

On the topological support of species sampling priors

Pier Giovanni Bissiri* and Andrea Ongaro

Dipartimento di Economia, Metodi Quantitativi e Strategie d'Impresa (DEMS)

Università degli Studi di Milano–Bicocca, Milan, Italy

e-mail: pier.bissiri@unimib.it; andrea.ongaro@unimib.it

Abstract: In Bayesian nonparametric statistics, it is crucial that the support of the prior is very large. Here, we consider species sampling priors. Such priors are widely used within mixture models and it has been shown in the literature that a large support for the mixing prior is essential to ensure the consistency of the posterior. In this paper, simple conditions are given that are necessary and sufficient for the support of a species sampling prior to be full. In particular, for proper species sampling priors, the condition is that the maximum size of the atoms of the corresponding process is small with positive probability. We apply this result to show that the main classes of species sampling priors known in literature have full support under mild conditions. Moreover, we find priors with a very simple construction still having full support.

MSC 2010 subject classifications: Primary 62F15, 62G99.

Keywords and phrases: Bayesian nonparametrics, consistency of posteriors, Dirichlet process, species sampling priors, topological support of measures.

Received January 2013.

Contents

1	Introduction	862
2	General results	863
3	Illustrations	867
3.1	Poisson–Kingman models	867
3.2	Stick–breaking priors	869
3.3	Geometric frequencies	870
3.4	Finite number of species	870
4	Discussion and conclusions	871
	Acknowledgment	873
A	Proof of Theorem 2	873
B	Proof of Proposition 6	878
	References	880

*Corresponding author.

1. Introduction

The main motive of Bayesian nonparametrics is to avoid restrictive parametric assumptions about the distribution generating the data. This is done constructing instead random probability measures whose distributions to be used as priors have a large support (see, for instance, Ghosal [10]).

This paper considers the support (with respect to the weak topology) of the distribution of species sampling processes, which are a relevant class of discrete random probability measures. Species sampling models, which have been developed by Pitman [31], Hansen and Pitman [14], have been extensively studied in Bayesian nonparametric statistical literature. See for instance Ishwaran and James [17], Ongaro [26], Navarrete, Quintana and Mueller [24], Jang, Lee and Lee [19], Hjort et al. [15] and citations provided therein.

So, the large support is an essential requirement for any Bayesian nonparametric model to be appropriate. Moreover, the support of species sampling processes plays a relevant role for the consistency of mixed models. Indeed, these processes are often used as mixing distributions in Bayesian mixture models for density estimation. Ghosal, Ghosh and Ramamoorthi [11] find sufficient conditions for the weak consistency of posteriors for normal mixture models. Their only requirement for the prior of the mixing distribution is to contain the true mixing distribution. More generally, the large support is usually the basic and unique assumption about the mixing prior in the literature about Bayesian consistency of Bayesian mixture models. Such assumption is made, for instance, by Datta [4], who considers a general mixture model and proves the consistency of the predictive distribution and the estimator of the mixing distribution under mild conditions. This is also the case of Wu and Ghosal [39], who obtain sufficient conditions, under which a prior obtained by mixing a wide range of kernels ensures weak consistency of the posterior. The same assumption is made by De Blasi, James and Lau [5], who consider mixed multinomial logit models, and by Bhattacharya and Dunson [1], who study consistency for a more general mixture models.

In spite of the amount of literature regarding species sampling priors, their support has not generally received much attention. To the best of our knowledge, this generally is the case of the most known priors in such class, apart from the Dirichlet prior. Furthermore, one might wonder if it is possible to consider other discrete random probability measures with a much simpler structure, still having a prior with large support.

The most well-known species sampling prior is the Dirichlet prior introduced and studied by Ferguson [8]. Ferguson [8] states that the topological support (with respect to the topology of weak convergence) of a Dirichlet prior is the set of all probability measures whose topological supports are contained in that one of the parameter measure of the Dirichlet prior, which is the marginal distribution of one observation. Later on, it will be shown that the support of a prior with the same marginal cannot be larger. For this reason, if a prior has the same support of the Dirichlet prior with the same marginal, then we can say that that prior has full support. Indeed, the support of the Dirichlet prior

contains essentially all probability measures on the relevant sample space, as it will be clear later on.

Ferguson's [8] statement about the support of the Dirichlet prior on the real line has been extended to broader classes of prior distributions. Majumdar [23] generalizes it replacing the real line with an arbitrary separable metric space. Ongaro and Cattaneo [27] find sufficient conditions for species sampling priors to have full support. The present paper provides two much weaker and simple conditions that are sufficient and necessary for such result. In particular, for proper species sampling priors, one such condition is that the maximum size of the atoms of the corresponding process is small with positive probability. We shall also show that if such condition is not satisfied, then the support is remarkably poor, not including for instance continuous probability measures.

Verifying this condition, it will be proved that the support is full under mild conditions for the most relevant species sampling priors, such as the stick-breaking priors and the Poisson–Kingman priors, which in turn include the normalized random measures with independent increments, the Pitman–Yor process priors and the Gibbs–type priors with infinitely many species. In particular, for the Poisson–Kingman model, this condition cannot be verified directly as there is not a closed expression for the distribution of the maximum atom size. A proof is given based on specific properties of non-homogeneous Poisson processes.

By means of our results, we shall find priors with full support that have a very simple construction and are expected to be easier to implement than the more complex Bayesian nonparametric models. This is the direction pointed by Fuentes-García, Mena and Walker [9], who consider the species sampling prior with geometric weights. Namely, they assess the random sizes of the atoms as $P_j = W(1 - W)^{j-1}$ ($j \geq 1$). They use such prior for density estimation showing that its inferential performance is at least as good as the usual nonparametric models, the estimation algorithm being more efficient. We prove that for this prior to have full support, the only condition is that zero belongs to the support of the random variable W . Another simple way to construct a species sampling prior with full support consists of taking $P_j = 1/K$, for $j = 1, \dots, K$, where K is an unbounded positive random variable valued into the natural numbers.

The outline of the paper is the following. Section 2 presents the general results, Section 3 shows that the support is full for different priors, including the most important nonparametric priors among the species sampling models. The Appendix contains the proof of Theorem 2 and Proposition 6.

2. General results

Let \mathbb{X} be separable metrizable space and \mathcal{X} its Borel σ -field. Let \mathcal{P} be the set of all probability measures on $(\mathbb{X}, \mathcal{X})$. Recall that \mathcal{P} equipped with the topology of weak convergence is metrizable as a separable metric space [see 28] and let $\mathcal{B}(\mathcal{P})$ be its Borel σ -field. Moreover, for every μ in \mathcal{P} , an open base of neighborhoods

of μ for the topology of weak convergence on \mathcal{P} is the class of sets

$$U_{\varepsilon_1, \dots, \varepsilon_k}(\mu; A_1, \dots, A_k) := \bigcap_{j=1}^k \{\nu \in \mathcal{P} : |\nu(A_j) - \mu(A_j)| < \varepsilon_j\},$$

where k is a positive integers, $(\varepsilon_1, \dots, \varepsilon_k)$ belongs to $(0, \infty)^k$, and (A_1, \dots, A_k) is a k -tuple of μ -continuity sets in \mathcal{X} . Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which all the random variables considered through the present paper are defined. A random probability measure is a measurable map from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$.

Recall that the topological support S_μ of a finite Borel measure μ on a second countable topological space is defined as the intersection of all closed set F such that $\mu(F^c) = 0$. As usual, the support of a random variable is defined to be the support of its probability distribution. It is known that s belongs to S_μ if and only if μ is positive on every neighborhood of s . So, a probability measure μ in \mathcal{P} belongs to the support of a random probability measure P if and only if

$$\mathbb{P}(P \in U_{\varepsilon_1, \dots, \varepsilon_k}(\mu; A_1, \dots, A_k)) > 0,$$

for every k -tuple (A_1, \dots, A_k) of μ -continuity subsets of \mathcal{X} and every integer $k \geq 1$.

Majumdar [23] proves that the topological support of the Dirichlet prior with parameter α is the class of all probability measures μ such that $S_\mu \subset S_\alpha$.

Species sampling processes are discrete random probability measures that generalize the Dirichlet Process. A species sampling process is a random discrete probability measure of the following form:

$$\sum_{j=1}^{\infty} P_j \delta_{Z_j} + \left(1 - \sum_{j=1}^{\infty} P_j\right) \alpha, \quad (1)$$

where α is a probability measure on \mathcal{X} which is diffuse (that is $\alpha\{x\} = 0$ for every x in \mathbb{X}), $(P_j)_{j \geq 1}$ and $(Z_j)_{j \geq 1}$ are two independent sequences of random variables such that $P_j \geq 0$ for every $j \geq 1$ and $\sum_{j=1}^{\infty} P_j \leq 1$, \mathbb{P} -a.s., and $(Z_j)_{j \geq 1}$ is an i.i.d. sequence whose common distribution is α .

