Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1–26

Large Sample Asymptotic Analysis for Normalized Random Measures with Independent Increments

Junxi Zhang* and Yaozhong Hu†,‡

Abstract. Normalized random measures with independent increments (NRMIs) represent a large class of Bayesian nonparametric priors and are widely used in the Bayesian nonparametric framework. In this paper, we provide the posterior consistency analysis for these NRMIs through their characterizing Lévy intensities. Assumptions are introduced on the Lévy intensities to analyse the posterior consistency and are verified with multiple interesting examples. Another focus of the paper is the Bernstein-von Mises theorem for a particular subclass of NRMIs, namely the normalized generalized gamma processes (NGGP). When the Bernstein-von Mises theorem is applied to construct credible sets, in addition to the usual form, there will be an additional bias term on the left endpoint closely related to the number of atoms of the true distribution in the discrete case. We also discuss the effect of the estimators for the model parameters of the NGGP under the Bernstein-von Mises convergence. Finally, to further illustrate the impact of the bias correction term in the construction of credible sets, we present a numerical example to demonstrate numerically how the bias correction affects the coverage of the true value.

Keywords: Bayesian nonparametrics, posterior consistency, Bernstein-von Mises theorem, normalized random measures with independent increments, Lévy intensity, normalized generalized gamma process, credible sets.

MSC2020 subject classifications: Primary 62G20, 60F05; secondary 62G15, 60G57.

1 Introduction

Bayesian nonparametrics has been undergone major investigation due to its various applications to diverse areas, such as biology, economics, machine learning and more. As a rich class of Bayesian nonparametric priors, normalized random measures with independent increments (NRMIs), introduced by Regazzini et al. (2003), include the famous Dirichlet process (Ferguson, 1973), the σ-stable NRMIs (Kingman, 1975), the normalized inverse Gaussian process (Lijoi et al., 2005b), the normalized generalized gamma process (Lijoi and Prünster, 2003; Lijoi et al., 2007b), and the generalized Dirichlet

*Department of Mathematical and Statistical Sciences, University of Alberta at Edmonton, Alberta, Canada, T6G 2G1, junxi3@ualberta.ca

†Department of Mathematical and Statistical Sciences, University of Alberta at Edmonton, Alberta, Canada, T6G 2G1, yaozhong@ualberta.ca

‡Supported by an NSERC discovery fund and a centennial fund of University of Alberta.

 https://doi.org/10.1214/23-BA1411

process (Lijoi et al., 2005a). We refer to Müller and Quintana (2004); Lijoi and Prünster (2010); Zhang and Hu (2021) for reviews of these processes with some important properties and applications.

In Bayesian nonparametric statistics, samples are drawn from a random probability measure with a prior distribution. To be more precise, let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space, let $\mathbb{X}$ be a complete, separable metric space whose σ-algebra is denoted by $\mathcal{X}$ and let $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ be the space of all probability measures on $\mathbb{X}$ with the σ-algebra generated by the topology of weak convergence. A sample $\mathbf{X} = (X_1, \ldots, X_n)$ that takes values in $\mathbb{X}^n$ is drawn iid (we use "iid" acronym to represent "independent and identically distributed" throughout the paper) from a random probability measure P conditional on P, which follows a prior distribution $\mathcal{Q}$ on $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$. That is to say,

$$X_1, \ldots, X_n | P \overset{iid}{\sim} P; \qquad P \sim \mathcal{Q}. \tag{1.1}$$

Two natural questions in the literature are raised as follows.

(i) A frequentist analysis of Bayesian consistency (Freedman and Diaconis, 1983): by assuming the "true" distribution of $\mathbf{X}$ is P_0, we are interested in whether the posterior law, that is the conditional law of $P|\mathbf{X}$, denoted by $\mathcal{Q}_n$, converges to δ_{P_0}, the Dirac measure with point mass at the "true" distribution, as $n \to \infty$.

(ii) What is the limiting distribution of centered and rescaled $P|\mathbf{X}$? In particular, are there Bernstein-von Mises like theorem and central limit theorem for P? If so, what are the limiting processes of $\sqrt{n}(P - \mathbb{P}_n)|\mathbf{X}$ and $\sqrt{n}(P|\mathbf{X} - \mathbb{E}[P|\mathbf{X}])$?

The above two questions play important roles in statistics. For question (i), posterior consistency is to guarantee that the model behaves "good" when the sample size is large. Here, the model behaves "good" means consistency, namely, under the assumption that $\mathbf{X} \overset{iid}{\sim} P_0$, the posterior distribution of $P|\mathbf{X}$ converges weakly to δ_{P_0} a.s.-P_0^∞. That said, $P|\mathbf{X}$ will converge to P_0 in distribution and this provides the validation of the Bayesian nonparametric procedure. For question (ii), the limiting distribution of the posterior process is the key ingredient to construct Bayesian credible sets and to conduct hypothesis tests.

Many inspiring works related to the above questions have been done. Regarding question (i), James (2008) obtains the posterior consistency analysis of the two-parameter Poisson-Dirichlet process, which is not an NRMI, but closely related to NRMIs (Pitman and Yor, 1997; Perman et al., 1992; Ghosal and Van der Vaart, 2017). The posterior consistency of species sampling priors (Pitman, 1996; Aldous, 1985) and Gibbs-type priors (De Blasi et al., 2015; Gnedin and Pitman, 2006) are discussed in Ho Jang et al. (2010) and De Blasi et al. (2013). It is worth to point out that there are overlaps among species sampling priors, Gibbs-type priors and the homogeneous NRMIs. Whereas, non-homogeneous NRMIs are different from species sampling priors and Gibbs-type priors. As for question (ii), Bernstein-von Mises results have been established for the Dirichlet process (Lo, 1983, 1986; Ray and van der Vaart, 2021; Hu and Zhang, 2022) and for the two-parameter Poisson-Dirichlet process (James, 2008; Franssen and van der Vaart, 2022). Along the same line, we would like to answer the two questions when P is an NRMI.

Since NRMIs are constructed by the normalization of completely random measures (Kingman, 1967, 1993) associated with their Lévy intensities (see e.g., Section 2), it is quite natural to study their properties based on the corresponding Lévy intensities. In this work, we discuss the posterior consistency of non-homogeneous NRMIs (including the homogeneous case as a particular case) and provide a simple condition to guarantee the posterior consistency of non-homogeneous NRMIs. As a result, when P_0 is continuous, posterior consistency does not generally hold for NRMIs, and when P_0 is discrete, posterior consistency holds as long as our proposed condition is satisfied. To compare our work with the studies in Ho Jang et al. (2010); De Blasi et al. (2013), our posterior consistency analysis is valid for non-homogeneous NRMIs, which are not covered in Ho Jang et al. (2010); De Blasi et al. (2013). For example, the posterior consistency analysis of this work covers a larger class of the Bayesian nonparametric priors, including the extended gamma NRMI (James et al., 2009, 2010) and the generalized extended gamma NRMI (defined in Section 3). On the other hand, the assumptions given in Ho Jang et al. (2010); De Blasi et al. (2013) for posterior consistency analysis depend on the associated random partition structure, which are not easy to verify as they are not always given explicitly. Our assumption (see Assumption 4) is simple and is based on the Lévy intensities used to define NRMIs, which are then always available explicitly. This assumption is very easy to verify, as we will explain and illustrate in Examples 12, 13, 14, 15 below for some interesting NRMIs, which makes our results immediately applicable.

In addition to the posterior consistency analysis, we further obtain the Bernstein-von Mises theorem for the normalized generalized gamma process (NGGP), which is a flexible subclass of NRMIs that includes the Dirichlet process, the normalized inverse-Gaussian process and the σ-stable process as special cases. Through the posterior consistency analysis, the NGGP is posterior consistent when the true distribution P_0 is discrete or when the true distribution P_0 is continuous and the parameter σ of the NGGP goes to 0. The parameter σ is one of the model parameters of the NGGP, it is a variance related parameter that controls the growth of the number of clusters induced by a sample of the NGGP as the sample size increases. The case when $\sigma \to 0$ would reduce the NGGP to the Dirichlet process. Thus, we should emphasize the case when the true distribution P_0 is discrete. However, there will be a bias term on the left hand side of the Bernstein-von Mises theorem for the NGGP when P_0 is discrete. It turns out that the bias term may not go to 0 when $n \to \infty$. Thus, in order to construct the "correct" Bayesian credible sets that cover the true parameter value, we suggest a bias correction to mitigate the bias term. The comparison of credible intervals with bias correction and without bias correction is illustrated in the numerical experiment. In application, the model parameters of the NGGP are chosen by some data driven estimators and we show that the Bayesian estimator or maximum likelihood estimators of the model parameters of the NGGP will not affect the convergences in the Bernstein-von Mises results. The Bernstein-von Mises results in this work cover the findings in Lo (1983, 1986); Ray and van der Vaart (2021); Hu and Zhang (2022) for the Dirichlet process, which is a special case of the NGGP. However, the two-parameter Poisson-Dirichlet process is not included in the NGGP, thus our Bernstein-von Mises results fill the gap of frequentist theoretical understanding of the NGGP and complement the works in James (2008); Franssen and van der Vaart (2022).

The outline of this paper is as follows. In Section 2, we recall the construction of NRMIs, their Lévy intensity, and their posterior distributions. In Section 3, we analyse posterior consistency of the homogeneous and non-homogeneous NRMIs under a simple assumption on the corresponding Lévy intensities. Verification of the introduced assumption is carried out for several well-known Bayesian nonparametric priors to demonstrate its applicability and advantage. In Section 4, we derive the Bernstein-von Mises theorem for the NGGP and provide an analysis of the bias correction, together with a numerical illustration. Finally, in Section 5, we provide a discussion of our results and some ideas that can be studied in the future. In order to ease the flow of the ideas, we delay the proofs to the Supplementary Materials (Zhang and Hu, 2024).

2 Normalized random measures with independent increments

2.1 Constructions of NRMIs

We start by recalling the notions of completely random measures (see e.g., (Kingman, 1967, 1993) and references therein for more details), which play an important role in the construction of NRMIs.

