Biometrika (2013), **xx**, x, *pp*. 1–14
© 2007 Biometrika Trust
Printed in Great Britain

Species sampling models: consistency for the number of species

BY P.G. BISSIRI, A. ONGARO

Dipartimento di Economia, Metodi Quantitative e Strategie d'Impresa

Università degli Studi di Milano–Bicocca

Edificio U7, via Bicocca degli Arcimboldi 8, 20126 Milano, Italy

pier.bissiri@unimib.it andrea.ongaro@unimib.it

AND S.G. WALKER

School of Mathematics, Statistics & Actuarial Sciences, University of Kent

Canterbury, Kent, CT2 7NZ, United Kingdom

s.g.walker@kent.ac.uk

SUMMARY

This paper considers species sampling models using constructions which arise from Bayesian nonparametric prior distributions. A discrete random measure, used to generate a species sampling model, can either have a countable infinite number of atoms, which has been the emphasis in the recent literature, or a finite number of atoms K, while allowing K to be assigned a prior probability distribution on the positive integers. It is the latter class of model we consider here, due to the existence and interpretation of K as the number of species. We demonstrate the consistency of the posterior distribution of K as the sample size increases.

Some key words: Bayesian consistency; Exchangeable random partition; Gibbs–type partition; Species sampling model.

1. INTRODUCTION

This paper is concerned with species sampling models. The idea we present here is motivated by recent work appearing in Lijoi et al. (2007, 2008) and Favaro et al. (2009). The problem is to estimate the number of species in a population, early work on which can be found in many papers. See, for instance, Efron & Thisted (1976), Hill (1979), Boender & Rinnooy Kan (1987), Chao & Lee (1992), Chao & Bunge (2002), Chao et al. (2009), Zhang & Stern (2005), Wang & Lindsay (2005), Wang (2010) and Barger & Bunge (2010).

Lijoi et al. (2007) are predominantly concerned with estimating the number of new species in a further sample of size m having previously observed a sample of size n. For this, Bayesian nonparametric models are employed and, specifically, discrete random probability measures are used, such as the Dirichlet process and the two parameter Poisson–Dirichlet process. More generally, two classes used are the class of normalized random measures, which are driven by non–decreasing Lévy processes, and Gibbs-type priors (Lijoi et al., 2008, Favaro et al., 2009). These models assume that the number of species is infinite, claiming that if the number of species in the population is large, then it is reasonable to assume that it is infinite (Favaro et al., 2009, Lijoi et al., 2007). Probably this was done because the mathematics is more attractive for such models. The model we use here assumes that the number of species K in the population is finite. Therefore, having assigned a prior for K, we can consider estimating it. Moreover, we can prove consistency of the posterior. In other words, the sequence of posterior distributions for K accumulates at the true value as the sample size increases.

2. THE MODEL

Let K be the random number of species in the population, and let $V_1, \ldots, V_K$ be the absolute frequencies of the K species in the population, where $\{V_j\}$ is a sequence of positive, independent and identically distributed random variables and $\{V_j\}$ is independent of K. Given that there are K species, let $P_{1,K}, \ldots, P_{K,K}$ be the relative frequencies of the species in the population, namely, $P_{j,K} = V_j / \sum_{l=1}^{K} V_l$ $(j = 1, \ldots, K)$. Clearly, the joint conditional distribution of $P_{1,K}, \ldots, P_{K,K}$ given K is exchangeable and $\sum_{j=1}^{K} P_{j,K} = 1$.

Now assume that observations X_i $(i \geq 1)$ take values in a measurable space $(\mathbb{X}, \mathscr{X})$, and that the observations $X_1, X_2, \ldots$ are sampled from the random probability measure

$$\sum_{j=1}^{K} P_{j,K}\, \delta_{Z_j}, \tag{1}$$

where $\{P_{j,K} : j = 1, \ldots, K\}$ and $\{Z_j\}$ are two independent sequences, the Z_j are independent and identically distributed random variables with values in $(\mathbb{X}, \mathscr{X})$ and the distribution α of Z_1 is diffuse, that is $\alpha\{x\} = 0$ for every x in $\mathbb{X}$. Let the prior π for K be such that $\pi(k) = \mathbb{P}(K = k)$ is positive for every $k \geq 1$, where $\mathbb{P}$ is the probability measure that underlines all the random variables above.