If $\sum_{j=1}^{\infty} P_j = 1$, \mathbb{P} -a.s., then (1) becomes

$$\sum_{j=1}^{\infty} P_j \delta_{Z_j}. \quad (2)$$

To be consistent with Pitman's terminology, here a species sampling process that admits a representation of the form (2) is called *proper*.

The Dirichlet Process is a proper species sampling process as shown by Ferguson [8] and Sethuraman [36]. The support of a Dirichlet prior is essentially as big as possible, as shown in the next Proposition.

Proposition 1. *Let Q be a random probability measure on $(\mathbb{X}, \mathcal{X})$ such that $\alpha(A) = \mathbb{E}(Q(A))$, for every A in \mathcal{X} . If μ is a probability measure on $(\mathbb{X}, \mathcal{X})$ belonging to the support (in the weak topology) of the distribution of Q , then $S_\mu \subset S_\alpha$.*

Proof. We need to show that the set $C := \{\mu \in \mathcal{P} : S_\mu \subset S_\alpha\}$ is closed and that Q belongs to C almost surely. To prove the second claim, consider that $Q(S_\alpha^c) \geq 0$ a.s. and $0 = \alpha(S_\alpha^c) = \mathbb{E}(Q(S_\alpha^c))$, where \mathbb{E} denotes the expectation w.r.t. \mathbb{P} . Hence, $Q(S_\alpha^c) = 0$ a.s. and therefore, $S_Q \subset S_\alpha$ a.s.

To prove closeness, we assume that $(Q_n)_{n \geq 1}$ is a sequence of elements of C such that Q_n weakly converges to some Q_∞ in \mathcal{P} and we show that Q_∞ belongs to C . To this aim, note that $Q_n(S_\alpha)$ is equal to one a.s. since Q_n belongs to C , for every $n \geq 1$. Moreover, being S_α a closed set, we can apply the Portmanteau theorem to obtain that

$$Q_\infty(S_\alpha) \geq \limsup_{n \rightarrow \infty} Q_n(S_\alpha) = 1.$$

Hence, $Q_\infty(S_\alpha)$ is equal to one a.s., i.e. $S_{Q_\infty} \subset S_\alpha$, and the proof is complete. \square

Proposition 1 leads to the following definition:

Definition. Let Π be a prior distribution on \mathcal{P} , let P be a random probability measure with distribution Π and let α the measure on $(\mathbb{X}, \mathcal{X})$, such that $\alpha(A) = \mathbb{E}(P(A))$ for every A in \mathcal{X} . Π (or equivalently P) is said to have *full support* if and only if the support (in the weak topology) of Π is the set of all measures μ on $(\mathbb{X}, \mathcal{X})$ such that $S_\mu \subset S_\alpha$.

To make clearer the sense of the previous definition, let X be an observation from P , i.e. P is the conditional distribution of X given P . Then, \mathbb{X} is the sample space for X and, in the most natural case, \mathbb{X} is also the support of the marginal distribution α of X . In this case, the support of P is full if it coincides with the whole class of probability measures on \mathbb{X} .

Our aim is to find necessary and sufficient conditions for a species sampling prior to have full support. Ongaro and Cattaneo [27] have essentially shown that a sufficient condition for a proper species sampling prior is that for every n , the support of the law of (P_1, \dots, P_n) is full, i.e. is equal to the n -dimensional simplex

$$\{(p_1, \dots, p_n) : \sum_{j=1}^n p_j \leq 1, p_j \geq 0, 1 \leq j \leq n\}.$$

The next theorem identifies two much weaker conditions that are necessary and sufficient for the distribution of a species sampling process to have full support.

Theorem 2. Consider a prior distribution Π on \mathcal{P} that is the distribution of the process (1).

The following facts are equivalent:

- i. The prior Π has full support,
- ii. For every $\varepsilon > 0$, $\mathbb{P}(\max_{j \geq 1} \bar{P}_j < \varepsilon, \sum_{l \geq 1} P_l > 1 - \varepsilon) > 0$, where $\bar{P}_j = P_j / \sum_{l \geq 1} P_l$ ($j \geq 1$),
- iii. For every $\varepsilon > 0$ there is an integer $m \geq 1$ such that

$$\mathbb{P}(P_j < \varepsilon, j = 1, \dots, m, \sum_{l=1}^m P_l > 1 - \varepsilon) > 0. \tag{3}$$

Clearly, condition iii. improves the sufficient condition given by Ongaro and Cattaneo [27]. Moreover, it is a convenient alternative when it is difficult to deal with the distribution of $\max_{j \geq 1} P_j$. Theorem 2, which is proved in the Appendix, yields the following corollary:

Corollary 3. *If $\sum_{j \geq 1} P_j = 1$, \mathbb{P} -a.s. then the corresponding (proper) species sampling prior has full support if and only if*

$$\mathbb{P}\left(\max_{j \geq 1} P_j < \varepsilon\right) > 0, \quad (4)$$

for every $\varepsilon > 0$.

In other words, for a proper species sampling prior to have full support is necessary and sufficient that zero belongs to the support of the maximum atom size. It will be now shown that if such condition is not satisfied then the support of the prior, besides being not full, is substantially smaller. Indeed, if this is the case, then neither continuous distributions nor discrete distributions whose atoms have too small sizes belong to the support of the prior, as shown in the following proposition:

Proposition 4. *Given a proper species sampling prior, if (4) is not satisfied for some $\varepsilon > 0$ and μ is a probability measure on \mathbb{X} such that $\mu(\{x\}) < \varepsilon$, for every $x \in \mathbb{X}$, then μ does not belong to the support of the prior.*

Proof. To begin with, recall that, for every μ in the space \mathcal{P} of probability measures on $(\mathbb{X}, \mathcal{X})$, a base of neighborhoods of μ generating the topology of weak convergence is made of the sets of the form

$$\{\nu \in \mathcal{P} : \nu(F_i) < \mu(F_i) + \varepsilon, i = 1, \dots, m\},$$

where $\varepsilon > 0$ and $\{F_1, \dots, F_m\}$ is an m -tuple of measurable closed subsets of \mathbb{X} [see 2, p. 236]. Therefore, to prove the statement we need to find a set of such form with zero prior probability under the assumption that $\mu(\{x\}) < \varepsilon$, for every $x \in \mathbb{X}$.

To this aim, let B the union of a countable dense subset of \mathbb{X} , which exists by separability of \mathbb{X} , and the set of the atoms of μ . So, B is countable and we denote by $(x_j)_{j \geq 1}$ an arrangement of the elements of B . Now, let d be a metric on \mathbb{X} which induces the Borel sigma-field \mathcal{X} and set $\bar{B}_j = \{x \in \mathbb{X} : d(x, x_j) \leq r_j\}$ where r_j is small enough to ensure that $\mu(\bar{B}_j) < \varepsilon$, for $j \geq 1$. Such r_j exists since $\bigcap_{n \geq 1} \{x \in \mathbb{X} : d(x, x_j) \leq 1/n\} = \{x_j\}$ and $\mu\{x_j\} < \varepsilon$, for every $j \geq 1$. Denote by B_j the open ball $\{x \in \mathbb{X} : d(x, x_j) < r_j\}$, for every $j \geq 1$. Clearly, the sets in the sequence $(B_j)_{j \geq 1}$ cover the space \mathbb{X} . In fact, B contains a dense subset of \mathbb{X} , say D , and therefore for every $r > 0$ and every x in \mathbb{X} , $d(x, x_j) < r$ for some x_j in D . So, $\lim_{m \rightarrow \infty} \mu(\bigcup_{j=1}^m B_j) = 1$, and therefore there is an integer $k \geq 1$ such that $\mu(\bigcup_{j=1}^k B_j) > 1 - \varepsilon$. Set $\bar{B}_0 = \mathbb{X} \setminus (\bigcup_{j=1}^k B_j)$ and $\eta = \max_{0 \leq j \leq k} \mu(\bar{B}_j) < \varepsilon$. Hence, the $(k+1)$ -tuple $\{\bar{B}_0, \dots, \bar{B}_k\}$ is a measurable cover of \mathbb{X} , where \bar{B}_j is closed and $\mu(\bar{B}_j) \leq \eta$, for $j = 0, \dots, k$. Therefore,

$$\begin{aligned} 0 &= \mathbb{P}(P_l \leq \varepsilon, l \geq 1) \geq \mathbb{P}(\sum_{l \geq 1: Z_l \in \bar{B}_j} P_l < \varepsilon, 0 \leq j \leq k) \\ &= \mathbb{P}(P(\bar{B}_j) < \varepsilon, 0 \leq j \leq k) \geq \mathbb{P}(P(\bar{B}_j) < \mu(\bar{B}_j) + \varepsilon - \eta, 0 \leq j \leq k). \end{aligned}$$

So, there is a neighborhood of μ with zero \mathbb{P} -probability and the proof is complete. \square

We now state the following proposition, which shows a connection between proper and not proper species sampling priors related to their supports.

