppose $X_1,X_2,\ldots$ is an i.i.d. sequence from $P_0$ of form* (3.3) *with separability condition* (4.3)*. Under the model* (3.1) *with a prior* (4.1) *that satisfies the smoothness condition* (4.2)*, the posterior given $X_1,\ldots,X_n$ is weakly consistent at $P_0$ if and only if the predictive probability function satisfies* (4.5) *and either $p_{k+1}(\mathbf{n})\to 0$ as $n\to\infty$, $P_0^\infty-a.s.$, or $P_0$ is a mixture of a discrete probability measure and the diffuse measure $\nu$.*

The proof of "If" part relies essentially on Lemma 1. The mean and variance of the predictive probability are represented by the prediction probability function of the prior and their limits calculated under the conditions of Theorem 4. The proof of "Only if" part proceeds similarly. First, one assumes that $P_0$ is a mixture of discrete probability measure and a diffuse probability measure $\mu$ different from $\nu$. The limit of the predictive probability is identified as $P_0-\phi(\mu-\nu)$, where $\phi$ is the limit of $p_{k+1}(\mathbf{n})$ in case it converges, and one concludes $\phi=0$. The details of the proof are given in the supplemental note.

**Remark 9.** The first condition in Theorem 4 is natural in the following sense. Since $p_{k+1}(\mathbf{n})$ is the predictive probability that $X_{n+1}$ is sampled from $\nu$, we expect that $p_{k+1}(\mathbf{n})\to 0$ as $n\to\infty$, if posterior consistency holds.

**Example 6 (N-IG Process Prior).** Recall that the predictive probability function $p_j(\mathbf{n})$ of the N-IG process is

$$p_j(\mathbf{n})=(n_j-\frac{1}{2})w_{1,n}=\frac{n_j-1/2}{n}\frac{\int_1^\infty(1-y^{-2})^n y^{k-1}e^{-ay}dy}{\int_1^\infty(1-y^{-2})^{n-1}y^{k-1}e^{-ay}dy}.$$

Note $nw_{1,n} \to 1$ by Lemma 7 in the Appendix, which guarantees (4.2) and (4.5) because

$$\sum_{j=1}^{k} \Big| p_j(\mathbf{n}) - p_j(\mathbf{n}^{i+}) \Big| = \sum_{j=1}^{k} \Big| \frac{n_j - 1/2}{n} nw_{1,n} - \frac{n_j^{i+} - 1/2}{n+1} (n+1) w_{1,n+1} \Big|$$
$$\leq |nw_{1,n} - nw_{1,n+1}| + |w_{1,n+1}| \to 0, \quad P_0^\infty - a.s.$$
$$\sum_{j:\tilde{X}_j \in \mathcal{Z}} \Big| p_j(\mathbf{n}) - \frac{n_j}{n} \Big| \leq |nw_{1,n} - 1| + \frac{k - k^*}{2n} nw_{1,n} \to 0, \qquad P_0^\infty - a.s..$$

The second condition of Theorem 4 implies that the N-IG process prior produces a consistent posterior at all discrete probability measures.

Now suppose $P_0$ is non-discrete or, equivalently, $\lambda > 0$. Note that $np_+(\mathbf{n}) = nw_{1,n}/2 \to 1/2 < 1$, $P_0^\infty - a.s.$. Thus, all N-IG processes produce inconsistent posteriors at all probability measures that do not satisfy the second condition of Theorem 4. In particular, the N-IG process produces inconsistent posterior at all diffuse probability measures except $\nu$.

**Example 7 (Poisson-Kingman process).** In this example, we consider the Poisson-Kingman process with generalized gamma Lévy density $\rho_{a,b,c}(x) = cx^{-a-1} e^{-bx}$ for $0 < a < 1, b \geq 0$ and $c > 0$; the class is large enough to contain Dirichlet Processes – $DP(\theta\nu)$ is equivalent to $PK(\rho_{0,1,\theta}, \nu)$. The EPPF of $PK(\rho_{a,b,c}, \nu)$ is

$$p(n_1, \ldots, n_k) = \frac{c^k \prod_{j=1}^{k} \Gamma(n_j - a)}{\Gamma(n)} \int_0^\infty u^{n-1} e^{-c\Gamma(1-a)[(b+u)^a - b^a]/a} (b+u)^{ak-n} du.$$

The predictive probability function is given by $p_j(\mathbf{n}) = [(n_j - a)/n] w(n,k)$, where

$$w(n,k) = \frac{\int_0^\infty u^n (b+u)^{ak-n-1} e^{-c\Gamma(1-a)(b+u)^a/a} du}{\int_0^\infty u^{n-1} (b+u)^{ak-n} e^{-c\Gamma(1-a)(b+u)^a/a} du}.$$