The above model belongs to the class of species sampling models introduced by Pitman (1996), which has been widely studied in the statistical literature. A species sampling process is a random probability measure of the form $\sum_{j=1}^{\infty} P_j\, \delta_{Z_j} + (1 - \sum_{j=1}^{\infty} P_j)\alpha$, where $\{P_j\}$ and $\{Z_j\}$ are two independent sequences of random variables such that $P_j \geq 0$ for every $j \geq 1$ and $\sum_{j=1}^{\infty} P_j \leq 1$, almost surely, the Z_j are independent and identically distributed random variables with values in $(\mathbb{X}, \mathscr{X})$ and α is the distribution of Z_1, and it is diffuse. So, the model under consideration is a species sampling model with finitely many positive weights, as considered by Ongaro & Cattaneo (2004) and Ongaro (2004, 2005). Whereas we will focus on the posterior distribution of K, Ongaro considered the posterior of the underlying random measure given by (1).

For our model (1), the posterior for K is

$$\pi_n(k) = \mathbb{P}(K = k \mid X_1, \ldots, X_n)$$
$$= \frac{\pi(k)k(k-1)\cdots(k-K_n+1)\mathbb{E}\left(\prod_{j=1}^{K_n} P_{j,k}^{n_j}\right)}{\sum_{l=K_n}^{\infty} \pi(l)l(l-1)\cdots(l-K_n+1)\mathbb{E}\left(\prod_{j=1}^{K_n} P_{j,l}^{n_j}\right)} \mathbb{I}_{\{k \geq K_n\}}, \quad (2)$$

where $n_j = \left|\{i = 1, \ldots, n : X_i = X_j^*\}\right|$, for $j = 1, \ldots, K_n$, $|A|$ denotes the cardinality of a set A, K_n is the number of different species among $X_1, \ldots, X_n$, and $X_1^*, \ldots, X_{K_n}^*$ are the distinct values of $X_1, \ldots, X_n$.

We can also provide predictive distributions for other quantities, most important of which would be the species of the next observation or the number of new species in a further sample. But to emphasize what sets our model apart from the previous ones, we focus on results for the number of species.

We briefly highlight the difference between our model and more classic models, such as the mixed Poisson model. While both rely on multinomial structures, they are different; in our model, and in fact for all species sampling models, it is the frequencies of species $P_{j,K}$ which are modeled conditional on K, but, with the classic models, it is the number of species with the same number of observations which is modeled conditional on K. If $f_{j,K}$ denotes the number of species with j observations, then $K_n = \sum_{j=0}^{K} f_{j,K}$ and $n = \sum_{j=0}^{K} j f_{j,K}$. In this way, the sample size n is random, and this is the practical difference between the classical models and the species sampling models.

3. CONSISTENCY

Let $\mathbb{P}_0$ denote the true population from which the observations $X_1, X_2, \dots$ are sampled with replacement. The observations are discrete, independent and identically distributed random variables under the probability measure $\mathbb{P}_0$. As before, $\mathbb{P}$ denotes the probability measure making $(X_i)_{i\geq 1}$ an exchangeable sequence directed by (1). Let $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$, the probability measure that yields the posterior and predictive distributions, while $\mathbb{P}_0$ generates the data.

Let k_0 be the true unknown number of species in the population, that is, the number of possible outcomes of each X_i under $\mathbb{P}_0$. We want to find conditions on the law of V_1 to ensure that the posterior π_n of K is consistent, that is, $\lim_{n\to\infty} \pi_n(k_0) = 1$, $\mathbb{P}_0$–almost surely. Before proceeding, denote $T_{l,t} = \{(x_1, \dots, x_l) \in \mathbb{R}^l : x_j > 0, 1 \leq j \leq l,\ \sum_{j=1}^{l} x_j < t\}$, for every $t > 0$ and $l \geq 1$. Moreover, let $T_l = T_{l,1}$, namely, the l–dimensional open simplex.