Proposition 5. *Let P be a species sampling process defined by (1) and let $Q = \sum_{j \geq 1} \bar{P}_j \delta_{z_j}$, where $\bar{P}_j = P_j / \sum_{l \geq 1} P_l$, for every $j \geq 1$, if $\sum_{l \geq 1} P_l > 0$.*

If $\mathbb{P}(\sum_{j \geq 1} P_j > w) > 0$ for every $w < 1$ and the conditional distribution of Q given $\sum_{l \geq 1} P_l$ has full support almost surely, then the distribution of P has full support, too.

We do not report the proof of this proposition, which is an immediate consequence of Lemma 10 in the Appendix. Proposition 5 suggests a simple way to construct a not-proper species sampling prior with full support from a proper species sampling prior with full support. In fact, if Q is a proper species sampling process, its prior has full support and its marginal is α , then one can take $WQ + (1 - W)\alpha$, where W is a $(0, 1)$ -valued random variable with 1 in its support and Q and W are independent.

3. Illustrations

We now apply Theorem 2 and Corollary 3 to show that several species sampling priors have full support. We shall consider the two most general and studied priors among the species sampling models, namely the Poisson–Kingman and the stick-breaking model, but also two priors with a very simple construction, one based on geometric frequencies and the other one on a finite number of species.

3.1. Poisson–Kingman models

The class of Poisson–Kingman priors was introduced and studied by Pitman [32] and it includes as special cases another known class, such as the homogeneous normalized random measures with independent increments [see 34]. Moreover, Poisson–Kingman models are related to the Gibbs-type random probability measures with infinitely many species [see 13, 22, 7]. Normalized random measures with independent increments and Gibbs-type priors are two of the most studied and used for applications.

To properly define the class of Poisson–Kingman priors, let us introduce some notation. Let ρ be a measure on \mathbb{R}^+ such that

$$\rho(1, \infty) < \infty, \quad \int_{(0,1)} s\rho(ds) < \infty, \tag{5}$$

and

$$\rho(0, \infty) = \infty. \tag{6}$$

Moreover, let α be a probability measure on \mathbb{X} , let $(T_j)_{j \geq 1}$ be the decreasing arrangement of the points of a Poisson process Π with mean measure ρ , and let $(Z_j)_{j \geq 1}$ be a sequence of independent and identically distributed random variables with common distribution α . Set $S = \sum_{j \geq 1} T_j$ and assume that the probability distribution of S is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^+ . Denote by $Q_{\rho,s}$ the regular conditional distribution of the sequence $(T_j/S)_{j \geq 1}$ given $S = s$ as constructed by Pitman [32]. Let $(P_j)_{j \geq 1}$ be a sequence of nonnegative random variables with distribution

$$\int_{\mathbb{R}^+} Q_{\rho,s} \mu(ds), \quad (7)$$

where μ is a probability measure on \mathbb{R}^+ . Then (7) is termed a Poisson–Kingman distribution with Lévy intensity ρ and mixing distribution μ and denoted by $\text{PK}(\rho, \mu)$. Moreover, the random probability measure $\sum_{j=1}^{\infty} P_j \delta_{Z_j}$ is termed a Poisson–Kingman random probability measure with Lévy intensity ρ and mixing distribution μ .

Condition (5) is required to ensure that S is finite almost surely [see 20, p. 28], and (6) to ensure that S is positive almost surely. To show this last statement, one can just apply Campbell’s theorem to obtain that

$$\mathbb{E}(e^{-uS}) = \exp\{-\int_{(0,\infty)}(1 - e^{-ux})\rho(dx)\}$$

for every $u > 0$, and then apply the monotone convergence theorem, letting u diverge to infinite, to obtain that $\mathbb{P}(S = 0) = e^{-\rho(0,\infty)}$.

An important class of priors related to the Poisson–Kingman priors is that one of Gibbs–type priors with infinitely many species. In fact, as proved by Gnedin and Pitman [13], a Gibbs–type prior with infinitely many positive frequencies is either a mixture of Dirichlet priors or a Poisson–Kingman prior with stable Lévy density $\rho_{\sigma}(x) = \sigma x^{-\sigma-1}/(\Gamma(1-\sigma))$, for some $0 < \sigma < 1$, and arbitrary mixing distribution μ . Another relevant prior belonging to the Poisson–Kingman class is the Pitman–Yor process prior, also known as two parameter Poisson–Dirichlet prior, introduced by Perman, Pitman and Yor [29] and further studied by Pitman [30], Pitman and Yor [33] and Ishwaran and James [17]. In fact, a Pitman–Yor process prior with parameters $0 < \sigma < 1$ and $\theta > -\sigma$ is a $\text{PK}(\rho_{\sigma}, \mu_{\sigma,\theta})$, and therefore a Gibbs–type prior, where $\mu_{\sigma,\theta} dt = \sigma \Gamma(\theta)/\Gamma(\theta/\sigma) t^{-\theta} f_{\sigma}(t) dt$, and f_{σ} is a density of a σ –stable random variable. The Pitman–Yor process prior is widely used in applications. See, for instance, Teh and Jordan [38] and citations provided therein.

We are now ready to state and prove the following proposition:

Proposition 6. *If μ is absolutely continuous with respect to the Lebesgue measure, then the $\text{PK}(\rho, \mu)$ prior has full support.*

By Proposition 6, the whole class of NRMII’s and Pitman–Yor process priors have full support. In fact, if μ coincides with the probability distribution of S , then the corresponding Poisson–Kingman random probability measure is

a normalized random measure with independent increments. To prove Proposition 6, it is convenient to split the half line of positive real numbers into two intervals and to consider separately the restriction of Π to each interval. By an asymptotic argument, it is possible to deal with the infinitely many jumps around zero, letting the interval containing zero decrease to the singleton $\{0\}$. Then, one can deal with the interval which does not contain zero converting the Poisson process Π into a Bernoulli process by conditioning. The details of the proof of Proposition 6 are deferred to the Appendix.

3.2. Stick-breaking priors

The class of stick-breaking priors is a relevant one in Bayesian nonparametrics [see 16, 17, 18, 6]. A stick-breaking prior is such that $P_1 = V_1$ and $P_j = V_j \prod_{l=1}^{j-1} (1 - V_l)$, $j > 1$, being $(V_j)_{j \geq 1}$ a sequence of $[0, 1]$ -valued random variables. In this way,

$$\sum_{l=1}^m P_l = 1 - \prod_{j=1}^m (1 - V_j), \tag{8}$$

for every integer $m \geq 1$, and therefore $\sum_{j=1}^\infty P_j = 1 - \prod_{j=1}^\infty (1 - V_j)$ is between zero and one. Stick breaking priors have full support under minimal assumptions as shown in the next proposition.

Proposition 7. *If $(V_j)_{j \geq 1}$ is a sequence of independent random variables and for every $\varepsilon > 0$ there is $\delta > 0$ such that $\mathbb{P}(\delta < V_j < \varepsilon)$ is positive for every integer $j \geq 1$, then the corresponding stick-breaking prior has full support.*

Proof. Fix $0 < \varepsilon < 1$ and $0 < w < 1$. Being $\mathbb{P}(\delta < V_j < \varepsilon) > 0$ for every $j \geq 1$ and recalling the identity (8), one obtains that

$$\begin{aligned} 0 < \prod_{j=1}^m \mathbb{P}(\delta < V_j < \varepsilon) &= \mathbb{P}(\delta < V_1 < \varepsilon, \delta < V_2 < \varepsilon, \dots, \delta < V_m < \varepsilon) \\ &\leq \mathbb{P}(P_1 < \varepsilon, \dots, P_m < \varepsilon(1 - \delta)^{m-1}, \sum_{l=1}^m P_l > 1 - (1 - \delta)^m) \\ &\leq \mathbb{P}(P_1 < \varepsilon, \dots, P_m < \varepsilon, \sum_{l=1}^m P_l > 1 - (1 - \delta)^m), \end{aligned} \tag{9}$$

for every integer $m \geq 1$. At this stage, fix m bigger than $\log(\varepsilon)/\log(1 - \delta)$ so that $(1 - \delta)^m < \varepsilon$. In this way, (9) implies (3), and condition iii. of Theorem 2 is therefore satisfied. \square

Corollary 8. *If $(V_j)_{j \geq 1}$ is a sequence of independent random variables and for some $\varepsilon > 0$, the support of each V_j includes the interval $(0, \varepsilon)$, then the corresponding stick-breaking prior has full support.*

Generally, for each $j \geq 1$, V_j has a Beta distribution with parameters a_j and b_j , where $(a_j)_{j \geq 1}$ and $(b_j)_{j \geq 1}$ are two sequences of positive numbers [see for instance 16]. Clearly, in this case the support of each V_j is the whole unit interval and Corollary 8 can be applied.