Let $\mathbb{B}_{\mathbb{X}}$ be the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$, in the sense that for any $\mu \in \mathbb{B}_{\mathbb{X}}$ and any bounded set $A \in \mathcal{X}$ one has $\mu(A) < \infty$. Let $\mathbb{B}_{\mathbb{X}}$ be endowed with a suitable topology so that the associated Borel σ-algebra $\mathcal{B}_{\mathbb{X}}$ can be introduced (Daley and Vere-Jones, 2008).

Definition 1. *Let $\tilde{\mu}$ be a measurable function defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$. We say that $\tilde{\mu}$ is a completely random measure (CRM) if the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_d)$ are mutually independent, for any pairwise disjoint sets $A_1, \ldots, A_d$ in $\mathcal{X}$, where $d \geq 2$ is a finite integer.*

Completely random measures play an important role in Bayesian nonparametric priors and we refer to Regazzini et al. (2003); Lijoi and Prünster (2010) for more detailed discussion.

One way to construct NRMIs is through Poisson random measures explained as follows. Denote $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$ and denote its Borel σ-algebra by $\mathcal{S}$. A Poisson random measure $\tilde{N}$ on $\mathbb{S}$ with finite mean measure $\nu(ds, dx)$ is a random measure from $\Omega \times \mathbb{S}$ to $\mathbb{R}_+$ satisfying:

(i) $\tilde{N}(B) \sim \text{Poisson}(\nu(B))$ for any B in $\mathcal{S}$ such that $\nu(B) < +\infty$;

(ii) for any pairwise disjoint sets $B_1, \ldots, B_m$ in $\mathcal{S}$, the random variables $\tilde{N}(B_1), \ldots, \tilde{N}(B_m)$ are mutually independent.

The Poisson mean measure ν satisfies the condition (see (Daley and Vere-Jones, 2008) for details of Poisson random measures) that

$$\int_0^\infty \int_{\mathbb{X}} \min(s,1)\nu(ds,dx) < \infty.$$

Let $\tilde{\mu}$ be a random measure defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$ defined as follows,

$$\tilde{\mu}(A) := \int_0^\infty \int_A s\tilde{N}(ds,dx), \quad \forall A \in \mathcal{X}. \tag{2.1}$$

It is trivial to verify that $\tilde{\mu}$ is a completely random measure. It is also well-known that for any $A \in \mathcal{X}$, $\tilde{\mu}(A)$ is discrete and is uniquely characterized by its Laplace transform as follows:

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(A)}\right] = \exp\left\{-\int_0^\infty \int_A \left[1 - e^{-\lambda s}\right]\nu(ds,dx)\right\}. \tag{2.2}$$

The measure ν is called the *Lévy intensity* of $\tilde{\mu}$ and we denote the Laplace exponent by

$$\psi_A(\lambda) = \int_0^\infty \int_A \left[1 - e^{-\lambda s}\right]\nu(ds,dx). \tag{2.3}$$

From the Laplace transform in (2.2), we are aware that the completely random measure $\tilde{\mu}$ is characterized completely by its Lévy intensity ν, which usually takes the following forms in the literature.

(a) $\nu(ds,dx) = \rho(ds|x)\alpha(dx)$, where α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ so that $\alpha(\mathbb{X}) = a < \infty$ and ρ is defined on $\mathcal{B}(\mathbb{R}^+) \times \mathbb{X}$ such that for any $x \in \mathbb{X}$, $\rho(\cdot|x)$ is a σ-finite measure on $\mathcal{B}(\mathbb{R}^+)$ and for any $A \in \mathcal{X}$, $\rho(A|\cdot)$ is $\mathcal{B}(\mathbb{R}^+)$ measurable. The corresponding $\tilde{\mu}$ is called *non-homogeneous* completely random measure.

(b) If the above $\rho(ds|x)$ is independent of x, namely, $\nu(ds,dx) = \rho(ds)\alpha(dx)$, where α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ so that $\alpha(\mathbb{X}) = a < \infty$ and $\rho : \mathcal{B}(\mathbb{R}^+) \to \mathbb{R}^+$ is some measure on $\mathbb{R}^+$. The corresponding $\tilde{\mu}$ is called *homogeneous* completely random measure.

To avoid confusion, it is worth to point out that case (b) is a special case of case (a). We single out case (b) since it is an important particular case (that is used to construct homogeneous NRMIs) that is frequently used. Usually, we assume that α is a finite measure so we may write $\alpha(dx) = aH(dx)$ for some probability measure H and some constant $a = \alpha(\mathbb{X}) \in (0, \infty)$.

To construct NRMIs, the completely random measures will be normalized, and thus one needs the total mass $\tilde{\mu}(\mathbb{X})$ to be finite and positive almost surely. This happens under the condition that $\rho(\mathbb{R}^+) = \infty$ in homogeneous case and that $\rho(\mathbb{R}^+|x) = \infty$ for all $x \in \mathbb{X}$ in non-homogeneous case (Regazzini et al., 2003). Under the above conditions, an NRMI P on $(\mathbb{X}, \mathcal{X})$ is a random probability measure defined by

$$P(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}. \tag{2.4}$$

P is discrete almost surely due to the discreteness of $\tilde{\mu}$. For notional simplicity, we let $T = \tilde{\mu}(\mathbb{X})$ and let $f_T(t)$ be the density of T throughout this paper.

2.2 Posterior of NRMIs

We will recall the posterior analysis (James et al., 2009) of NRMIs, which is a key topic in Bayesian nonparametric analysis. As commonly assumed in Bayesian models and throughout the paper, we shall also assume that our sample is exchangeable. We briefly recall the concept of exchangeable sequence (see e.g., (De Finetti, 1937; Aldous, 1985; Kallenberg, 2005)). Let us consider an infinite sequence of random variables $\mathbf{X}^\infty = (X_i)_{i\geq 1}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with each X_i taking values in $\mathbb{X}$. The infinite sequence $\mathbf{X}^\infty$ is called *exchangeable* if the probability distribution of $(X_1, \ldots, X_n)$ coincides with the probability distribution of $(X_{\varrho(1)}, \ldots, X_{\varrho(n)})$ for any $n \geq 1$ and any permutation ϱ of $(1, \ldots, n)$. The exchangeability assumption is usually formulated in terms of conditional iid as introduced in (1.1) and is given in its following equivalent form:

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n | P] = \prod_{i=1}^{n} P(A_i), \tag{2.5}$$

for any $n \geq 1$ and any measurable $A_1, \ldots, A_n$ in $\mathcal{X}$.

Let P be an NRMI on $\mathbb{X}$. Due to the almost surely discreteness of P as mentioned in (2.4), it is possible that $\mathbb{P}(X_i = X_j) > 0$ for $i \neq j$ and hence there is a random partition structure associated with the exchangeable random sequence. To represent the associated random partition, for any $n \geq 1$, let $\mathbf{Y} = (Y_1, \ldots, Y_{n(\pi)})$ be the distinct observations of the exchangeable sequence $\mathbf{X} = (X_1, \ldots, X_n)$, where $n(\pi)$ is the number of distinct values of $\mathbf{X}$. This gives a partition $\pi = (i_1, \ldots, i_{n_1}, \ldots, i_{n_{n(\pi)-1}}, \ldots, i_{n_{n(\pi)}})$ of indices $(1, \ldots, n)$ of size $n(\pi)$, so that $\sum_{j=1}^{n(\pi)} n_j = n$, and $Y_1 := X_{i_1} = \cdots = X_{i_{n_1}}, \ldots, Y_{n(\pi)} := X_{i_{n_{n(\pi)-1}+1}} = \cdots = X_{i_{n_{n(\pi)}}}$. To state the posterior analysis result we let

$$\tau_k(u, Y) = \int_0^\infty s^k e^{-us} \rho(ds|Y) \quad \text{for any positive integer } k \text{ and } Y \in \mathbb{X}. \tag{2.6}$$

With these notations, the posterior distribution of P conditional on the observations of the sample $(X_1, \ldots, X_n)$ is given by the following theorem.

Theorem 2 (James et al., 2009)**.** *Let P be an NRMI with intensity $\nu(ds, dx) = \rho(ds|x)\alpha(dx)$. The posterior distribution of P, given a latent random variable U_n, is* *an NRMI that coincides in distribution with the random measure*

$$\kappa_n \frac{\tilde{\mu}_{(U_n)}}{T_{(U_n)}} + (1-\kappa_n) \sum_{j=1}^{n(\pi)} \frac{J_j \delta_{Y_j}}{\sum_{j=1}^{n(\pi)} J_j}, \tag{2.7}$$

where

(i) the random variable U_n has density

$$f_{U_n}(u) = \frac{u^{n-1}}{\Gamma(n)} \int_0^\infty t^n e^{-ut} f_T(t) dt; \tag{2.8}$$

(ii) given U_n, $\tilde{\mu}_{(U_n)}$ is the conditional completely random measure of $\tilde{\mu}$ with the Lévy intensity $\nu_{(U_n)} = e^{-U_n s}\rho(ds|x)\alpha(dx)$;

(iii) $\{J_1, \dots, J_{n(\pi)}\}$ are random variables depending on U_n and Y_j and having density

$$f_{J_j}(s|U_n = u, \mathbf{X}) = \frac{s^{n_j} e^{-us}\rho(ds|Y_j)}{\int_0^\infty s^{n_j} e^{-us}\rho(ds|Y_j)}; \tag{2.9}$$

(iv) the random elements $\tilde{\mu}_{(U_n)}$ and J_j, $j \in \{1, \dots, n(\pi)\}$ are independent;

(v) $T_{(U_n)} = \tilde{\mu}_{(U_n)}(\mathbb{X})$ and $\kappa_n = \frac{T_{(U_n)}}{T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j}$;

(vi) the conditional distribution of U_n given $\mathbf{X}$ admits a density function coinciding with

$$f_{U_n|\mathbf{X}}(u|\mathbf{X}) \propto u^{n-1} e^{-\psi(u)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(u, Y_j). \tag{2.10}$$

The above theorem shows that, given the latent variable U_n, the posterior of P is a weighted sum of another NRMI $\frac{\tilde{\mu}_{(U_n)}}{T_{(U_n)}}$ and the normalization of Dirac measure δ_{Y_j} of distinct observations Y_j, multiplied by its corresponding jumps J_j. This gives a rather complete description of the posterior distribution of NRMIs. More details of the posterior analysis of $\tilde{\mu}$ and P can be found in James et al. (2009).