THEOREM 1. *Assume that:*

a) π has a finite k_0-th moment and $\pi(k_0) > 0$;

b) the distribution of V_1 is absolutely continuous with respect to Lebesgue measure and it has a density f_{V_1} that is positive on $(0, M)$ or on (M, ∞), for some $M > 0$;

c) for every $l \geq 2$, the density of $(P_{1,l}, \dots, P_{l-1,l})$, that is

$$g_l(x_1, \dots, x_{l-1}) = \int_{[0,\infty)} y^{l-1} f_{V_1}(y(1 - \textstyle\sum_{j=1}^{l-1} x_j)) \prod_{j=1}^{l-1} f_{V_1}(yx_j)\,\mathrm{d}y, \qquad (3)$$

is continuous on T_{l-1};

d) each k–dimensional marginal of g_l, that is

$$g_l^{(k)}(x_1, \dots, x_k) = \int_{T_{l-1-k,1-\sum_{j=1}^{k} x_j}} g_l(x_1, \dots, x_{l-1})\,\mathrm{d}x_{k+1} \cdots \mathrm{d}x_{l-1}, \qquad (4)$$

is continuous on T_k*, for* $k = 1, \ldots, l-1$ *and* $l \geq 3$.

Then the posterior π_n *is consistent.*

COROLLARY 1. *If the assumptions of Theorem 1 hold, and* π *admits the* (k_0+1)*–th moment, then the Bayes estimator is consistent:* $\lim_{n\to\infty} \mathbb{E}(K \mid X_1, \ldots, X_n) = k_0$, $\mathbb{P}_0$*-almost surely.*

The proof of the Theorem is deferred to the Appendix. The proof of the Corollary is similar and is omitted.

4. GIBBS MODELS

4·1. *Gibbs–type priors: definition and main properties*

A relevant case for our model is given by the Gibbs–type priors with finitely many species, studied by Gnedin & Pitman (2006) and Pitman (2006). We shall now introduce them, and we shall show how they can be used for the estimation of the number of species in a population.

For each integer $n \geq 1$, denote by Π_n the random partition of $\{1, \ldots, n\}$ generated by $(X_1, \ldots, X_n)$ in the sense that any $i \neq j$ belong to the same partition set if and only if $X_i = X_j$. Recall that the probability distribution of a species sampling sequence is characterized by the marginal distribution α of a single observation and the exchangeable partition probability functions for each $n \geq 1$, that is, the probability distribution of the random partition Π_n,

$$p(n_1, \ldots, n_k) = \mathbb{P}(\Pi_n \in \{A_1, \ldots, A_k\}) = \sum_{(i_1,\ldots,i_k)\in E_k} \mathbb{E}\left(\prod_{j=1}^k P_{i_j}^{n_j}\right),$$

where $\{A_1, \ldots, A_k\}$ is a partition of $\{1, \ldots, n\}$, n_j is the cardinality of A_j, for $j = 1, \ldots, k$, $n = \sum_{j=1}^k n_j$ and E_k is the set of all ordered k-tuples of distinct positive integers. A Gibbs–type prior is obtained if for each $n \geq 1$ the exchangeable partition probability function is

$p(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^{k} (1-\sigma)_{n_j - 1}$, for every $n \geq 1$, and some $\sigma < 1$, where $(a)_n = a(a+1) \cdots (a+n-1)$ for any $n \geq 1$ and $(a)_0 = 1$.