The sequence $(V_j)_{j \geq 1}$ is often taken i.i.d., which yields, by the way, a proper species sampling model. This is the most common case, which is considered in the following corollary:

Corollary 9. *Let $(V_j)_{j \geq 1}$ be a sequence of independent and identically distributed random variables. Then, $\mathbb{P}(0 < V_1 < \varepsilon)$ is positive for every $\varepsilon > 0$ if and only if the corresponding stick-breaking prior has full support.*

Proof. To prove the “only if” part, consider that $\lim_{c \rightarrow 0} \mathbb{P}(c < V_1 < \varepsilon) = \mathbb{P}(0 < V_1 < \varepsilon) > 0$. This implies that $\mathbb{P}(\delta < V_1 < \varepsilon)$ is positive for some $\delta > 0$ and then it is sufficient to apply Proposition 7. To prove the “if” part, we can just show that $\mathbb{P}(V_1 \in \{0\} \cup (c, \infty)) = 1$ implies that the support of the corresponding stick-breaking prior is not full. To this aim, denote by ϕ the map from $[0, 1]^\infty$ into itself such that $\phi((v_j)_{j \geq 1})$ is the sequence obtained from $(v_j)_{j \geq 1}$ removing the zeroes and let $(W_j)_{j \geq 1} = \phi((V_j)_{j \geq 1})$. Clearly, substituting $(V_j)_{j \geq 1}$ with $(W_j)_{j \geq 1}$, one obtains the same stick-breaking prior. Moreover, $\mathbb{P}(W_1 > c) = 1$. Being $W_1 = P_j$ for some $j \geq 1$ a.s., we have that $P_j > c$ for some $j \geq 1$ a.s. By Corollary 3, this implies that the corresponding stick-breaking prior does not have full support. \square

3.3. Geometric frequencies

Let us now introduce an example of a proper species sampling prior with a very simple structure still having the full support. Let

$$P_j = W(1 - W)^{j-1} \tag{10}$$

for $j \geq 1$, where W is a random variable such that $0 < W < 1$ with probability one and zero belongs its support. This prior has been proposed and used by Fuentes-García, Mena and Walker [9], letting W having a Beta distribution. To apply Corollary 3 it is sufficient to note that $(P_j)_{j \geq 1}$ is a decreasing sequence almost surely and $\mathbb{P}(0 < W < \varepsilon) > 0$ for every $\varepsilon > 0$. For this prior, all the weights P_j , $j \geq 1$, are obtained from a real valued random variable W and the support of (P_1, \dots, P_n) is not full, for $n \geq 1$, i.e. the condition for the full support given by Ongaro and Cattaneo [27] is not satisfied. As shown by Fuentes-García, Mena and Walker [9], such model is easier to implement than standard models and performs well at least in the density estimation context. In the following subsection, other examples are given of priors with a simple structure and full support, whose inferential behaviour are worth exploring.

3.4. Finite number of species

Let us consider a finite number of species, in other words a proper species sampling process (2) such that the number of P_j 's which are positive is finite. Define the random variable $K = \max\{j : P_j > 0\}$ with the convention that $K = \infty$ if P_j is infinitely often positive. When K is finite, (2) can be equivalently and more conveniently expressed as

$$\sum_{j=1}^K P_{j,K} \delta_{Z_j}, \tag{11}$$

where $(P_{1,k}, \dots, P_{k,k})$ is a random vector whose distribution coincides with the conditional distribution of (P_1, \dots, P_K) given $K = k$, for every positive integer k . So, $(P_{1,k}, \dots, P_{k-1,k})$ takes values into the $(k - 1)$ -dimensional simplex, for every $k \geq 1$. This formulation is the one used for instance by Ongaro [26]. It is reasonable that K is finite almost surely, i.e. the number of species is finite. This model is used within Bayesian nonparametric mixture models Richardson and Green [35]. See also Stephens [37] and Nobile and Fearnside [25]. Moreover, a model of this kind has been studied by Gnedin [12]. Priors which arise from processes of the form (11) are considered by Bissiri, Ongaro and Walker [3]. This family of priors includes the Gibbs-type priors with finitely many species, which are obtained if $(P_{1,k}, \dots, P_{k-1,k})$ has a symmetric finite-dimensional Dirichlet distribution. This is proved by Gnedin and Pitman [13].

It is not hard to verify that this prior has full support if and only if for every $\varepsilon > 0$, there is a positive integer k such that $\mathbb{P}(K = k) > 0$ and $\mathbb{P}(P_{j,k} < \varepsilon, 1 \leq j \leq k \mid K = k) > 0$. In fact,

$$\mathbb{P}\left(\max_{j \geq 1} P_j < \varepsilon\right) = \sum_{k=1}^{\infty} \mathbb{P}(K = k) \cdot \mathbb{P}(P_{j,k} < \varepsilon, 1 \leq j \leq k \mid K = k)$$

is positive if and only if at least one term of the sum is positive and one can just apply Corollary 3. Therefore, being the support of the Dirichlet distribution the whole simplex, Gibbs-type priors have full support provided that K is unbounded. A remarkably simple example where the prior has full support is obtained taking $P_{j,K} = 1/K$ where K is a random variable such that $\mathbb{P}(K > k) > 0$ for every positive integer k .

4. Discussion and conclusions

A large support is clearly an essential requirement for any Bayesian nonparametric model to be appropriate. In particular, it is an important requirement for species sampling priors that are used as mixing distributions in Bayesian mixtures models for density estimation to ensure consistency for such models. For instance, if the support (in the weak topology) of the mixing distribution is full and the true density to be estimated satisfies some mild conditions, then consistency is satisfied for a wide range of mixing location scale kernels (see Wu and Ghosal [39]).

This paper provides necessary and sufficient conditions for a species sampling prior to have full support (Theorem 2). In particular, for proper species sampling priors, it turns out to be a very simple condition (Corollary 3). These results can be applied to show that the most relevant classes of species sampling priors have full support, but they are also useful to construct species sampling priors with a much simpler structure still having full support. This is the case of the prior based on geometric weights (10), which has full support as shown in Section 3.3 and was used by Fuentes-García, Mena and Walker [9] within a Bayesian mixed model for density estimation. Namely, the prior for the unknown density is

assessed as the distribution of the following random density function:

$$\begin{aligned} f(x) &= \int h(x, \theta) P(d\theta) \\ &= \sum_{j=1}^{\infty} P_j h(x, Z_j), \end{aligned}$$

where P is a proper species sampling process with geometric weights (10), the distribution of W is Beta, and $h(\cdot, z)$ is a density function for each z .

The simple structure of this prior for P sets it apart from the priors mostly studied for Bayesian nonparametric mixture modeling, which are generalizations of the Dirichlet process model and in most cases yield complex models hard to implement and to apply in real situations. See Lijoi, Mena and Prünster [21] for an example of how complex these generalizations can be. Fuentes-García, Mena and Walker [9] propose a relatively easy Gibbs sampler algorithm for the model based on the prior with geometric weights, which results in a simpler alternative to those typically used for Bayesian nonparametric mixtures, but with similar inferential performance. To this aim, they consider the following alternative representation for f :

$$f(x) = \sum_{m=1}^{\infty} \frac{1}{m} \sum_{j=1}^m h(x, Z_j) q_W(m),$$

where $q_W(m) = mW^2(1-W)^{m-1}$ for $m \geq 1$. This representation suggests to write the model in hierarchical form:

- W is Beta distributed;
- the Z_j 's are i.i.d. with a given density function and are independent of W ;
- K_1, \dots, K_n are conditionally i.i.d. given W with common probability mass function q_W and are independent of the Z_j 's;
- the observations X_1, \dots, X_n are conditionally i.i.d. given K_1, \dots, K_n, W and the Z_j 's, and the conditional density of X_i is

$$\frac{1}{K_i} \sum_{j=1}^{K_i} h(\cdot, Z_j),$$

for $i = 1, \dots, n$.

Fuentes-García, Mena and Walker [9] derive the full conditionals of the random variables defined in each level of the hierarchy to construct a Gibbs sampler for the model.

In Section 3.4, another prior is considered with an even simpler structure, namely with finitely many species, and equal weights $P_{j,K} = 1/K$, for $j = 1, \dots, K$. Indeed, this is probably the simplest Bayesian nonparametric model. When it is used within mixed models, the random variables K_1, \dots, K_n in the

above representation are replaced by a single one, that is K . In this case, the random density takes the form:

$$f(x) = \frac{1}{K} \sum_{j=1}^K h(x, Z_j).$$

This model resembles the form of the basic classical kernel density estimator. In spite of its simplicity, for an appropriate choice of the class of densities h , it approximates a large variety of densities, ensuring posterior consistency within mixed models. To our knowledge, this model has not been studied for Bayesian mixture modeling and it would deserve further investigation.