3 Posterior consistency analysis for the NRMIs

In this section, we discuss the posterior consistency for NRMIs as pointed out in question (i) in the introduction. Recall that $\mathbb{M}_{\mathbb{X}}$ is the space of probability measures on $\mathbb{X}$ and $\mathcal{M}_{\mathbb{X}}$ is the corresponding σ-algebra generated by the topology of weak convergence. Assume that $\mathbf{X} = (X_1, \dots, X_n)$ is a sample from the "true" distribution P_0 in $\mathbb{M}_{\mathbb{X}}$. Namely, $\mathbf{X}$ is iid P_0- distributed. Let $\mathcal{Q}_n$ denote the probability law of the posterior random probability measure $P|\mathbf{X}$. The posterior distribution is said to be weakly consistent if for any neighbourhood $O \in \mathcal{M}_{\mathbb{X}}$ of P_0 one has

$$\mathcal{Q}_n(O) \to 1 \qquad \text{a.s.-}P_0^{\infty},$$

as $n \to \infty$. Here and throughout the paper, $P_0^{\infty} = P_0 \times P_0 \cdots$ is the infinite product measure on $\mathbb{X}^{\infty}$, that makes the random variables $X_1, X_2, \ldots$ independent with common true distribution P_0.

Before presenting the main result, we shall give the following lemma, which provides the moments of the posterior P. The lemma plays an important role in the proof of the main theorem. By recalling ψ_A in (2.3), we denote

$$V_{\alpha(A)}^{(k)}(y) = (-1)^k e^{\psi_A(y)} \frac{d^k}{dy^k} e^{-\psi_A(y)}, \tag{3.1}$$

for any $A \in \mathcal{X}$.

Lemma 3. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from an NRMI P. The moments and the mixed moments of the posterior of P given $\mathbf{X}$ are given as follows.*

(i) For any $A \in \mathcal{X}$ and $m \in \mathbb{N}$, the posterior m-th moment of P is given by

$$\begin{aligned}\mathbb{E}[(P(A))^m|\mathbf{X}] = \frac{\Gamma(n)}{\Gamma(m+n)} \sum_{0 \le l_1 + \cdots + l_{n(\pi)} \le m}^{m} \binom{m}{l_1, \ldots, l_{n(\pi)}} \int_0^{\infty} u^m f_{U_n|\mathbf{X}}(u|\mathbf{X}) \\ \times V_{\alpha(A)}^{(m-(l_1+\cdots+l_{n(\pi)}))}(u) \left(\prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A) \right) du. \end{aligned} \tag{3.2}$$

(ii) For any family of pairwise disjoint subsets $\{A_1, \ldots, A_q\}$ of $\mathcal{X}$ and any integers $(m_1, \ldots, m_q)$, we have

$$\begin{aligned}\mathbb{E}\left[P(A_1)^{m_1} \cdots P(A_q)^{m_q}|\mathbf{X}\right] = \frac{\Gamma(n)}{\Gamma(m+n)} \int_0^{\infty} u^m f_{U_n|\mathbf{X}}(u|\mathbf{X}) \\ \times \prod_{i=1}^{q+1} \Bigg\{ \sum_{0 \le l_1 + \cdots + l_{\#(\lambda_i)} \le m_i}^{m_i} \binom{m_i}{l_1, \ldots, l_{\#(\lambda_i)}} \\ \times V_{\alpha(A_i)}^{(m_i-(l_1+\cdots+l_{\#(\lambda_i)}))}(u) \left(\prod_{j \in \lambda_i} \frac{\tau_{n_j+l_j}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \right) \Bigg\} du, \end{aligned} \tag{3.3}$$

where $m = \sum_{i=1}^q m_i$, $A_{q+1} = (\cup_{i=1}^q A_i)^c$, $m_{q+1} = 0$, $\lambda_i = \{j : Y_j \in A_i\}$ is the set of the index of Y_j's that are in A_i, and $\#(\lambda_i)$ is the number of components in λ_i.

The above lemma provides the posterior moments of NRMIs. Such results can be reduced to the moments of NRMIs by letting the sample size $n = 0$. The proof of

Lemma 3 is inspired by the idea in James et al. (2006) and the details are given in the Supplementary Materials. To apply the above lemma, one needs to deal with the term $V_{\alpha(A)}^{(k)}(y)$ defined by (3.1). We give the following recursion formula for this quantity:

$$V_{\alpha(A)}^{(k)}(y) = \sum_{i=0}^{k-1} \binom{k-1}{i} \xi_{k-i}(y) V_{\alpha(A)}^{(i)}(y),$$

where $\xi_i(y) = \int_A \tau_i(y,x)\alpha(dx)$.

To answer question (i) mentioned in the introduction, we shall introduce the following assumption that is the key to analyse the posterior consistency of NRMIs.

Assumption 4. *Let $\tau_k(u,x)$ be defined by (2.6) and assume that for each $k \in \mathbb{Z}^+$ and $x \in \mathbb{X}$, $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)}$ is nondecreasing in u and satisfying*

$$k - 1 < \sup_{u>0} u \frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} \leq k. \tag{3.4}$$

Remark 5. *Condition (3.4) is equivalent to assuming that for each $k \in \mathbb{Z}^+$ the function*

$$C_k(x) = k - \sup_{u>0} u \frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} \tag{3.5}$$

takes values in $[0,1)$.

We shall need $C_1(x)$ to represent the bias in the following posterior consistent analysis. In the examples of application, we shall find $C_k(x)$ from the Lévy intensities.

Theorem 6. *Let P be an NRMI with Lévy intensity $\nu(ds,dx) = \rho(ds|x)\alpha(dx)$, where $\rho(ds|x)$ satisfies Assumption 4. Recall that $H(\cdot) = \frac{\alpha(\cdot)}{\alpha(\mathbb{X})}$. Then*

(i) If P_0 is continuous, then the posterior of P converges weakly to a point mass at $\bar{C}_1 H(\cdot) + (1-\bar{C}_1)P_0(\cdot)$ a.s.-P_0^∞, where $\bar{C}_1 = \lim_{n\to\infty} \frac{\sum_{i=1}^n C_1(X_i)}{n}$.

(ii) If P_0 is discrete with $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ almost surely, then P is weakly consistent, i.e., the posterior of P converges weakly to a point mass at $P_0(\cdot)$ a.s.-P_0^∞.

With the help of the moment results in Lemma 3, Assumption 4 plays the key role to make the posterior variance of NRMIs go to 0 when $n \to \infty$. Thus, the posterior distribution of P will converge to the point mass at the posterior expectation of P (more details with explanation are given in the proof of Theorem 6 in the Supplementary Materials). Although Assumption 4 looks complicated, it is quite easy to check as long as $\rho(ds|x)$ is given. For instance, the intensities $\rho(ds|x)$ for almost all popular NRMIs are gamma type, the corresponding $\tau_k(u,x)$ are gamma integrals, thus can be calculated directly. We shall check Assumption 4 for some popular specific NRMIs in Examples 12, 13, 14 and 15 to show how Assumption 4 works for these processes. This demonstrates the wide applicability of Theorem 6.

As a comparison between Theorem 6 and the results in Ho Jang et al. (2010) for the species sampling priors and in De Blasi et al. (2013) for the Gibbs-type priors, Theorem 6 considers the consistency results for the non-homogeneous NRMIs, which is a very general class of Bayesian nonparametric priors that are not covered by the species sampling priors and the Gibbs-type priors. For instance, the posterior consistency analysis of some non-homogeneous NRMIs, like the extended gamma NRMI (James et al., 2009, 2010) and the generalized extended gamma NRMI (defined in Example 15), are not included in Ho Jang et al. (2010) and De Blasi et al. (2013), however, they are covered by our Theorem 6. It is worth noting that Bayesian nonparametric priors based on non-homogeneous CRMs, for example the extended gamma NRMI, play important role in Bayesian nonparametric inference for modelling survival data and spatial phenomena (Ferguson, 1974; Hjort, 1990; James et al., 2010; Lijoi and Prünster, 2010). On the other hand, although the predictive distributions for homogeneous NRMIs are given (Pitman, 2003; James et al., 2006), the conditions in Ho Jang et al. (2010) and De Blasi et al. (2013) are not trivial to verify, however, our Assumption 4 is quite easy to verify as long as $\rho(ds|x)$ is given.

In Theorem 6, we require $\lim_{n\to\infty}\frac{n(\pi)}{n}=0$ a.s. as a condition to guarantee the posterior consistency result when P_0 is discrete. This condition is always true by the following proposition (which coincides with the results in Lemma 2 in Ho Jang et al. (2010)).

Proposition 7. *When P_0 is discrete, we have $\lim_{n\to\infty}\frac{n(\pi)}{n}=0$, almost surely. When P_0 is continuous, we have $\lim_{n\to\infty}\frac{n(\pi)}{n}=1$, almost surely.*

By the identity that $\frac{d}{du}\tau_k(u,x)=\frac{d}{du}\int_0^\infty s^k e^{-us}\rho(ds|x)=-\tau_{k+1}(u,x)$, the following assumption is equivalent to Assumption 4.

Assumption 8. *Let $\tau_k(u,x)$ be defined by (2.6) and assume that for each $k\in\mathbb{Z}^+$ and $x\in\mathbb{X}$, $u\frac{d}{du}\ln(\tau_k(u,x))$ is nonincreasing in u and satisfying*

$$-k\le\inf_{u>0}u\frac{d}{du}\ln(\tau_k(u,x))<-k+1. \tag{3.6}$$

Remark 9. *Condition (3.6) is equivalent to assuming that for each $k\in\mathbb{Z}^+$ the function*

$$C_k(x)=k+\inf_{u>0}u\frac{d}{du}\ln(\tau_k(u,x)) \tag{3.7}$$

takes values in $[0,1)$.