In the case of Gibbs–type priors, the representation (1) with finite K holds true if and only if $\sigma < 0$. This is the setup we examine in this paper. Gnedin & Pitman (2006) prove that each Gibbs–type prior with $-\infty < \sigma < 0$ is such that the conditional distribution of $(P_1, \ldots, P_{K-1})$ given K is symmetric Dirichlet with K parameters equal to $a = |\sigma|$. Conditionally on K, the directing random probability measure is distributed as a two–parameter Poisson–Dirichlet process, introduced by Pitman (1995) and widely studied (Prünster & Lijoi, 2009). For $a < \infty$, this is equivalent to letting $\{V_j\}$ be a sequence of independent and identically distributed random variables with a common Gamma distribution, having shape parameter a and scale parameter 1. The limiting case $a = +\infty$ is obtained by taking $P_{j,k} = 1/k$, for every integer $k \geq 1$, namely, $V_j = 1$ for every $j \geq 1$. This model is called coupon collecting by Pitman (2006). The exchangeable partition probability function depends on $(n_1, \ldots, n_{K_n})$ only through n and K_n. Therefore, any inference based on this model with $a = \infty$ does not take into account the frequencies of the species observed in the sample.

For finite a, the posterior for K is

$$\pi_n(k) = \frac{\pi(k)k(k-1)\cdots(k-K_n+1)\Gamma(ka)/\Gamma(n+ka)}{\sum_{l \geq K_n} \pi(l)l(l-1)\cdots(l-K_n+1)\Gamma(la)/\Gamma(n+la)} \mathbb{I}_{\{k \geq K_n\}}.$$

For this model, two different samples of the same size n and with the same number of distinct values K_n yield the same posterior for K, the same predictive distribution, and clearly also the same Bayes estimator.

4·2. *Consistency and rate of convergence of the posterior*

By Theorem 1, the posterior π_n for this model is consistent. However, in this case, consistency can be proved directly without resorting to the assumptions of Theorem 1. In particular, no assumption about the existence of the moments of π is required. Moreover, it is possible to obtain the convergence rate of $\pi_n(k_0)$. In fact, we can state the following result:

PROPOSITION 1. *Let the distribution of* $(P_{1,k}, \dots, P_{k-1,k})$ *be symmetric Dirichlet with* k *parameters equal to* a*, for some* $a > 0$ *and every integer* $k \geq 1$*. Then* π_n *is consistent and*

$$\pi_n(k_0) \sim 1 - c(k_0)\frac{\Gamma(k_0 a + a)}{\Gamma(k_0 a)}\frac{1}{n^a}, \tag{5}$$

as n *diverges* $\mathbb{P}_0$*–almost surely for* $a < \infty$*, where* $c(k_0) = (1 + k_0)\pi(k_0 + 1)/\pi(k_0)$*, and*

$$\pi_n(k_0) \sim 1 - c(k_0)\left(\frac{k_0}{1 + k_0}\right)^n, \tag{6}$$

as n *diverges,* $\mathbb{P}_0$*–almost surely, for* $a = \infty$*.*

The proof of Proposition 1 is deferred to the Appendix.

A similar result for mixture models, where the number of mixtures replaces the number of species, is obtained by Rousseau & Mengersen (2011). In species sampling models we are interested in the weights corresponding to distinct locations and not where the locations are. Typically, in mixture models, when the number of mixtures replaces the number of species, locations are important.

ACKNOWLEDGEMENT

The authors would like to thank three anonymous reviewers who provided comments and suggestions on an earlier version of the paper. This work was partially supported by the European Social Fund and Regione Lombardia within the grant Dote Ricercatori.

APPENDIX

We now state a lemma, whose proof can be obtained by Jacobi's transformation formula.