In general, we believe that for many applicative purposes, such as density estimation, a sophisticated modelization for the weights does not always provide a real gain in terms of inferential performance, implying on the other hand an heavy computational cost.

Acknowledgment

The authors are grateful to Stephen G. Walker, who inspired this paper. This work was partially supported by ESF and Regione Lombardia (by the grant ‘‘Dote Ricercatori’’).

Appendix A: Proof of Theorem 2

In order to prove Theorem 2, the three following lemmas are useful.

Lemma 10. *Let P be a species sampling process defined by (1) and let $Q = \sum_{j \geq 1} \bar{P}_j \delta_{Z_j}$, where $\bar{P}_j = P_j / \sum_{l \geq 1} P_l$, for every $j \geq 1$, if $\sum_{l \geq 1} P_l > 0$. If μ is a probability measure on $(\mathbb{X}, \mathcal{X})$ such that*

$$\mathbb{P}(Q \in U_{\delta_1, \dots, \delta_k}(\mu; A_1, \dots, A_k), \sum_{l \geq 1} P_l > w) > 0, \tag{12}$$

for every k -tuple (A_1, \dots, A_k) of μ -continuity subsets of \mathcal{X} and every $(\delta_1, \dots, \delta_k, w)$ in $(0, 1)^{k+1}$, then μ belongs to the support of P .

Proof. Clearly, $P = \sum_{l \geq 1} P_l \cdot Q + (1 - \sum_{l \geq 1} P_l) \cdot \alpha$. So, if $A \in \mathcal{X}$, $w > 1 - \varepsilon / |\alpha(A) - \mu(A)|$ and $|Q(A) - \mu(A)| < \varepsilon / w - |\alpha(A) - \mu(A)|(1/w - 1)$, then $|P(A) - \mu(A)| < \varepsilon$ (\mathbb{P} -a.s.). Therefore, for every integer $k \geq 1$, every k -tuple (A_1, \dots, A_k) of μ -continuity subsets of \mathcal{X} and every $(\varepsilon_1, \dots, \varepsilon_k) \in (0, \infty)^k$,

$$\begin{aligned} & \mathbb{P}(P \in U_{\varepsilon_1, \dots, \varepsilon_k}(\mu; A_1, \dots, A_k)) \\ & \geq \mathbb{P}(Q \in U_{\delta_1, \dots, \delta_k}(\mu; A_1, \dots, A_k), \sum_{l \geq 1} P_l > w) > 0 \end{aligned}$$

if $w > \max\{0, 1 - \varepsilon_j/2\}$ (for every $1 \leq j \leq k$), and $\delta_j = \varepsilon_j/w - |\alpha(A_j) - \mu(A_j)|(1/w - 1)$ ($1 \leq j \leq k$), and the proof is complete. \square

Lemma 11. *Let k be a positive integer and let (p_1, \dots, p_k) be a k -tuple of non-negative real numbers that sum up to one.*

For every $\varepsilon > 0$, there is $\varepsilon > 0$ such that

$$\left\{ \sum_{h=1}^{\infty} P_h = 1, 0 \leq P_l < \varepsilon : l \geq 1 \right\} \subset \bigcup_{n_1 \geq 1, \dots, n_k \geq 1} \bigcap_{j=1}^k \left\{ \left| p_j - \sum_{h=1}^{n_j} P_{m_j+h} \right| < \frac{\varepsilon}{k+1} \right\}, \tag{13}$$

where $m_1 = 0$ and $m_j = \sum_{l=1}^{j-1} n_l$ for $2 \leq j \leq k+1$.

Proof. Fix $\varepsilon \leq \min\{\varepsilon/[(k-1)(k+1)], p_l - \varepsilon/(k+1) : 1 \leq l \leq k\}$ and without loss of generality let

$$0 < \varepsilon < \min\{p_j : 1 \leq j \leq k\}. \tag{14}$$

To prove the statement, fix $\omega \in \Omega$ such that $0 \leq P_h(\omega) < \varepsilon$ (for every $h \geq 1$) and $\sum_{h=1}^{\infty} P_h(\omega) = 1$. In the rest of the proof, let P_h stand for $P_h(\omega)$ for $h \geq 1$. We shall prove that for some k -tuple (n_1, \dots, n_k) of positive integers and every $1 \leq j \leq k$,

$$p_j - \varepsilon/(k+1) < \sum_{h=1}^{n_j} P_{m_j+h} < p_j + \varepsilon/(k+1). \tag{15}$$

To this aim, we shall show by induction that

$$p_j - \varepsilon/[2(k-1)(k+1)] < \sum_{h=1}^{n_j} P_{m_j+h} < p_j + \varepsilon/[2(k-1)(k+1)], \tag{16}$$

for some k -tuple (n_1, \dots, n_k) of positive integers and every $1 \leq j \leq k-1$.

For convenience, set $\delta = \varepsilon/[2(k-1)(k+1)]$. Let us start with the case $j = 1$. Being $\sum_{h=1}^{\infty} P_h = 1$, there is an integer $l \geq 1$ such that $\sum_{h=1}^l P_h > p_1 - \delta$. Therefore, we can set

$$n_1 = \min\{l \geq 1 : \sum_{h=1}^l P_h > p_1 - \varepsilon/(k+1)\}.$$

Being $P_1 < \varepsilon \leq p_1 - \varepsilon/(k+1) < p_1 - \delta$, one can see that $n_1 \geq 2$. Clearly, $\sum_{h=1}^{n_1-1} P_h \leq p_1 - \delta$, and being $P_{n_1} < \varepsilon < \varepsilon/[(k+1)(k-1)] = 2\delta$, we obtain that $\sum_{h=1}^{n_1} P_h < p_1 + \delta$. Hence, (16) is true for $j = 1$.

At this stage, assume that (16) is true for j ($1 \leq j \leq k-1$) and let us show that it holds when j is replaced by $j+1$. By summation, (16) yields:

$$\sum_{h=1}^{m_{j+1}} P_h = \sum_{l=1}^j \sum_{h=1}^{n_l} P_{m_l+h} < \sum_{l=1}^j p_l + j\delta. \tag{17}$$

Being $\sum_{h=1}^{\infty} P_h = 1$ and being $\sum_{l=1}^{j+1} p_l + (j+1)\delta < 1$ by (14), there is an integer $l > 1$ such that

$$\sum_{h=1}^l P_h > \sum_{l=1}^{j+1} p_l + (j+1)\delta. \tag{18}$$

Combination of (17) with (18) yields that $\sum_{h=1}^l P_{m_{j+1}+h} = \sum_{h=1}^l P_h - \sum_{h=1}^{m_{j+1}} P_h > p_{j+1} - \delta$, for some integer $l \geq 1$. So, we can set

$$n_{j+1} = \min\{l \geq 1 : \sum_{h=1}^l P_{m_{j+1}+h} > p_{j+1} - \delta\}.$$

Being $P_{m_{j+1}+1} < \varepsilon \leq p_j - \varepsilon/(k+1) < p_j - \delta$, one can see that $n_{j+1} \geq 2$. Clearly, $\sum_{h=1}^{n_{j+1}-1} P_{m_{j+1}+h} \leq p_{j+1} - \delta$, and being $P_{m_{j+1}+n_{j+1}} < \varepsilon \leq \varepsilon/[(k-1)(k+1)] = 2\delta$,

$$\sum_{h=1}^{n_{j+1}} P_{m_{j+1}+h} < p_{j+1} + \delta.$$

Hence, (16) is true for $j + 1$.