By Lemma 8, we get $w(n,k) \to 1$ as $n \to \infty$. Then, (4.5) can be shown using the above result and Lemma 2 (i),

$$\sum_{j:\tilde{X}_j \in \mathcal{Z}} \Big| p_j(\mathbf{n}) - \frac{n_j}{n} \Big| \leq \frac{a(k-k^*)}{n} + \frac{n - a(k-k^*)}{n} \Big| w(n,k) - 1 \Big| \to 0. \quad P_0^\infty - a.s.$$

Also, it is not hard to see (4.2) holds. By Theorem 4 (ii), $PK(\rho_{a,b,c}, \nu)$ is consistent at all discrete probability measures. For the non-discrete probability measures, we can check the first condition of Theorem 4, or equivalently Proposition 3 (iii): $np_+(\mathbf{n}) = (1-a)w(n,k) \to 1 - a. P_0^\infty - a.s.$. Therefore, the Poisson-Kingman prior $PK(\rho_{a,b,c}, \nu)$ is inconsistent at all continuous probability measures except $\nu$ when $a > 0$.

**Example 8 (Gibbs partition).** The predicted probability function is

$$p_j(\mathbf{n}) = \frac{V_{n+1,k} W_{n_j+1}}{V_{n,k} W_{n_j}} = \frac{nbV_{n+1,k}}{V_{n,k}} \frac{n_j - a}{n}.$$

It is not hard to see that (4.5) is equivalent to $nbV_{n+1,k}/V_{n,k} \to 1$, which also implies (4.2). Thus, Theorem 4 implies the species sampling prior generated by a Gibbs partition is consistent at all discrete probability measures if $nbV_{n+1,k}/V_{n,k} \to 1$. For a non-discrete $P_0$, assume $nbV_{n+1,k}/V_{n,k} \to 1$. Since as $n \to \infty$

$$np_+(\mathbf{n}) = n \frac{V_{n+1,k}}{V_{n,k}} \frac{W_2}{W_1} = \frac{nbV_{n+1,k}}{V_{n,k}} (1-a) \to 1-a,$$

the posterior is consistent if and only if $a = 0$, thus a mixture of Dirichlet processes.

## 5. Conclusion

The species sampling prior, which is gaining increasing interest as a class of nonparametric priors, was tested for consistency in this paper. We found that, among all the Pitman-Yor priors, the only priors consistent at diffuse probability measures are Dirichlet processes. The same conclusion holds for the popular subclasses of species sampling priors, the N-IG process, the Poisson-Kingman process, and the Gibbs partition. This does not mean that the only consistent priors among the species sampling priors are Dirichlet processes and that the species sampling priors are not useful. First of all, the species sampling priors can be useful in mixture modelling as we discussed in Section 3. Second, as we have characterized the class of consistent species sampling priors in Section 4, the class of consistent species sampling priors can be still large. We believe that more research is necessary to develop flexible subclasses of the species sampling priors that are consistent.

## Acknowledgement

The authors are grateful for the helpful comments from an associate editor and the referee. Jaeyong Lee was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-070-C00021).

## References

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour*, XIII—1983, 1-198.

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

Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 209-230.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics.* Springer-Verlag.

Gnedin, A. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* **138**, 5674-5685.

Ishwaran, H. and James, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica* **13**, 1211-1235.

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* (Edited by B. Clarke and S. Ghosal), 187-199. Institute of Mathematical Statistics.

Kingman, J. F. C. (1975). Random discrete distribution. *J. Roy. Statist. Soc. B* **37**, 1-22.

Lijoi, A., Mena, R. H. and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100**, 1278-1291.

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.

Navarrete, C., Quintana, A. F. and Müller, P. (2008). Some issues on nonparametric Bayesian modeling using species sampling models. *Statist. Model.* **8**, 3-21.

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*, 245-267. Inst. Math. Statist., Hayward, CA.

Pitman, J. (2003). Poisson-Kingman partitions. In *Statistics and Science: a Festschrift for Terry Speed*, 1-34. Inst. Math. Statist., Beachwood, OH.

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

Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10–26.

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

Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada.

E-mail: gunho@utstat.utoronto.ca

Department of Statistics, Seoul National University, Seoul 151-747, Korea.

E-mail: leejyc@gmail.com

Department of Statistics, Seoul National University, Seoul 151-747, Korea.

E-mail: ylee@stats.snu.ac.kr

(Received January 2008; accepted January 2009)