LEMMA 1. *If* $(W_1, \ldots, W_l)$ *is* $(0,\infty)^l$*–valued random vector with density* h *with respect to the* l*–dimensional Lebesgue measure, then a density for* $(W_1/\sum_{j=1}^l W_j, \ldots, W_{l-1}/\sum_{j=1}^l W_j, \sum_{j=1}^l W_j)$ *is:*

$$\bar{h}(t_1, \ldots, t_{l-1}, s) = s^{l-1} h(st_1, \ldots, st_{l-1}, s(1 - \sum_{j=1}^{l-1} t_j) \mathbb{I}_{T_{l-1}\times(0,\infty)}(t_1, \ldots, t_{l-1}, s).$$

The following lemma will be useful for the proof of Theorem 1:

LEMMA 2. *Assume that* $\pi(k_0) > 0$. *The posterior* π_n *is consistent if and only if*

$$\lim_{n\to\infty} \sum_{l>k_0} \frac{\pi(l)}{\pi(k_0)} C(l, k_0) \frac{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)}{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)} = 0. \quad \text{(A1)}$$

for every $l \geq 1$*, where* p_j *is the* $\mathbb{P}_0$*-probability that* X_1 *is equal to* X_j^**,* $j = 1, \ldots, k_0$*, and* $C(m, k)$ *is the binomial coefficient of choosing* k *from* m*, that is* $m!/\{k!(m-k)!\}$.

Proof. Let $a_{l,n} = \pi(l) l(l-1)\cdots(l-k_0+1)\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)$. Since $K_n = k_0$ for big n almost surely,

$$\pi_n(k_0) \sim a_{k_0,n}/\sum_{l\geq k_0} a_{l,n} = 1 - \sum_{l>k_0} a_{l,n}/a_{k_0,n}\{1 + \sum_{l>k_0} a_{l,n}/a_{k_0,n}\}^{-1}, \quad \text{(A2)}$$

as $n \to \infty$, $\mathbb{P}_0$–almost surely. Hence, as n diverges, $\pi_n(k_0)$ goes to one if and only if $\sum_{l>k_0} a_{l,n}/a_{k_0,n}$ goes to zero and the proof is complete. □

Proof of Theorem 1. For every $l > k_0$, let $S_{l,k_0} = \sum_{j=1}^{k_0} V_j / \sum_{j=1}^{l} V_j$, for every $l > k_0$, and $Z_n = S_{l,k_0}^n Y_n$, where $Y_n = \mathbb{E}(n^{(k_0-1)/2} \exp\{-n \sum_{j=1}^{k_0} p_j \ln(p_j/P_{j,k_0})\} \mid S_{l,k_0})$, for every $n \geq 1$. Moreover, it is convenient to rewrite the ratio in (A1):

$$\frac{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)}{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)} = \frac{\mathbb{E}\left[\exp\{-n\sum_{j=1}^{k_0} p_j \ln(p_j/P_{j,l})\}\right]}{\mathbb{E}\left[\exp\{-n\sum_{j=1}^{k_0} p_j \ln(p_j/P_{j,k_0})\}\right]} = \frac{\mathbb{E}(Z_n)}{\mathbb{E}(Y_n)}. \quad \text{(A3)}$$

We shall deal with the numerator and the denominator separately. Let us deal with the denominator first. By Lemma 1 in the Appendix, g_{k_0} is a density for the distribution of $(P_{1,k_0}, \ldots, P_{k_0-1,k_0})$. By hypothesis c), taking $l = k_0$, such density is continuous on T_{k_0-1}. Moreover, by hypothesis b), the support of $(P_{1,k_0}, \ldots, P_{k_0-1,k_0})$ is the (k_0-1)-dimensional closed simplex. In fact, the transformation

$(v_1,\dots,v_{k_0}) \longrightarrow (v_1/\sum_{j=1}^{k_0} v_j,\dots,v_{k_0-1}/\sum_{j=1}^{k_0} v_j)$ maps $(0,M]^{k_0}$ onto the (k_0-1)–dimensional simplex and the same is true for $[M,\infty)^{k_0}$, for every $M>0$. Hence, the density g_{k_0} is positive on T_{k_0-1}.