Remark 10. *There are more general NRMIs. For example, James (2002) introduced the h-biased random measures $\tilde{\mu}$ by $\int_{\mathbb{Y}\times\mathbb{X}}g(s)\tilde{N}(ds,dx)$, where $g:\mathbb{Y}\to\mathbb{R}^+$ is an integrable function on any complete and separable metric space $\mathbb{Y}$. Theorem 6 can be extended to this type of NRMIs.*

One interesting quantity to be considered is $n(\pi)$, the number of distinct observations of the sample $(X_1,\ldots,X_n)$ from a Bayesian nonparametric model. In Bayesian nonparametric mixture models, $n(\pi)$ is the number of clusters in the sample observations and thus is studied in a number of works that are concerned with the clustering problems and so on. Among the literature let us mention that the distribution of $n(\pi)$ is obtained in Korwar and Hollander (1973) when $(X_1, \dots, X_n)$ is a sample from the Dirichlet process; in Antoniak (1974) when $(X_1, \dots, X_n)$ is a sample from the mixture of Dirichlet process; in Pitman (2003) when $(X_1, \dots, X_n)$ is a sample from the two-parameter Poisson-Dirichlet process. It is interesting to point out that the referred distributions of $n(\pi)$ are priors of the number of clusters. When $(X_1, \dots, X_n)$ is a sample from the general NRMIs, we have by a result of James et al. (2009):

Proposition 11. *For any positive integer n, the distribution of $n(\pi)$ is*

$$\begin{aligned} &\mathbb{P}(n(\pi) = k) \\ &\quad = \int_0^\infty \frac{nu^{n-1}}{k!} e^{-\int_{\mathbb{X}} \int_0^\infty (1-e^{-us}) \rho(ds|x) \alpha(dx)} \sum_{(n_1, \dots, n_k)} \prod_{j=1}^k \frac{\int_{\mathbb{X}} \tau_{n_j}(u, x) \alpha(dx)}{n_j!} du, \end{aligned} \tag{3.8}$$

where $k = 1, \dots, n$, and the summation is over all vectors of positive integers $(n_1, \dots, n_k)$ such that $\sum_{j=1}^k n_j = n$.

As we mentioned above, Assumption 4 is in fact quite easy to verify. We provide in the following examples to see the applicability of Theorem 6.

Example 12. *The normalized generalized gamma process NGGP(a, σ, θ, H) (Lijoi and Prünster, 2003; Lijoi et al., 2007b) is an NRMI with the following homogeneous Lévy intensity*

$$\nu(ds, dx) = \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\theta s} ds \alpha(dx), \tag{3.9}$$

where the parameters $\sigma \in (0,1)$ and $\theta > 0$. It is easy to see that the Laplace transform for $\tilde{\mu}(A)$ is

$$\mathbb{E}\left[e^{-\lambda \tilde{\mu}(A)}\right] = \exp\left\{-\frac{\alpha(A)}{\sigma}[(\lambda+\theta)^\sigma - \theta^\sigma]\right\}.$$

When $\theta \to 0$, this NRMI yields the homogeneous σ-stable NRMI introduced by Kingman (1975). Letting $\sigma \to 0$, this NRMI becomes the Dirichlet process (Ferguson, 1973). If we let $\sigma = \theta = \frac{1}{2}$, this NRMI becomes the normalized inverse-Gaussian process (Lijoi et al., 2005b).

It is easy to check that for any nonnegative integer k,

$$\tau_k(u, x) = \tau_k(u) = \frac{1}{\Gamma(1-\sigma)} \int_0^\infty s^{k-\sigma-1} e^{-(u+\theta)s} ds = \frac{\Gamma(k-\sigma)}{\Gamma(1-\sigma)(u+\theta)^{k-\sigma}}.$$

It is obvious that $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k-\sigma}{u+\theta}$ is increasing in u with the supremum $k-\sigma$ that belongs to $(k-1, k]$. In this case, the sequence of functions $\{C_k(x)\}$ is a sequence of constants σ, namely, $C_k(x) = \sigma$ for any $k \in \mathbb{Z}^+$ and $x \in \mathbb{X}$. Thus, Assumption 4 is verified and Theorem 6 implies the normalized generalized gamma process is posterior consistent when $\sigma \to 0$ (i.e. the Dirichlet process), or when P_0 is discrete.

Example 13. *The generalized Dirichlet process GDP(a, γ, H) (Lijoi et al., 2005a) is an NRMI with the following homogeneous Lévy intensity*

$$\nu(ds, dx) = \sum_{j=1}^{\gamma} \frac{e^{-js}}{s} ds\alpha(dx), \tag{3.10}$$

where γ is a positive integer. The corresponding Laplace transform of $\tilde{\mu}(A)$ is

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(A)}\right] = \left(\frac{(\gamma!)}{(\lambda+1)_{\gamma}}\right)^{\alpha(A)},$$

where for $c > 0$, $(c)_k = \frac{\Gamma(c+k)}{\Gamma(c)}$ is the ascending factorial of c for any positive integer k. When $\gamma = 1$, the generalized Dirichlet process is reduced to the Dirichlet process.

It is trivial to obtain for any nonnegative integer k,

$$\tau_k(u, x) = \tau_k(u) = \sum_{j=1}^{\gamma} \frac{k}{(u+j)^k}.$$

It follows $\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = k\frac{\sum_{j=1}^{\gamma}(u+j)^{-k-1}}{\sum_{j=1}^{\gamma}(u+j)^{-k}} \in (\frac{k}{u+\gamma}, \frac{k}{u+1})$, which implies $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k}{u+c(\gamma)}$ with some constant $c(\gamma) \in (1, \gamma)$. Therefore, $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)}$ is increasing in u with the supremum k. In this case, $C_k(x) = 0$ for any $z \in \mathbb{Z}^+$ and $x \in \mathbb{X}$. Theorem 6 can then be used to conclude that the generalized Dirichlet process is posterior consistent.

Example 14. *As a non-homogeneous example, we consider the extended gamma NRMI whose non-homogeneous Lévy intensity is given by*

$$\nu(ds, dx) = \frac{e^{-\beta(x)s}}{s} ds\alpha(dx), \tag{3.11}$$

where $\beta(x) : \mathbb{X} \to \mathbb{R}^+$ is an integrable function (with respect to $\alpha(dx)$). Such NRMI is constructed by the normalization of the extended gamma process on $\mathbb{R}$ introduced by Dykstra and Laud (1981). More generally, Lo (1982) studied the extended Gamma process, called weighted Gamma process on abstract spaces.

By a trivial computation, for any nonnegative integer k, $\tau_k(u, x) = \frac{\Gamma(k)}{(u+\beta(x))^k}$ and thus $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k}{u+\beta(x)}$, which is increasing in u with the supremum k. Therefore, $C_k(x) = 0$ for any $k \in \mathbb{Z}^+$ and $x \in \mathbb{X}$, Assumption 4 is satisfied. Theorem 6 implies that the extended gamma NRMI is posterior consistent when $\beta(x)$ is integrable with respect to $\alpha(dx)$.

Our theorem can also be applied to more general NRMIs which have not been investigated in previous works. In the next example, we naturally define a new non-homogeneous NRMI that is called the generalized extended gamma NRMI.

Example 15. *We say that the non-homogeneous NRMI P in (2.4) is the generalized extended gamma NRMI, if the corresponding Lévy intensity is*

$$\nu(ds, dx) = \sum_{i=1}^{r} \frac{e^{-\beta_i(x)s}}{s} ds\alpha(dx),$$

where $r \in \mathbb{Z}^+$ and $\beta_i(x) : \mathbb{X} \to \mathbb{R}^+$ are integrable functions (with respect to $\alpha(dx)$).

A similar argument to that of Example 13 and that of Example 14 implies that the generalized extended gamma NRMI is posterior consistent when $\beta_i(x)$ is integrable (with respect to $\alpha(dx)$) for all $i \in \{1, \dots, r\}$.

To summarize the previous discussion up to now we have answered the question (i) raised in the introduction. The posterior distribution of NRMIs when P_0 is continuous is consistent only in the case when $\bar{C}_1 = 0$ or $H = P_0(\mathbb{P}_n)$. However, it is rare to choose H to be the "true" distribution P_0 and it is not possible to let $H = \mathbb{P}_n$ before a sample is observed. Therefore, the assumption $\bar{C}_1 = 0$ should be made to guarantee the posterior consistency for NRMIs when P_0 is continuous. Furthermore, whenever $\rho(ds|x)$ is gamma type, $\bar{C}_1 = 0$ would reduce the corresponding P to the Dirichlet process, the extended gamma NRMI, the generalized Dirichlet process or the generalized extended gamma NRMI. However, posterior inconsistency of P when P_0 is continuous is not a big issue, as P is discrete and it is hardly used as a prior for the distribution of continuous data. On the other hand, posterior consistency of P when P_0 is discrete is more important.

4 Bernstein-von Mises theorem for the generalized normalized gamma process

The celebrate Bernstein-von Mises theorem links Bayesian inference with frequentist inference. The Bernstein-von Mises theorem plays important role in Bayesian parametric model (Vaart, 1998; Le Cam, 2012). To better explain this theorem let us consider a parametric model $(p_\theta : \theta \in \Theta)$, where Θ is finite dimensional and the parameter is assumed to follow a prior distribution, $\theta \sim \Pi$. Suppose we have iid observations $\mathbf{X}$ from p_{θ_0}. The Bernstein-von Mises theorem states that, under some mild and universal assumptions on the prior, the conditional distribution of $\sqrt{n}(\theta - \hat{\theta})|\mathbf{X}$ is asymptotically $N(0, V^2)$, where $\hat{\theta}$ is an efficient estimator (for example, the maximum likelihood estimator) with a variance V^2 that attains the Cramér–Rao bound. As a consequence, posterior-based inference asymptotically coincides with inference based on frequentist standard efficient, $\frac{1}{\sqrt{n}}$-consistent estimators $\hat{\theta}$, giving asymptotic efficiency of Bayesian methods.