We have just proved by induction that (16) holds true for $j = 1, \dots, k - 1$. This entails that (15) also holds true for $j = 1, \dots, k - 1$. To complete the proof, we need to show that (15) holds for $j = k$. To this aim, let us sum up (16) for $j = 1, \dots, k - 1$ obtaining that

$$\sum_{h=1}^{m_k} P_h = \sum_{j=1}^{k-1} \sum_{h=1}^{n_j} P_{m_j+h} < 1 - p_k + \varepsilon/[2(k+1)]. \tag{19}$$

Clearly, for some integer $l \geq 1$,

$$\sum_{h=1}^l P_h > 1 - \varepsilon/[2(k+1)]. \tag{20}$$

Subtracting (19) from (20) yields that $\sum_{h=1}^l P_{m_k+h} = \sum_{h=1}^l P_h - \sum_{h=1}^{m_k} P_h > p_k - \varepsilon/(k+1)$, for some integer $l \geq 1$. So, we can set:

$$n_k = \min\{l \geq 1 : \sum_{h=1}^l P_{m_k+h} > p_k - \varepsilon/(k+1)\}.$$

Being $P_{m_k+1} < \varepsilon < p_k - \varepsilon/(k+1)$, one can see that $n_k \geq 2$. Clearly, $\sum_{h=1}^{n_k-1} P_{m_k+h} \leq p_k - \varepsilon/(k+1)$, and being $P_{m_k+n_k} < \varepsilon \leq \varepsilon/[(k-1)(k+1)] < 2\varepsilon/(k+1)$,

$$\sum_{h=1}^{n_k} P_{m_k+h} < p_k + \varepsilon/(k+1).$$

Hence, (15) is true for $j = k$. □

Lemma 12. *Let P be a proper species sampling process with frequencies $(P_j)_{j \geq 1}$ such that, for every $\varepsilon > 0$,*

$$\mathbb{P}(\cap_{j \geq 1} \{P_j < \varepsilon\} \cap C) > 0, \tag{21}$$

for some C in the sigma-algebra generated by $(P_j)_{j \geq 1}$. Then for every measurable partition $\{A_1, \dots, A_k\}$ of \mathbb{X} with $\alpha(A_j) > 0$ ($1 \leq j \leq k$), every k -tuple (p_1, \dots, p_k) of positive real numbers that sum up to one and every $\varepsilon > 0$,

$$\mathbb{P}(\{|P(A_j) - p_j| < \varepsilon, j = 1, \dots, k\} \cap C) > 0.$$

Proof. If

$$p_j - \varepsilon/(k+1) < \sum_{1 \leq l \leq n: Y_l \in A_j} P_l < p_j + \varepsilon/(k+1), \tag{22}$$

for every $1 \leq j \leq k$ and some positive integer n , then, summing up each term w.r.t. $j = 1, \dots, k$, the first inequality in (22) yields:

$$\sum_{l > n} P_l \leq k\varepsilon/(k+1),$$

and therefore, being $P(A_j) = \sum_{l \geq 1: Y_l \in A_j} P_l$ (for $1 \leq j \leq k$),

$$p_j - \varepsilon / (k+1) \leq \sum_{1 \leq l \leq n: Y_l \in A_j} P_l \leq P(A_j) \leq \sum_{1 \leq l \leq n: Y_l \in A_j} P_l + \sum_{l > n} P_l \leq p_j + \varepsilon. \quad (23)$$

So, (22) implies (23) and therefore, denoting $\mathbb{P}_C(A) = \mathbb{P}(A \cap C)$, for every $A \in \mathcal{F}$,

$$\begin{aligned} & \mathbb{P}_C(|P(A_j) - p_j| < \varepsilon, j = 1, \dots, k) \\ & \geq \mathbb{P}_C \left(\bigcup_{n_1 \geq 1, \dots, n_k \geq 1} \bigcap_{j=1}^k \left\{ \left| p_j - \sum_{h=1}^{n_j} P_{m_j+h} \right| < \frac{\varepsilon}{k+1}, \right. \right. \\ & \qquad \qquad \qquad \left. \left. Y_{m_j+l} \in A_j, l = 1, \dots, n_j \right\} \right), \quad (24) \end{aligned}$$

where $m_1 = 0$ and $m_j = \sum_{l=1}^{j-1} n_l$ for $j = 2, \dots, k$.

To complete the proof, it is sufficient to show that the last term in (24) is positive. To this aim, note that there is a set in the union in (24) with positive \mathbb{P}_C -probability. In fact, by Lemma 11,

$$\mathbb{P}_C \left(\bigcup_{n_1 \geq 1, \dots, n_k \geq 1} \bigcap_{j=1}^k \left\{ \left| p_j - \sum_{h=1}^{n_j} P_{m_j+h} \right| < \frac{\varepsilon}{k+1} \right\} \right) > 0,$$

and therefore, for some k -tuple (n_1, \dots, n_k) ,

$$\mathbb{P}_C \left(\bigcap_{j=1}^k \left\{ \left| p_j - \sum_{h=1}^{n_j} P_{m_j+h} \right| < \frac{\varepsilon}{k+1} \right\} \right) > 0. \quad \square$$

Proof of Theorem 2. The condition ii. is equivalent to the following condition:

ii'. For every $\varepsilon > 0$ and every $w < 1$, $\mathbb{P}(\max_{j \geq 1} \bar{P}_j < \varepsilon, \sum_{l \geq 1} P_l > w) > 0$.

Let us prove that i. implies ii'. Denote by P the species sampling process (1) and let $\{A_1, \dots, A_k\}$ be a partition of \mathbb{X} such that $0 < \alpha(A_j) < 1$, A_j is α -continuous and define

$$\alpha_k(A) := \sum_{j=1}^k \frac{\alpha(A \cap A_j)}{k\alpha(A_j)}$$

for every $A \in \mathcal{X}$ and every integer $k \geq 1$. Hence, for every $k \geq 1$, α_k is a probability measure such that $S_{\alpha_k} = S_\alpha$ and belongs to the support of P by ii'. At this stage fix $\varepsilon > 0$, an integer $k > 1/\varepsilon$, and

$$0 < \delta < \min\{\varepsilon\alpha^*/(\alpha^* + \varepsilon), (1-w)\alpha^*\}, \quad (25)$$

where $\alpha^* = \max_{j=1, \dots, k} \alpha(A_j)$. Hence, one can write:

$$\begin{aligned} 0 &< \mathbb{P}(P \in U_{\delta-1/k, \dots, \delta-1/k}(\alpha_k; A_1, \dots, A_k)) \\ &= \mathbb{P}(\cap_{j=1}^k \{|P(A_j) - 1/k| < \delta - 1/k\}) = \mathbb{P}(\cap_{j=1}^k \{P(A_j) < \delta\}) \\ &= \mathbb{P}(\cap_{j=1}^k \{\sum_{l: Z_l \in A_j} P_l + (1 - \sum_{j \geq 1} P_j)\alpha(A_j) < \delta\}) \\ &\leq \mathbb{P}(\cap_{j=1}^k \{\sum_{l: Z_l \in A_j} P_l < \delta, (1 - \sum_{j \geq 1} P_j)\alpha(A_j) < \delta\}) \\ &\leq \mathbb{P}(\cap_{l \geq 1} \{P_l < \delta\} \cap \{(1 - \sum_{j \geq 1} P_j)\alpha^* < \delta\}) \\ &\leq \mathbb{P}(\cap_{l=1}^\infty \{\bar{P}_l < \delta/(1 - \delta/\alpha^*)\} \cap \{\sum_{j \geq 1} P_j > 1 - \delta/\alpha^*\}), \end{aligned}$$

which by (25) yields:

$$\mathbb{P}(\cap_{j \geq 1} \{\bar{P}_j < \varepsilon\} \cap \{\sum_{l \geq 1} P_l > w\}) > 0.$$

Let us prove that ii'. implies i. To this aim, let $S_\mu \subset S_\alpha$. By Lemma 10, we can complete the proof showing that

$$\mathbb{P}(Q \in U_{\varepsilon_1, \dots, \varepsilon_k}(\mu; A_1, \dots, A_k), \sum_{l \geq 1} P_l > w) \tag{26}$$

is positive, for every $(\varepsilon_1, \dots, \varepsilon_k, w) \in (0, 1)^{k+1}$, for every k -tuple (A_1, \dots, A_k) of μ -continuity subsets of \mathcal{X} and every integer $k \geq 1$.

Let $\{B_1, \dots, B_{2^k}\}$ be the partition generated by A_1, \dots, A_k . Since $S_\mu \subset S_\alpha$ and A_j is μ -continuous ($1 \leq j \leq k$), if $\alpha(B_l) = 0$, then $\mu(B_l) = 0$ ($1 \leq l \leq 2^k$). To show this note that B_1, \dots, B_{2^k} are μ -continuous, being μ -continuity closed under intersection (as $\partial(A \cap B) \subset \partial A \cup \partial B$ for every two sets A and B), and therefore $\mu(\partial B_l) = 0$, for $1 \leq l \leq 2^k$. Moreover, for every $1 \leq l \leq 2^k$, the interior of B_l (say B_l°) is an open set with zero α -measure, which implies that $(B_l^\circ)^c \supset S_\alpha \supset S_\mu$, and therefore $\mu(B_l^\circ) = 0$.

So $\alpha(B_l) = 0$ implies that $\mu(B_l) = 0$ ($1 \leq l \leq 2^k$), and therefore (26) is greater than or equal to

$$\mathbb{P}(\sum_{l \geq 1} P_l > w, |Q(B_l) - \mu(B_l)| < 2^{-k}\varepsilon, 1 \leq l \leq 2^k : \alpha(B_l) > 0),$$

which is positive by assumption i., in virtue of Lemma 12 with $C = \{\sum_{l \geq 1} P_l > w\}$.