In particular, this density is positive and continuous in $(p_1,\dots,p_{k_0-1})$. Therefore, it is possible to apply the multi–dimensional Laplace method (Hsu, 1948) to obtain:

$$\lim_{n\to\infty} \mathbb{E}(Y_n) = c_2 g_{k_0}(p_1,\dots,p_{k_0-1}), \tag{A4}$$

where $c_2 = (2\pi)^{(k_0-1)/2}\,|h_\phi(p_1,\dots,p_{k_0-1})|^{-1/2}$, and h_ϕ is the determinant of the Hessian matrix of the function $\phi(x_1,\dots,x_{k_0-1}) = \sum_{j=1}^{k_0-1} p_j \ln(p_j/x_j) + p_{k_0}\ln\left\{p_{k_0}/(1-\sum_{j=1}^{k_0-1} x_j)\right\}$.

By (A3) and (A4), there is a constant c_1 such that

$$\frac{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)}{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)} \le c_1 n^{(k_0-1)/2}\,\mathbb{E}\left[\exp\{-n\sum_{j=1}^{k_0} p_j \ln(p_j/P_{j,l})\}\right] = c_1\,\mathbb{E}\left(Z_n\right), \tag{A5}$$

for every $n \ge 1$.

A density for $(P_{1,k_0},\dots,P_{k_0-1,k_0},S_{l,k_0})$ is

$$g_{l,k_0}(x_1,\dots,x_{k_0-1},s) = s^{k_0-1} g_l^{(k_0)}\{sx_1,\dots,sx_{k_0-1},s(1-\sum_{j=1}^{k_0-1} x_j)\}. \tag{A6}$$

In fact, $S_{l,k_0} = \sum_{j=1}^{k_0} P_{j,l}$, $P_{j,k_0} = P_{j,l}/\sum_{j=1}^{k_0} P_{j,l}$ for $1\le j\le k_0$, and one can apply Lemma 1 in the Appendix taking $W_j = P_{j,l}$ $(1\le j\le k_0)$ to obtain (A6). Hence, a conditional density of $(P_{1,k_0},\dots,P_{k_0-1,k_0})$ given S_{l,k_0} is

$$g_{l,k_0}(x_1,\dots,x_{k_0-1},s)/\bar{g}_{l,k_0}(s)\mathbb{I}_{\{\bar{g}_{l,k_0}>0\}}(s), \tag{A7}$$

where $\bar{g}_{l,k_0}$ is a density for S_{l,k_0}. By hypotesis d), (A6) is continuous as a function of $(x_1,\dots,x_{k_0-1})$ on T_{k_0-1} and so is (A7). Moreover, by hypothesis a), (A7) is also positive on T_{k_0-1}. Hence, by the multi–dimensional Laplace method,

$$\lim_{n\to\infty} Y_n = c_2 g_{l,k_0}(x_1,\dots,x_{k_0-1},S_{l,k_0})/\bar{g}_{l,k_0}(S_{l,k_0})\mathbb{I}_{\{\bar{g}_{l,k_0}(S_{l,k_0})>0\}}, \tag{A8}$$

almost surely. Moreover,

$$\mathbb{E}(\lim_{n\to\infty} Y_n) = c_2 g_{k_0}(p_1,\dots,p_{k_0-1}). \tag{A9}$$

To prove (A9), it is sufficient to verify that:

$$\mathbb{E}\{g_{l,k_0}(p_1,\dots,p_{k_0-1},S_{l,k_0})/\bar g_{l,k_0}(S_{l,k_0})\mathbb{I}_{\{\bar g_{l,k_0}(S_{l,k_0})>0\}}\} = \int_{[0,1]\cap\{\bar g_{l,k_0}>0\}} g_{l,k_0}(p_1,\dots,p_{k_0-1},y)\mathrm{d}y$$
$$= g_{k_0}(p_1,\dots,p_{k_0-1}).$$

This can be done combining (A6), (4) and (3) and then computing the integral by substitution.