In Bayesian nonparametric framework, it is natural to ask whether the Bernstein-von Mises theorem still holds true, as it would give a further justification for the use of Bayesian nonparametric models, for example, in the construction of credible sets. The nonparametric maximum likelihood estimator of P_0 is well-known to be the empirical process $\mathbb{P}_n = \frac{\sum_{i=1}^n \delta_{X_i}}{n}$ (van der Vaart and Wellner, 1996; Vaart, 1998; Shao, 2003), and the limit law of $\sqrt{n}(\mathbb{P}_n - P_0)$ is normal distribution. The Bernstein-von Mises theorem in this setting is to give the limit law of the posterior distribution of $\sqrt{n}(P - \mathbb{P}_n)$ given $\mathbf{X}$ by the normal distribution obtained as the limit law of $\sqrt{n}(\mathbb{P}_n - P_0)$. More generally, we temporarily let $P \in \mathbb{M}_{\mathbb{X}}$ be any random probability measure and define the functional as follows:

$$Pf = \int f dP, \qquad P_0 f = \int f dP_0, \qquad \mathbb{P}_n f = \int f d\mathbb{P}_n = \frac{\sum_{i=1}^n f(X_i)}{n},$$

where $f : \mathbb{X} \to \mathbb{R}$ is any measurable function.

Let $\mathbb{F}$ be a collection of functions f, the Bernstein-von Mises theorem in the Bayesian nonparametric case considers the distribution of $\{\sqrt{n}(Pf - \mathbb{P}_n f)|\mathbf{X} : f \in \mathbb{F}\}$ and $\{\sqrt{n}(\mathbb{P}_n f - P_0 f) : f \in \mathbb{F}\}$. It is worth to point out that there have been many works for the weak convergence of stochastic processes indexed by elements of Banach space of functions, we refer the readers to van der Vaart and Wellner (1996); Vaart (1998) for further reading. When the collection $\mathbb{F}$ is finite, both $\{\sqrt{n}(Pf - \mathbb{P}_n f)|\mathbf{X} : f \in \mathbb{F}\}$ and $\{\sqrt{n}(\mathbb{P}_n f - P_0 f) : f \in \mathbb{F}\}$ are random vectors in Euclidean space. Otherwise, it is convenient to consider the $\mathbb{F}$ to be P_0-Donsker. Here we recall that $\mathbb{F}$ is P_0-Donsker if the sequence $\sqrt{n}(\mathbb{P}_n f - P_0 f)$ converges to $\mathbb{B}^o_{P_0}$ in distribution in the metric space $l^\infty(\mathbb{F})$ of bounded functions $g : \mathbb{F} \to \mathbb{R}$, equipped with the uniform norm $||g||_{\mathbb{F}} = \sup_{f \in \mathbb{F}} |g(f)|$. Here and throughout the paper, $\mathbb{B}^o_{P_0}$ is a Brownian bridge with parameter P_0 or P_0-Brownian bridge, namely, $\mathbb{E}[\mathbb{B}^o_{P_0} f] = 0$ and $\mathbb{E}[\mathbb{B}^o_{P_0} f_1 \mathbb{B}^o_{P_0} f_2] = P_0(f_1 f_2) - P_0 f_1 P_0 f_2$. A notable result is that a finite set $\mathbb{F}$ is P_0-Donsker if and only if $P_0 f^2 < \infty$ for every $f \in \mathbb{F}$. For the infinite P_0-Donsker classes, one can find details and examples in van der Vaart and Wellner (1996).

In order to define the weak convergence of $\sqrt{n}(P - \mathbb{P}_n)$ conditional on $\mathbf{X}$ to $\mathbb{B}^o_{P_0}$, we can use the conditional weak convergence in the bounded Lipschitz metric (van der Vaart and Wellner, 1996) as follows:

$$\sup_{h \in \mathrm{BL}_1} \left| \mathbb{E}\left[h(\sqrt{n}(P - \mathbb{P}_n))|\mathbf{X} \right] - \mathbb{E}[h(\mathbb{B}^o_{P_0})] \right| \to 0, \tag{4.1}$$

as $n \to \infty$. The expectation in (4.1) is taken for the random probability measure P, and thus the left side of (4.1) is a function of $\mathbf{X}$. The convergence in (4.1) refers to the iid sample $\mathbf{X}$ from P_0 and can be in probability or almost surely. The supremum is taken over the set BL_1 of all functions $h : l^\infty(\mathbb{F}) \to [0, 1]$ such that $|h(f_1) - h(f_2)| \leq ||f_1 - f_2||_{\mathbb{F}}$, for all $f_1, f_2 \in l^\infty(\mathbb{F})$. We denote the above convergence as

$$\sqrt{n}(P - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}. \tag{4.2}$$

Under the convergence criteria we explained above, we will present the Bernstein-von Mises theorem when $P \sim \mathrm{NGGP}(a, \sigma, \theta, H)$. For simplicity of interpretation, let $\tilde{\mathbb{P}}_n = \frac{\sum_{i=1}^{n(\pi)} \delta_{Y_i}}{n(\pi)}$.

Theorem 16. *Let $\mathbf{X}$ be a sample as defined in (1.1) with $P \sim NGGP(a, \sigma, \theta, H)$. Let $\mathbb{F}$ be the finite collection of functions such that $P_0 f^2 < \infty$ and $H f^2 < \infty$ for any $f \in \mathbb{F}$. We have the following convergences almost surely under P_0^∞.*

(i) If P_0 is discrete,

$$\sqrt{n}\left(P - \left[\mathbb{P}_n + \frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n)\right]\right)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}, \tag{4.3}$$

$$\sqrt{n}\left(P - \mathbb{E}[P|\mathbf{X}]\right)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}. \tag{4.4}$$

(ii) If P_0 is continuous,

$$\begin{aligned}\sqrt{n}\left(P - [(1-\sigma)\mathbb{P}_n + \sigma H]\right)|\mathbf{X}& \\ \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}^o_{P_0} &+ \sqrt{\sigma(1-\sigma)}\mathbb{B}^o_H + \sqrt{\sigma}Z(P_0 - H),\end{aligned} \tag{4.5}$$

$$\begin{aligned}\sqrt{n}(P - \mathbb{E}[P|\mathbf{X}])|\mathbf{X}& \\ \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}^o_{P_0} &+ \sqrt{\sigma(1-\sigma)}\mathbb{B}^o_H + \sqrt{\sigma}Z(P_0 - H).\end{aligned} \tag{4.6}$$

Here $\mathbb{B}^o_{P_0}$, $\mathbb{B}^o_H$ are independent Brownian bridges, independent of the standard normal random variable Z. Moreover, if $\mathbb{F}$ is any P_0-Donsker class of functions, then the convergences hold in probability in $l^\infty(\mathbb{F})$. In this case, the convergence is also P_0^∞-almost surely under an additional condition that $P_0||f - P_0 f||^2_{\mathbb{F}} < \infty$.

We refer to Theorem 2.11.1 and 2.11.9 in van der Vaart and Wellner (1996) for more details of the discussion for $\mathbb{F}$ such that the convergence holds in $l^\infty(\mathbb{F})$. It is worth noting that the Bernstein-von Mises results in Theorem 16 are not exactly in the form of standard Bernstein-von Mises theorem in (4.2).

When P_0 is continuous, there is a “bias” term $\sigma(H - \mathbb{P}_n)$ in the convergence result (4.5). Such “bias” term vanishes only when $\sigma = 0$, under which P becomes the Dirichlet process, or when $H = \mathbb{P}_n$ ($H = P_0$), which is unrealistic. Moreover, σ equals $\bar{C}_1$ in Theorem 6. Thus, it suggests that one is not expected to use NGGP for continuous P_0.

On the other hand, when P_0 is discrete, it is interesting to see that there is a “bias” term $\frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n)$ on the left hand side of the convergence result (4.3) to make the limiting process to be $\mathbb{B}^o_{P_0}$. That said, the convergence (4.3) is the usual form as in (4.2) as long as $\sqrt{n}\frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n) \to 0$. We can not drop this “bias” term directly, even if $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ a.s.. Apart from the case when $\sigma = 0$, the “bias” term can be dropped when $\lim_{n\to\infty} \frac{n(\pi)}{\sqrt{n}} = 0$, in the sense that the number of atoms $\{x_j\}$ in P_0 should decrease fast enough when $n \to \infty$. For a formal condition of P_0 to make $\lim_{n\to\infty} \frac{n(\pi)}{\sqrt{n}} = 0$, we have the following proposition.

Proposition 17. *Under the conditions in Theorem 16, when P_0 is discrete, we have the following results.*

(i) If $P_0(\{x_j\}) \le \frac{C}{j^\alpha}$, for some positive constant C and $\alpha > 2$ and $\mathbb{F}$ is the class of uniformly bounded functions, then $\sqrt{n}(P - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ in probability in $l^\infty(\mathbb{F})$.

(ii) If the function $h(t) := \#\{x : P_0(\{x\}) \geq \frac{1}{t}\}$ is regularly varying at ∞ of exponent η with $\eta < \frac{1}{2}$ and $\mathbb{F}$ is the class of uniformly bounded functions, then $\sqrt{n}(P - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ a.s. in $l^\infty(\mathbb{F})$.

(iii) If $\mathbb{F}$ is a class of functions f such that $f(\{x_j\}) \asymp j^\beta$ for some $\beta > 0$ and $P_0(\{x_j\}) \leq \frac{C}{j^\alpha}$, for some positive constant C and $\alpha > 2 + 2\beta$, then $\sqrt{n}(P - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ in probability in $l^\infty(\mathbb{F})$.