Let us prove that ii'. implies iii. By definition, the series $\sum_{j \geq 1} P_j$ is less than or equal to one and therefore it converges, almost surely. Hence,

$$\mathbb{P}(\cup_{m \geq 1} \{\sum_{j > m} P_j < \varepsilon/2\}) = 1. \tag{27}$$

Applying ii'. with $w = 1 - \varepsilon/2$ and then (27), one obtains that:

$$\begin{aligned} 0 &< \mathbb{P}(P_j \leq \varepsilon \sum_{l \geq 1} P_l, j \geq 1, \sum_{l \geq 1} P_l > 1 - \varepsilon/2) \\ &= \mathbb{P}(\cup_{m \geq 1} \{P_j \leq \varepsilon \sum_{l \geq 1} P_l, j \geq 1, \sum_{l \geq 1} P_l > 1 - \varepsilon/2, \sum_{l > m} P_l < \varepsilon/2\}) \\ &\leq \sum_{m \geq 1} \mathbb{P}(P_j \leq \varepsilon \sum_{l \geq 1} P_l, j \geq 1, \sum_{l \geq 1} P_l > 1 - \varepsilon/2, \sum_{l > m} P_l < \varepsilon/2). \end{aligned}$$

This implies that for some $m \geq 1$,

$$\mathbb{P}(P_j \leq \varepsilon \sum_{l \geq 1} P_l, j \geq 1, \sum_{l \geq 1} P_l > 1 - \varepsilon/2, \sum_{l > m} P_l < \varepsilon/2) > 0 \quad (28)$$

Being $\sum_{l=1}^m P_j = \sum_{l \geq 1} P_l - \sum_{l > m} P_l$ and $\sum_{l \geq 1} P_l \leq 1$, almost surely, (28) implies that:

$$\mathbb{P}(P_j \leq \varepsilon, j \geq 1, \sum_{l=1}^m P_l > 1 - \varepsilon) > 0,$$

which in turn implies iii.

Let us now prove that iii. implies ii'. To this aim, fix $\varepsilon > 0$ and $0 < w < 1$ and set $\delta < \min\{\varepsilon/(1 + \varepsilon), 1 - w\}$. If $P_j < \delta$ for $j = 1, \dots, m$ and $\sum_{l=1}^m P_l > 1 - \delta$, then $P_j < \delta < \delta/(1 - \delta) \sum_{l=1}^m P_l < \delta/(1 - \delta) \sum_{l \geq 1} P_l$ for $j = 1, \dots, m$. Moreover, for every $j > m$, $P_j < \sum_{l > m} P_l = \sum_{l \geq 1} P_l - \sum_{l=1}^m P_l \leq 1 - \sum_{l=1}^m P_l < \delta < \delta/(1 - \delta) \sum_{l=1}^m P_l \leq \delta/(1 - \delta) \sum_{l \geq 1} P_l$. Being $\sum_{l=1}^m P_l \leq \sum_{l \geq 1} P_l$, these two facts allow us to write the following:

$$\begin{aligned} & \{P_j \leq \delta, j = 1, \dots, m, \sum_{l=1}^m P_l > 1 - \delta\} \\ & \subset \{P_j < \delta/(1 - \delta) \sum_{l \geq 1} P_l, j \geq 1, \sum_{l \geq 1} P_l > 1 - \delta\}. \end{aligned}$$

Therefore, being $\delta/(1 - \delta) < \varepsilon$ and $1 - \delta > w$, one can see that iii. implies ii'. \square

Appendix B: Proof of Proposition 6

To begin with, let us point out two facts about (5) and (6) which will be useful in the proof of Proposition 6. Being $1 - e^{-x} < \min\{x, 1\}$, for every $x > 0$, (5) implies that

$$\int_{\mathbb{R}^+} (1 - e^{-us}) \rho(ds) < \infty,$$

for every $u > 0$. Moreover, by the Markov inequality,

$$\int_{(0,1]} s \rho(ds) \geq t \cdot \rho(t, 1],$$

for every $0 < t \leq 1$. Therefore, (5) entails that $\rho(t, \infty) < \infty$, for every $t > 0$, which in turn by (6) implies that

$$\rho(0, t) = \infty, \quad (29)$$

for every $t > 0$.

Proof of Proposition 6. Our aim is to prove that

$$\mathbb{P}(T_1/S < \varepsilon, a < S < b) > 0, \quad (30)$$

for every triplet (ε, a, b) , where $\varepsilon > 0$ and $0 < a < b$. In fact, we can show that (30) implies the thesis. Indeed, (30) entails that

$$\mathbb{P}(T_1/S < \varepsilon, S \in B) = \mathbb{E}[\mathbb{P}(T_1/S < \varepsilon \mid S) \mathbb{I}_{\{S \in B\}}] > 0, \quad (31)$$

for every Borelian $B \subset \mathbb{R}$ containing a non degenerate interval, i.e. with positive Lebesgue measure. Hence, $\mathbb{P}(T_1/S < \varepsilon \mid S)$ is positive almost everywhere on \mathbb{R}^+ with respect to the Lebesgue measure. So, the set $\{\mathbb{P}(T_1/S < \varepsilon \mid S) = 0\}$

has zero Lebesgue measure. Being μ absolutely continuous with respect to the Lebesgue measure, the same set has also zero μ -measure. In other words, $\mathbb{P}(T_1/S < \varepsilon \mid S)$ is positive μ -almost surely. Therefore,

$$\mathbb{P}(\sup_{j \geq 1} P_j < \varepsilon) = \int_{\mathbb{R}^+} \mathbb{P}(T_1 < \varepsilon S \mid S = s) \mu(ds) > 0,$$

and Corollary 3 can be applied.

To ensure (30), it will be sufficient to prove that

$$\mathbb{P}(T_1 < \varepsilon a, a < S < b) > 0, \tag{32}$$

for every pair (a, b) where $0 < a < b$.

To begin with, fix $0 < a < b$, $\varepsilon > 0$. Moreover, let $0 < c < a\varepsilon$ and set $q \in (a/b, 1)$. Therefore, being $qb > a$ and $(1 - q)b > 0$,

$$\begin{aligned} \mathbb{P}(T_1 < \varepsilon a, a < S < b) &\geq \mathbb{P}(T_1 < \varepsilon a, a < \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j > c\}} < qb, \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j \leq c\}} < (1 - q)b) \\ &\geq \mathbb{P}(c < T_1 < \varepsilon a, a < \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j > c\}} < qb, \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j \leq c\}} < (1 - q)b). \end{aligned} \tag{33}$$

At this stage, note that $\Pi|_{(0,c]} := \{T_j : T_j \leq c, j \geq 1\}$ and $\Pi|_{(c,\infty)} := \{T_j : T_j > c, j \geq 1\}$, which are the restriction of Π to the intervals $(0, c]$ and (c, ∞) respectively, are independent (Poisson) processes. Hence,

$$\begin{aligned} &\mathbb{P}(c < T_1 < \varepsilon a, a < \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j > c\}} < qb, \sum_{j \geq 1} T_j \cdot \mathbb{I}_{\{T_j \leq c\}} < (1 - q)b) \\ &= \mathbb{P}(\Pi|_{(c,\infty)}(c, \infty) \geq 1, \Pi|_{(c,\infty)}(\varepsilon a, \infty) = 0, a < \int_{\mathbb{R}^+} x \Pi|_{(c,\infty)}(dx) < qb) \\ &\quad \cdot \mathbb{P}(\int_{\mathbb{R}^+} x \Pi|_{(0,c]}(dx) < (1 - q)b). \end{aligned} \tag{34}$$

Being $\Pi|_{(0,c]}$ a Poisson process with intensity $\rho(\cdot \cap (0, c])$, the Laplace transform of the sum of the points of $\Pi|_{(0,c]}$ is:

$$\mathbb{E}[\exp\{-t \int_{\mathbb{R}^+} x \Pi|_{(0,c]}(dx)\}] = \exp\{-\int_{(0,c]} (1 - e^{-tx}) \rho(dx)\},$$

which converges to one as $c \downarrow 0$, for every $t > 0$. Therefore, $\int_{\mathbb{R}^+} x \Pi|_{(0,c]}(dx)$ converges to zero in probability. Hence, the probability $\mathbb{P}(\int_{\mathbb{R}^+} x \Pi|_{(0,c]}(dx) < (1 - q)b)$ is positive for every c in $(0, \bar{c})$, for some $\bar{c} > 0$. On the basis of this fact, together with (33) and (34), the proof can be completed showing that

$$\mathbb{P}(\Pi|_{(c,\infty)}(c, \infty) \geq 1, \Pi|_{(c,\infty)}(\varepsilon a, \infty) = 0, a < \int_{\mathbb{R}^+} x \Pi|_{(c,\infty)}(dx) < qb) > 0, \tag{35}$$

for some $c > 0$ smaller than \bar{c} and $a\varepsilon$.