Combination of (A4) and (A9) yields that $\mathbb{E}(\lim_{n\to\infty} Y_n) = \lim_{n\to\infty}\mathbb{E}(Y_n)$. Since $0 \le Z_n \le Y_n$,for every $n \ge 1$, and $\lim_{n\to\infty} Z_n = 0$, $\mathbb{P}$–almost surely, this fact allow us to apply the Pratt's lemma (Gut, 2005, page 221–222) to obtain that $\lim_{n\to\infty}\mathbb{E}(Z_n) = 0$. Therefore, by (A5),

$$\lim_{n\to\infty} \frac{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)}{\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)} = 0. \tag{A10}$$

Since $S_{l,k_0} \le 1$, the ratio $\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)/\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)$, which is equal to $\mathbb{E}\left(S_{l,k_0}^n \prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)/\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)$ is bounded by one from above, for every $l > k_0$. Hence,

$$C(l,k_0)\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,l}^{np_j}\right)/\left\{\pi(k_0)\mathbb{E}\left(\prod_{j=1}^{k_0} P_{j,k_0}^{np_j}\right)\right\} \le l^{k_0}/\{k_0!\pi(k_0)\},$$

for every $l > k_0$ and by hypothesis $\sum_{l>k_0} l^{k_0}\pi(l) < \infty$. Therefore, it is possible to apply the dominated convergence theorem to obtain (A1) from (A10) and by Lemma 2 the proof is complete. □

Proof of Proposition 1. Consider first the case of finite a. In this case, $\mathbb{E}\left(\prod_{j=1}^{k} P_{j,l}^{n_j}\right) = \Gamma(la)\prod_{j=1}^{k}\Gamma(n_j+a)/\big(\Gamma(n+la)\Gamma(a)^k\big)$, for every integer $k, l \ge 1$ and every k-tuple $(n_1,\dots,n_k)$. Therefore, the left hand side of (A1) becomes

$$\lim_{n\to\infty}\sum_{l>k_0}\frac{\pi(l)}{\pi(k_0)}C(l,k_0)\frac{\Gamma(la)}{\Gamma(k_0a)}\frac{\Gamma(n+k_0a)}{\Gamma(n+la)}. \tag{A11}$$

As we noted above, for this model, we do not need assumptions about the moments of π, which were useful to ensure the convergence of the series in (A1) dealing with the general case. In fact, the series in (A11) converges for large enough n and for any π, its general term being of order l^{k_0-n} as $l \to \infty$, by Stirling's formula, that is, $\Gamma(x) \sim (2\pi)^{1/2} x^{x-1/2} e^{-x}$, $x \to \infty$.

At this stage, let us prove consistency. To this aim, note that with $c_n(l) = \Gamma(n + k_0 a)/\Gamma(n + la)$ for every $n \geq 1$ and every $l > k_0$, the general term of the series in (A11) depends on n only through $c_n(l)$, which is a nonnegative decreasing sequence since $c_{n+1}(l)/c_n(l) = (ak_0 + n)/(al + n) < 1$, for every $l > k_0$. Therefore, one can apply the monotone convergence theorem.

In order to obtain the convergence rate, note that by (A2), $\pi_n(k_0) \sim 1 - \sum_{l>k_0} b_n(l)$, as $n \to \infty$, where $b_n(l) = \pi(l)C(l, k_0)\Gamma(la)/\{\Gamma(k_0 a)\pi(k_0)\}c_n(l)$, for $l > k_0$. Moreover, since the Gamma function is increasing on $(2, \infty)$, for $n \geq 2$,

$$\sum_{l>k_0+1} \frac{b_n(l)}{b_n(k_0+1)} \leq \sum_{l>k_0+1} \frac{\pi(l)}{\pi(k_0+1)} \frac{l!}{(l-k_0)!(k_0+1)} \frac{\Gamma(la)}{\Gamma\{(k_0+1)a\}} \frac{\Gamma\{n+(k_0+1)a\}}{\Gamma(n+la)},$$

which goes to zero as n diverges, by the monotone convergence theorem. Hence, $\sum_{l>k_0} b_n(l) \sim b_n(k_0+1)$ and therefore, $\pi_n(k_0) \sim 1 - b_n(k_0+1)$, as n diverges, almost surely, that is equal to $1 - c(k_0)\{\Gamma(k_0 a + a)/\Gamma(n + k_0 a + a)\}\{\Gamma(n + k_0 a)/\Gamma(k_0 a)\}$. This implies (5) by Stirling's formula.