The proof of Proposition 17 follows directly from Corollary 2 in Franssen and van der Vaart (2022) and Theorem 9 in Karlin (1967). Here we recall that if h is regularly varying at ∞ with exponent $\eta \in (0, 1)$, then for any $t > 0$, we have $\lim_{n\to\infty} \frac{h(nt)}{h(n)} = t^\eta$. Moreover, for such regularly varying function h, we have $\frac{n(\pi)}{h(n)} \to \Gamma(1 - \eta)$ a.s., and $h(n)$ is n^η up to a slowly varying factor. We refer to the appendix in Haan and Ferreira (2006) and Bingham et al. (1987) for more details of the regularly varying function. The parameter α in Proposition 17 is related to the number of clusters in the population, larger α means less clusters in the population. In fact, it controls the order of $n(\pi)$ (see Theorems 1, 9 and Example 4 in Karlin, 1967) relative to n. More precisely, $\lim_{n\to\infty} \frac{n(\pi)}{n^{1/\alpha}} = C$, where C is some constant. With a large α, observations from P_0 would concentrate to the atoms with small indices, and $n(\pi)$ would have a small order with respect to n.

As one application of the Bernstein-von Mises results in Theorem 16, we may construct Bayesian credible sets for Pf when $n \to \infty$. The choices of f determine the parameters, for which the credible sets are constructed. For example, if $f(x) = x$, the credible interval is for the mean. Since the posterior consistency does not hold for the case when P_0 is continuous, the credible sets for Pf is not correct in this case, thus we shall only give the credible sets for Pf when P_0 is discrete.

Corollary 18. *If P_0 is discrete, under the conditions in Theorem 16, we have the probability of $P_0 f \in (L_{n,\alpha} f - \frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f), L_{n,\beta} f - \frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f))$ is $\beta - \alpha$ for any f such that $P_0 f^2 < \infty$ and $Hf^2 < \infty$. Here $L_{n,\alpha}$ is the α-quantile of the posterior distribution of $Pf|\mathbf{X}$ and $\beta > \alpha$.*

The credible interval in Corollary 18 holds for any discrete P_0. A straightforward consequence of Proposition 17 is that if $\frac{n(\pi)}{\sqrt{n}} \to 0$ in probability, then the "bias" term vanishes and therefore the credible interval for $P_0 f$ becomes the usual form $(L_{n,\alpha} f, L_{n,\beta} f)$. This is true under the restrictive assumptions of P_0 in Proposition 17. However, these assumptions of P_0 are not realistic since the "true" distribution P_0 is unknown. Thus, one can always keep the correction term $\frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f)$ as a bias correction to construct credible intervals as we state in Corollary 18. Furthermore, the bias correction is necessary when the "bias" term does not vanish. We provide a numerical illustration corresponding to this scenario in Section 4.1.

As we have mentioned, P_0 is of course unknown in real application and we shall consider Theorem 16 without the information from P_0. In this case, one needs to pay special attention to the parameter σ, which plays a remarkable role in determining the number of clusters $n(\pi)$ of a sample from the NGGP. As shown in Proposition 3 in

Lijoi et al. (2007b), the number of clusters $n(\pi)$ is of the type n^σ asymptotically with respect to the NGGP prior. That said, similarly as we explained for the parameter α of P_0 in Proposition 17, the larger σ the larger the growth rate of the number of clusters $n(\pi)$ with respect to the NGGP prior. It is easy to see from both Theorem 6 and Theorem 16 that if $\sigma \to 0$, P is posterior consistent and the Bernstein-von Mises results hold without the bias terms for any P_0. But this corresponds to the case that P becomes the Dirichlet process. Thus, one should at least expect the parameter σ to be small. Usually, the model parameters are chosen by the empirical Bayesian method, and people can estimate the model parameters by using the maximum likelihood estimators conditional on the observations $\mathbf{X}$. A well known conclusion in Bayesian nonparametric framework is the observation $\mathbf{X}$ from NRMIs induces a random partition structure for $\{1, \dots, n\}$ as we introduced in Section 2.2. The random partition structure is characterized by the exchangeable partition probability function (EPPF) (Pitman, 1995), which also plays the rule as the likelihood function of σ. The idea of using the EPPF as a likelihood function for empirical Bayesian estimation of parameters of Gibbs-type priors has been introduced in Lijoi et al. (2007a), in which the parameters of the two-parameter Poisson-Dirichlet process are estimated in such procedure in the analysis of genomic data. Similar use of EPPFs as likelihood functions can be found and interpreted in various works (e.g., Favaro and Naulet, 2023; Ghosal and Van der Vaart, 2017; Franssen and van der Vaart, 2022). An alternative way of obtaining empirical Bayesian estimation of parameters of the two-parameter Poisson-Dirichlet process is to place a specific prior to the parameters as introduced in Lijoi et al. (2008). The EPPF for the NGGP is given as

$$\Pi_\sigma(n_1, \dots, n_{n(\pi)}) = \frac{\prod_{j=1}^{n(\pi)} (1-\sigma)_{(n_j-1)}}{\Gamma(n)} \int_0^\infty u^{n-1} (u+\theta)^{n(\pi)\sigma - n} e^{\frac{a}{\sigma}((u+\theta)^\sigma - \theta^\sigma} du,$$

where $(1-\sigma)_{(n_j-1)} = \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}$. From Theorem 1 in Favaro and Naulet (2023), the maximum likelihood estimator $\hat{\sigma}_n$ exists uniquely. Furthermore, the results in Theorem 2 in Favaro and Naulet (2023) implies that $\hat{\sigma}_n \to \sigma_0$ in probability with a rate $\sqrt{\log(n)} n^{-\frac{\sigma_0}{2}}$, when P_0 is discrete with atoms x satisfying $h(t) = \#\{x : P_0(\{x\}) \geq \frac{1}{t}\}$ is a regularly varying function of exponent $\sigma_0 \in [0, 1)$. In this case, the number of clusters of a sample $\mathbf{X}$ from P_0 is of type n^{σ_0} asymptotically. If the sample $\mathbf{X}$ is assumed to be a sample from the NGGP with parameter σ, the number of clusters $n(\pi)$ is of type n^σ asymptotically. Thus, the coefficient of regular variation σ_0 is the true value of σ, and the parameter σ can be estimated by the maximum likelihood estimator $\hat{\sigma}_n$.

Theorem 19. *Under the assumptions in Theorem 16, we have the following results.*

(i) If $\hat{\sigma}_n$ is an estimator based on $\mathbf{X}$ that converges to σ_0 in probability, then the convergences in Theorem 16 hold in probability by replacing σ_n with $\hat{\sigma}_n$ and replacing σ with σ_0. In particular, this is true for the maximum likelihood estimator $\hat{\sigma}_n$, if P_0 is discrete with atoms x satisfying the condition that $h(t) = \#\{x : P_0(\{x\}) \geq \frac{1}{t}\}$ is a regularly varying function of exponent $\sigma_0 \in [0, 1)$.

(ii) If $\sigma \sim \mathcal{L}_\sigma$, where $\mathcal{L}_\sigma$ is a probability law on $[0, 1]$ that plays the prior distribution of σ, then the Bayesian model becomes

$$\mathbf{X}|P, \sigma \sim P; \quad P|\sigma \sim NGGP(a, \sigma, \theta, H)$$

The convergences in Theorem 16 hold by replacing σ_n with σ on the left hand side, and replacing σ with σ_0 on the limiting processes. The σ on the left hand side is the posterior random variable.

It is worth to point out that the parameter σ_0 in the assumption of discrete P_0 in (i) of the last theorem controls the number of clusters of atoms of P_0 as shown in Karlin (1967). More precisely, the number of clusters $n(\pi)$ of observations from P_0 is of the type n^{σ_0} asymptotically (see Theorem 1 and 9 in Karlin, 1967). The proof of the above theorem follows the same procedures as the proof in Section 4.2 of Franssen and van der Vaart (2022). For the posterior consistency of $\hat{\sigma}_n$, we refer to the details with proofs in Section 4.3 of Franssen and van der Vaart (2022). The maximum likelihood estimator is not quite interesting as $\hat{\sigma}_n \to \sigma_0$ with $\sigma_0 = 1$ when P_0 is continuous, and $\sigma_0 \neq 0$ when P_0 is discrete (Favaro and Naulet, 2023).

Besides the parameter σ, the parameters a and θ do not appear in the asymptotic results in Theorem 6 and Theorem 16, and thus estimators of a and θ based on prior distributions or maximum likelihood method will not affect the convergences when $a \ll \sqrt{n}$ and $\theta \ll n^{\sigma}$. The cases when $\hat{a}_n$ and $\hat{\theta}_n$ converge to ∞ as $n \to \infty$ are not usual and beyond the scope of this work and can be considered in the future works.

4.1 Numerical illustration

We present the credible intervals for $P_0 f$ when P_0 is discrete with different types of the number of atoms. To be more precise, let $P_0 f = P_0([2,\infty])$ for $P_0 = P_1, P_2, P_3, P_4$, where we describe P_1, P_2, P_3, P_4 on $\mathbb{Z}^+$ as follows.

$$P_1(X=1)=0.2,\ P_1(X=2)=0.2,\ P_1(X=3)=0.2,\ P_1(X=4)=0.3,\ P_1(X=5)=0.1,$$
$$P_2(X=k) \propto k^{-3}, \qquad P_3(X=k) \propto k^{-2}, \qquad P_4(X=k) \propto k^{-\frac{3}{2}}.$$

Obviously, $n(\pi) = 5$ for P_1. From the result (see e.g., Example 4) in Karlin (1967), we have the regularly varying functions $h(t)$ corresponding to P_2, P_3, P_4 are proportional to $t^{\frac{1}{3}}, t^{\frac{1}{2}}, t^{\frac{2}{3}}$ respectively. Moreover, when $n \to \infty$, the distinct numbers $n(\pi)$ of P_2, P_3, P_4 are proportional to $n^{\frac{1}{3}}, n^{\frac{1}{2}}, n^{\frac{2}{3}}$, respectively, from Theorem 1 in Karlin (1967). Thus, the "bias" term for P_1, P_2, P_3, P_4 goes to $0, 0$, some constant, ∞, respectively.