Recall that by (29), $\rho(0, x)$ is infinite for every $x > 0$. So, in particular, it is not zero. Hence, for every $x > 0$, there is a point of the support of ρ inside the interval $(0, x)$. On the basis of this fact, we can fix a point x in the support of ρ such that $0 < x < \min\{a\varepsilon, qb - a\}$. Let $k > 0$ be integer such that $a < kx < qb$. Moreover,

fix $c > 0$ smaller than x and set $0 < \delta < \min\{b/(2k) - x, x - a/k, x - c, \varepsilon a - x\}$. In this way,

$$(k(x - \delta), k(x + \delta)) \subset (a, qb) \quad (36)$$

and

$$(x - \delta, x + \delta) \subset (c, \varepsilon a). \quad (37)$$

By (36), to prove (35), it will be sufficient to show that:

$$\mathbb{P}(\Pi(c, \infty) \geq 1, \Pi(\varepsilon a, \infty) = 0, k(x - \delta) < \int_{\mathbb{R}^+} x \Pi|_{(c, \infty)}(dx) < k(x + \delta)) > 0, \quad (38)$$

for some $c > 0$ smaller than \bar{c} and $a\varepsilon$.

Recall that $\Pi|_{(c, \infty)}$ is distributed as a Bernoulli process conditionally on $\Pi(c, \infty)$. For every integer $m > 0$, let $(T_{1,m}, \dots, T_{m,m})$ be an m -tuple of independent and identically distributed random variables on $(0, \infty)$ such that $\mathbb{P}(T_1 \in A) = \rho(A \cap (c, \infty))/\rho(c, \infty)$, for every Borelian $A \subset (0, \infty)$. So, the probability in (38) becomes:

$$\begin{aligned} & \sum_{m \geq 1} e^{-\rho(c, \infty)} \cdot \frac{\rho(c, \infty)^m}{m!} \cdot \mathbb{P}\left(\max_{1 \leq j \leq m} T_{j,m} \leq \varepsilon a, k(x - \delta) < \sum_{j=1}^m T_{j,m} < k(x + \delta)\right) \\ & \geq e^{-\rho(c, \infty)} \cdot \frac{\rho(c, \infty)^k}{k!} \cdot \mathbb{P}\left(\max_{1 \leq j \leq k} T_{j,k} \leq \varepsilon a, k(x - \delta) < \sum_{j=1}^k T_{j,k} < k(x + \delta)\right), \end{aligned}$$

which by (37) is greater than or equal to

$$\begin{aligned} & \geq e^{-\rho(c, \infty)} \cdot \frac{\rho(c, \infty)^k}{k!} \cdot \mathbb{P}\left((x - \delta) < T_{j,k} < (x + \delta), j = 1, \dots, k\right) \\ & = e^{-\rho(c, \infty)} \cdot \frac{\rho(x - \delta, x + \delta)^k}{k!}, \end{aligned}$$

which in turn is positive since x belongs to the support of ρ . This entails (38). \square

References

- [1] BHATTACHARYA, A. and DUNSON, D. (2012). Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds. *Ann. I. Stat. Math.* **64** 687–714. [MR2927766](#)
- [2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York. [MR0233396](#)
- [3] BISSIRI, P. G., ONGARO, A. and WALKER, S. G. (2013). Species sampling models: consistency for the number of species. *Biometrika* **100** 771–777. [MR3094452](#)
- [4] DATTA, S. (1991). On the consistency of posterior mixtures and its applications. *Ann. Statist.* **19** 338–353. [MR1091855](#)
- [5] DE BLASI, P., JAMES, L. F. and LAU, J. W. (2010). Bayesian nonparametric estimation and consistency of mixed multinomial logit choice models. *Bernoulli* **16** 679–704. [MR2730644](#)
- [6] DUNSON, D. B. (2009). Nonparametric Bayes local partition models for random effects. *Biometrika* **96** 249–262. [MR2507141](#)

- [7] FAVARO, S., PRÜNSTER, I. and WALKER, S. G. (2011). On a class of random probability measures with general predictive structure. *Scand. J. Stat.* **38** 359–376. [MR2829605](#)
- [8] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. [MR0350949](#)
- [9] FUENTES-GARCÍA, R., MENA, R. H. and WALKER, S. G. (2010). A new Bayesian nonparametric mixture model. *Commun. Stat. Simulat.* **39** 669–682. [MR2785657](#)
- [10] GHOSAL, S. (2010). The Dirichlet process, related priors and posterior asymptotics. In *Bayesian Nonparametrics* (N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds.), 35–79. Cambridge Univ. Press, Cambridge. [MR2730660](#)
- [11] GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158. [MR1701105](#)
- [12] GNEDIN, A. (2010). A species sampling model with finitely many types. *Elect. Comm. in Probab.* **15** 79–88. [MR2606505](#)
- [13] GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sciences* **138** 5674–5685. [MR2160320](#)
- [14] HANSEN, B. and PITMAN, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statist. Probab. Lett.* **46** 251–256. [MR1745692](#)
- [15] HJORT, N. L., HOLMES, C., MÜLLER, P. and WALKER, S. G., eds. (2010). *Bayesian Nonparametrics. Cambridge Series in Statistical and Probabilistic Mathematics* **28**. Cambridge University Press, Cambridge. [MR2722987](#)
- [16] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Stat. Assoc.* **96** 161–173. [MR1952729](#)
- [17] ISHWARAN, H. and JAMES, L. F. (2003a). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Stat. Sinica* **13** 1211–1235. [MR2026070](#)
- [18] ISHWARAN, H. and JAMES, L. F. (2003b). Some further developments for stick-breaking priors: finite and infinite clustering and classification. *Sankhyā* **65** 577–592. [MR2060608](#)
- [19] JANG, G. H., LEE, J. and LEE, S. (2010). Posterior consistency of species sampling priors. *Stat. Sinica* **20** 581–593. [MR2682631](#)
- [20] KINGMAN, J. F. C. (1993). *Poisson Processes. Oxford Studies in Probability* **3**. The Clarendon Press Oxford University Press, New York. Oxford Science Publications. [MR1207584](#)
- [21] LIJOI, A., MENA, R. and PRÜNSTER, I. (2005). Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Stat. Infer. Stoch. Process.* **8** 283–309. [MR2177315](#)
- [22] LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica* **18** 1653–1668. [MR2469329](#)
- [23] MAJUMDAR, S. (1992). On topological support of Dirichlet prior. *Stat. Probabil. Lett.* **15** 385–388. [MR1193898](#)

- [24] NAVARRETE, C., QUINTANA, F. A. and MUELLER, P. (2008). Some issues in nonparametric Bayesian modeling using species sampling models. *Stat. model.* **8** 3–21. [MR2750628](#)
- [25] NOBILE, A. and FEARNISIDE, A. T. (2007). Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Stat. Comput.* **17** 147–162. [MR2380643](#)
- [26] ONGARO, A. (2005). Size-biased sampling and discrete nonparametric Bayesian inference. *J. Statist. Plann. Inference* **128** 123–148. [MR2110181](#)
- [27] ONGARO, A. and CATTANEO, C. (2004). Discrete random probability measures: a general framework for nonparametric Bayesian inference. *Stat. Probabil. Lett.* **67** 33–45. [MR2039931](#)
- [28] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York. [MR0226684](#)
- [29] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Rel.* **92** 21–39. [MR1156448](#)
- [30] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Rel.* **102** 145–158. [MR1337249](#)
- [31] PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. IMS Lecture Notes Monogr. Ser.* **30** 245–267. Inst. Math. Statist., Hayward, CA. [MR1481784](#)
- [32] PITMAN, J. (2003). Poisson-Kingman partitions. *Lecture Notes-Monograph Series* **40** 1–34. [MR2004330](#)
- [33] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. [MR1434129](#)
- [34] PRÜNSTER, I. and LIJOI, A. (2009). Models beyond the Dirichlet process. In *Bayesian Nonparametrics: Principles and Practice* (N. Hjort, C. Holmes, P. Müller and S. Walker, eds.), 80–136. Cambridge University Press, Cambridge, UK.
- [35] RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. R. Statist. Soc. B* **59** 731–792. [MR1483213](#)
- [36] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. [MR1309433](#)
- [37] STEPHENS, M. (2000). Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. *Ann. Stat.* **28** 40–74. [MR1762903](#)
- [38] TEH, Y. W. and JORDAN, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds.), 158–207. Cambridge Univ. Press, Cambridge. [MR2730663](#)
- [39] WU, Y. and GHOSAL, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Stat.* **2** 298–331. [MR2399197](#)