At this stage, let $d_n(l) = C(l, k_0)\pi(l)k_0^n/\{\pi(k_0)l^n\}$, for every $n \geq 1$ and every $l > k_0$. If $a = \infty$, then π_n is consistent since $\lim_{n\to\infty} \sum_{l>k_0} d_n(l)$ is zero, by the monotone convergence theorem. In fact, the series converges for large n, since its general term is of order of l^{k_0-n} as l diverges. Therefore, by (A2), $\pi_n(k_0) \sim 1 - \sum_{l>k_0} d_n(l)$. Moreover, $\sum_{l>k_0} d_n(l) \sim d_n(k_0+1)$ as $n \to \infty$, which completes the proof. □

BIBLIOGRAPHY

BARGER, K. & BUNGE, J. (2010). Objective Bayesian estimation for the number of species. *Bayes. Anal.* **5**, 765–785.

BOENDER, C. G. E. & RINNOOY KAN, A. H. G. (1987). A multinomial Bayesian approach to the estimation of population and vocabulary size. *Biometrika* **74**, 849–856.

CHAO, A. & BUNGE, J. (2002). Estimating the number of species in a stochastic abundance model. *Biometrics* **58**, 531–539.

CHAO, A., COLWELL, R. K., LIN, C.-W. & GOTELLI, N. J. (2009). Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* **90**, 1125–1133.

CHAO, A. & LEE, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87**, 210–217.

EFRON, B. & THISTED, R. (1976). Estimating the number of unseen species: how many words did Shakespeare know? *Biometrika* **63**, 435–447.

FAVARO, S., LIJOI, A., MENA, R. H. & PRÜNSTER, I. (2009). Bayesian non-parametric inference for species variety with a two-parameter Poisson-Dirichlet process prior. *J. R. Stat. Soc. Ser. B* **71**, 993–1008.

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

GUT, A. (2005). *Probability: a Graduate Course*. New York: Springer.

HILL, B. M. (1979). Posterior moments of the number of species in a finite population and the posterior probability of finding a new species. *J. Am. Stat. Assoc.* **74**, 668–673.

HSU, L. C. (1948). A theorem on the asymptotic behavior of a multiple integral. *Duke Math. J.* **15**, 623–632.

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

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

ONGARO, A. (2004). Distribuzioni a priori discrete con pesi dirichlet scambiabili. In *Studi in ricordo di Marco Martini*. Milan, Italy: Giuffrè, pp. 293–314. In Italian.

ONGARO, A. (2005). Size–biased sampling and discrete nonparametric Bayesian inference. *J. Statist. Plann. Inference* **128**, 123–148.

ONGARO, A. & CATTANEO, C. (2004). Discrete random probability measures:a general framework for nonparametric Bayesian inference. *Stat. Probabil. Lett.* **67**, 33–45.

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

PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory*, vol. 30 of *IMS Lecture Notes Monogr. Ser.* Hayward, CA: Inst. Math. Statist., pp. 245–267.

PITMAN, J. (2006). *Combinatorial Stochastic Processes*. Berlin: Springer.

PRÜNSTER, I. & LIJOI, A. (2009). Models beyond the Dirichlet process. In *Bayesian Nonparametrics: Principles and Practice*, N. Hjort, C. Holmes, P. Müeller & S. Walker, eds. Cambridge, UK: Cambridge University Press, pp. 80–136.

ROUSSEAU, J. & MENGERSEN, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B* **73**, 689–710.

WANG, J. P. (2010). Estimating species richness by a Poisson–compound gamma model. *Biometrika* **97**, 727–740.

WANG, J. P. & LINDSAY, B. G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *Journal of the American Statistical Association* **100**, 942–959.

ZHANG, H. & STERN, H. (2005). Investigation of a generalized multinomial model for species data. *J. Statist. Comp. Simul.* **75**, 347–362.