For the NGGP, we let $P \sim \mathrm{NGGP}(1, \sigma = 0.5, 1, H)$ (namely, the normalized inverse-Gaussian process), where H is standard normal distribution. We simulate P through its stick-breaking representation (Favaro et al., 2012) with the generating algorithm in Favaro et al. (2016). To make sure the simulation of $P = \sum_{i=1}^{\infty} w_i \delta_{X_i}$ is accurate, we truncate the infinite sum at some N such that the weight of the tail $\sum_{i=N}^{\infty} w_i < \frac{1}{\sqrt{n}}$, where n is the sample size. We simulate 10000 replications of the sample $\mathbf{X}$ from P_1, P_2, P_3, P_4 with the sample size $n = 10, 100, 1000, 10000, 100000$ respectively. For each sample from P_1, we construct one 95% credible interval for $P_1([2,\infty))$ with the "bias" correction as stated in Corollary 18 and compute the proportion that the true value $P_1([2,\infty))$ belongs to the intervals of 10000 replications. We also compute the same proportion without the "bias" correction. The results of P_1, P_2, P_3, P_4 are given in Tables 1 and 2.

n	10	100	1000	10000	100000
P_1	0.791	0.952	0.961	0.967	0.986
P_2	0.695	0.857	0.928	0.917	0.931
P_3	0.712	0.785	0.811	0.727	0.754
P_4	0.601	0.292	0.078	0.000	0.000

Table 1: Proportion of coverage of the true value for the 95% credible interval without "bias" correction.

n	10	100	1000	10000	100000
P_1	0.977	0.989	0.991	0.995	0.997
P_2	0.914	0.938	0.951	0.933	0.941
P_3	0.863	0.931	0.962	0.960	0.978
P_4	0.901	0.955	0.969	0.966	0.956

Table 2: Proportion of coverage of the true value for the 95% credible interval with "bias" correction.

Since the "bias" terms for P_1 and P_2 vanish as $n \to \infty$, the proportions of the coverage of the true value are large for both with and without "bias" correction. However, the 95% credible intervals for $P_3 f$ and $P_4 f$ are not performing good without "bias" correction. Thus, the credible intervals with "bias" correction as stated in Corollary 18 work well for all types of discrete P_0.

As for the normality convergence, we draw the marginal density plots in Figure 1 for $P_1([2,\infty))$ given sample $\mathbf{X}$ with size $n = 10, 100, 1000, 10000, 100000$ respectively. Both plots are generated from 1000000 replicates, the true mean of $P_1([2,\infty))$ is 0.8 The marginal density for $P_1([2,\infty))$ is skewed when $n = 10, 100$, and symmetric when $n = 1000$ and larger.

5 Discussion

To the best of our knowledge, the jump component of the Lévy intensities of the well-studied NRMIs up-to-date are given in the form of the gamma density: $s^{-\sigma-1}e^{-\beta s}$. It turns out that with the shape parameter $\sigma = 0$, the posterior consistency is always guaranteed for any "true" distribution P_0. Otherwise, the posterior consistency only holds for discrete P_0 but not for continuous P_0. Such phenomenon does naturally make sense due to the discreteness of NRMIs. If P_0 is diffuse and the prior guess for the sample distribution $\alpha \neq P_0$, the prior information will always contribute to the posterior, no matter how large is the sample size. In such sense, the Bayesian nonparametric models never behave "better" than the empirical models asymptomatically when P_0 is diffuse. As we have explained in the end of Section 3, with continuous data, one would hardly use P as a prior of the data distribution and other nonparametric models should be considered (e.g. mixture Bayesian nonparametric models (Lo, 1984; Escobar and West, 1995; Müller and Quintana, 2004; Gershman and Blei, 2012), Gaussian process models (Gershman and Blei, 2012; Seeger, 2004; Williams and Rasmussen, 2006)). On the other hand, we are not able to know the "true" distribution of a given sample with any size n,

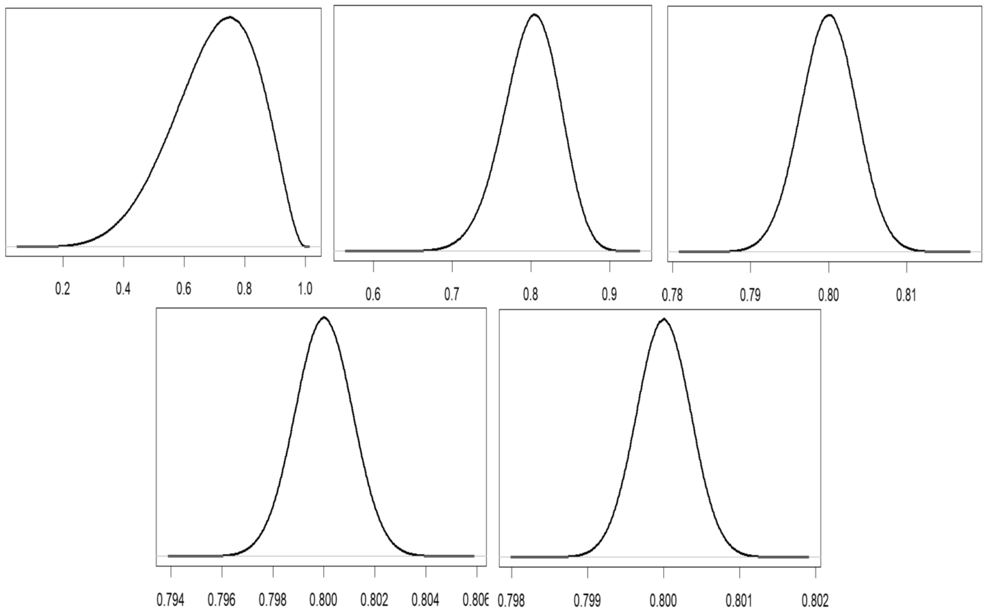

Figure 1: The marginal densities for $P_1([2,\infty))$ with sample size $n = 10, 100, 1000, 10000$, 100000 follow the order from top left to bottom right.

also the sample size n will never be ∞, a prior guess of the random probability measure based on experience could make the model suitable. Furthermore, the mixture and hierarchical Bayesian nonparametric models (Lo, 1984; Escobar and West, 1995) based on NRMIs are showing great success in the applications and consistency behaviours (Ghosal et al., 1999; Lijoi et al., 2005c).

Importantly, the posterior consistency result of NRMIs when P_0 is discrete in this work provides strong theoretical support of using NRMIs. The results in this work also provide a guideline of choosing the proper intensity $\rho(ds|x)$, for example, the generalized Dirichlet process and the generalized extended gamma NRMI are good choices in Bayesian nonparametric applications and they both show some flexibilities. Besides, we may let $\sigma \to 0$ by assigning a randomness on σ, or one may construct α to depend on $\rho(ds|x)$ to deduct $\bar{C}_1$. Moreover, we shall develop other subclasses of NRMIs and other NRMIs like classes that are both flexible and satisfying posterior consistency property to make the Bayesian nonparametric class rich.

Bayesian nonparametric mixture models have been successfully applied in probabilistic clustering and density estimation. One interesting open question is whether consistency for the number of clusters of Bayesian nonparametric mixture models can be ensured. Given the fact that the Dirichlet process mixture model and the two-parameter Poisson-Dirichlet process mixture model are not consistent for the number of clusters when the observed data are generated from a finite mixture and the concentration parameter a is fixed (Miller and Harrison, 2013, 2014), this query may be explored for the following two circumstances: 1) when the concentration parameter converges to 0 at an appropriate rate as n goes to infinity; 2) when the concentration parameter is estimated by Bayesian approaches.

Due to the complexity of the posterior of NRMIs, it is not easy to present a Bernstein-von Mises like result to give the limiting process of the posterior of general NRMIs. The result for the normalized generalized gamma process, along with the works in Lo (1983, 1986); Ray and van der Vaart (2021); Hu and Zhang (2022); James (2008); Franssen and van der Vaart (2022), shed some light in discovering the Bernstein-von Mises theorem for general NRMIs.

Supplementary Material

Supplementary Material of "Large Sample Asymptotic Analysis for Normalized Random Measures with Independent Increments" contains all proofs of the results provided in the main paper (DOI: 10.1214/23-BA1411SUPP; .pdf).

References

Aldous, D. J. (1985). "Exchangeability and related topics." In Hennequin, P. L. (ed.), *École d'Été de Probabilités de Saint-Flour XIII — 1983*, 1–198. Berlin: Springer. MR0883646. doi: https://doi.org/10.1007/BFb0099421. 2, 6

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems." *The Annals of Statistics*, 1152–1174. MR0365969. 11

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge. MR0898871. doi: https://doi.org/10.1017/CBO9780511721434. 16

Daley, D. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and its Applications. Springer, New York, second edition. MR2371524. doi: https://doi.org/10.1007/978-0-387-49835-5. 4, 5

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prunster, I., and Ruggiero, M. (2015). "Are Gibbs-type priors the most natural generalization of the Dirichlet process?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2): 212–229. 2

De Blasi, P., Lijoi, A., and Pruenster, I. (2013). "An asymptotic analysis of a class of discrete nonparametric priors." *Statistica Sinica*, 23(3): 1299–1322. MR3114715. 2, 3, 10

De Finetti, B. (1937). "La prévision: ses lois logiques, ses sources subjectives." *Annales de l'institut Henri Poincaré*, 7(1): 1–68. MR1508036. 6

Dykstra, R. and Laud, P. (1981). "A Bayesian nonparametric approach to reliability." *The Annals of Statistics*, 9(2): 356–367. MR0606619. 12

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90(430): 577–588. MR1340510. 19, 20

Favaro, S., Lijoi, A., Nava, C., Nipoti, B., Prünster, I., and Teh, Y. W. (2016). "On the stick-breaking representation for homogeneous NRMIs." *Bayesian Analysis*, 11(3): 697–724. MR3498043. doi: https://doi.org/10.1214/15-BA964. 18

Favaro, S., Lijoi, A., and Prünster, I. (2012). "On the stick-breaking representation of normalized inverse Gaussian priors." *Biometrika*, 99(3): 663–674. MR2966776. doi: https://doi.org/10.1093/biomet/ass023. 18

Favaro, S. and Naulet, Z. (2023). "Near-optimal estimation of the unseen under regularly varying tail populations." *Bernoulli*, 29(4): 3423–3442. MR4632144. doi: https://doi.org/10.3150/23-bej1589. 17, 18

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1(2): 209–230. MR0350949. 1, 11

Ferguson, T. S. (1974). "Prior distributions on spaces of probability measures." *The Annals of Statistics*, 2(4): 615–629. MR0438568. 10

Franssen, S. and van der Vaart, A. (2022). "Bernstein-von Mises theorem for the Pitman-Yor process of nonnegative type." *Electronic Journal of Statistics*, 16(2): 5779–5811. MR4505382. doi: https://doi.org/10.1214/22-ejs2077. 2, 4, 16, 17, 18, 21

Freedman, D. and Diaconis, P. (1983). "On inconsistent Bayes estimates in the discrete case." *The Annals of Statistics*, 1109–1118. MR0720257. doi: https://doi.org/10.1214/aos/1176346325. 2

Gershman, S. J. and Blei, D. M. (2012). "A tutorial on Bayesian nonparametric models." *Journal of Mathematical Psychology*, 56(1): 1–12. MR2903470. doi: https://doi.org/10.1016/j.jmp.2011.08.004. 19

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (1999). "Posterior consistency of Dirichlet mixtures in density estimation." *The Annals of Statistics*, 27(1): 143–158. MR1701105. doi: https://doi.org/10.1214/aos/1018031105. 20

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge: Cambridge University Press, first edition. MR3587782. doi: https://doi.org/10.1017/9781139029834. 2, 17

Gnedin, A. and Pitman, J. (2006). "Exchangeable Gibbs partitions and Stirling triangles." *Journal of Mathematical Sciences*, 138: 5674–5685. MR2160320. doi: https://doi.org/10.1007/s10958-006-0335-z. 2

Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, volume 3. Springer. MR2234156. doi: https://doi.org/10.1007/0-387-34471-3. 16

Hjort, N. L. (1990). "Nonparametric Bayes estimators based on Beta processes in models for life history data." *The Annals of Statistics*, 18(3): 1259–1294. MR1062708. doi: https://doi.org/10.1214/aos/1176347749. 10

Ho Jang, G., Lee, J., and Lee, S. (2010). "Posterior consistency of species sampling priors." *Statistica Sinica*, 20(2): 581–593. MR2682631. 2, 3, 10

Hu, Y. and Zhang, J. (2022). "Functional central limit theorems for stick-breaking priors." *Bayesian Analysis*, 17(4): 1101–1120. MR4506023. doi: https://doi.org/10.1214/21-ba1290. 2, 3, 21

James, L. F. (2002). "Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics." *arXiv preprint*, arXiv:math/0205093. 10

James, L. F. (2008). "Large sample asymptotics for the two-parameter Poisson–Dirichlet process." In *Clarke, B. and Ghosal, S. (eds.), Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, 187–199. Cambridge, MA: MIT Press. MR2459225. doi: https://doi.org/10.1214/074921708000000147. 2, 4, 21

James, L. F., Lijoi, A., and Prünster, I. (2006). "Conjugacy as a distinctive feature of the Dirichlet process." *Scandinavian Journal of Statistics*, 33(1): 105–120. MR2255112. doi: https://doi.org/10.1111/j.1467-9469.2005.00486.x. 9, 10

James, L. F., Lijoi, A., and Prünster, I. (2009). "Posterior analysis for normalized random measures with independent increments." *Scandinavian Journal of Statistics*, 36(1): 76–97. MR2508332. doi: https://doi.org/10.1111/j.1467-9469.2008.00609.x. 3, 6, 7, 10, 11

James, L. F., Lijoi, A., and Prünster, I. (2010). "On the posterior distribution of classes of random means." *Bernoulli*, 16(1): 155–180. MR2648753. doi: https://doi.org/10.3150/09-BEJ200. 3, 10

Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*, volume 9. Springer. MR2161313. 6

Karlin, S. (1967). "Central limit theorems for certain infinite urn schemes." *Journal of Mathematics and Mechanics*, 17(4): 373–401. MR0216548. doi: https://doi.org/10.1512/iumj.1968.17.17020. 16, 18

Kingman, J. (1967). "Completely random measures." *Pacific Journal of Mathematics*, 21(1): 59–78. MR0210185. 3, 4

Kingman, J. F. (1975). "Random discrete distributions." *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(1): 1–15. MR0368264. 1, 11

Kingman, J. F. C. (1993). *Poisson Processes*. Oxford Studies in Probability, volume 3. Clarendon Press, Oxford University Press, New York. MR1207584. 3, 4

Korwar, R. M. and Hollander, M. (1973). "Contributions to the theory of Dirichlet processes." *The Annals of Probability*, 1(4). MR0350950. doi: https://doi.org/10.1214/aop/1176996898. 11

Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer Science & Business Media. MR0856411. doi: https://doi.org/10.1007/978-1-4612-4946-7. 13

Lijoi, A., Mena, R. H., and Prünster, I. (2005a). "Bayesian nonparametric analysis for a generalized Dirichlet process prior." *Statistical Inference for Stochastic Processes*, 8(3): 283–309. MR2177315. doi: https://doi.org/10.1007/s11203-005-6071-z. 2, 12

Lijoi, A., Mena, R. H., and Prünster, I. (2005b). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: https://doi.org/10.1198/016214505000000132. 1, 11

Lijoi, A., Mena, R. H., and Prünster, I. (2007a). "Bayesian nonparametric estimation of the probability of discovering new species." *Biometrika*, 94(4): 769–786. MR2416792. doi: https://doi.org/10.1093/biomet/asm061. 17

Lijoi, A., Mena, R. H., and Prünster, I. (2007b). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 715–740. MR2370077. doi: https://doi.org/10.1111/j.1467-9868.2007.00609.x. 1, 11, 17

Lijoi, A., Mena, R. H., and Prünster, I. (2008). "A Bayesian nonparametric approach for comparing clustering structures in EST libraries." *Journal of Computational Biology*, 15(10): 1315–1327. MR2461978. doi: https://doi.org/10.1089/cmb.2008.0043. 17

Lijoi, A. and Prünster, I. (2003). "On a normalized random measure with independent increments relevant to Bayesian nonparametric inference." In *13th European Young Statisticians Meeting*, 123–134. Bernoulli Society. 1, 11

Lijoi, A. and Prünster, I. (2010). "Models beyond the Dirichlet process." In Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (eds.), *Bayesian Nonparametrics*, volume 28, chapter 3, 80–163. Cambridge University Press. MR2730661. 2, 4, 10

Lijoi, A., Prünster, I., and Walker, S. G. (2005c). "On consistency of nonparametric normal mixtures for Bayesian density estimation." *Journal of the American Statistical Association*, 100(472): 1292–1296. MR2236442. doi: https://doi.org/10.1198/016214505000000358. 20

Lo, A. Y. (1982). "Bayesian nonparametric statistical inference for Poisson point processes." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59(1): 55–66. MR0643788. doi: https://doi.org/10.1007/BF00575525. 12

Lo, A. Y. (1983). "Weak convergence for Dirichlet processes." *Sankhyā: The Indian Journal of Statistics, Series A*, 105–111. MR0749358. 2, 3, 21

Lo, A. Y. (1984). "On a class of Bayesian nonparametric estimates: I. Density estimates." *The Annals of Statistics*, 351–357. MR0733519. doi: https://doi.org/10.1214/aos/1176346412. 19, 20

Lo, A. Y. (1986). "A remark on the limiting posterior distribution of the multiparameter Dirichlet process." *Sankhyā: The Indian Journal of Statistics, Series A*, 247–249. MR0905464. 2, 3, 21

Miller, J. W. and Harrison, M. T. (2013). "A simple example of Dirichlet process mixture inconsistency for the number of components." In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 1, 199–206. NY: Curran Associates, Inc. MR3277163. 20

Miller, J. W. and Harrison, M. T. (2014). "Inconsistency of Pitman-Yor process mixtures for the number of components." *The Journal of Machine Learning Research*, 15(1): 3333–3370. MR3277163. 20

Müller, P. and Quintana, F. A. (2004). "Nonparametric Bayesian data analysis." *Statistical Science*, 19(1): 95–110. MR2082149. doi: https://doi.org/10.1214/088342304000000017. 2, 19

Perman, M., Pitman, J., and Yor, M. (1992). "Size-biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: https://doi.org/10.1007/BF01205234. 2

Pitman, J. (1995). "Exchangeable and partially exchangeable random partitions." *Probability Theory and Related Fields*, 102(2): 145–158. MR1337249. doi: https://doi.org/10.1007/BF01213386. 17

Pitman, J. (1996). "Some developments of the Blackwell-MacQueen urn scheme." In *Statistics, Probability and Game Theory, Lecture Notes-Monograph Series*, volume 30, 245–267. Hayward, CA. MR1481784. doi: https://doi.org/10.1214/lnms/1215453576. 2

Pitman, J. (2003). "Poisson-Kingman partitions." In *Statistics and Science: a Festschrift for Terry Speed*. Lecture Notes-Monograph Series, volume 40, 1–34. Beachwood, OH. MR2004330. doi: https://doi.org/10.1214/lnms/1215091133. 10, 11

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 25(2): 855–900. MR1434129. doi: https://doi.org/10.1214/aop/1024404422. 2

Ray, K. and van der Vaart, A. (2021). "On the Bernstein-von Mises theorem for the Dirichlet process." *Electronic Journal of Statistics*, 15: 2224–2246. MR4255307. doi: https://doi.org/10.1214/21-ejs1821. 2, 3, 21

Regazzini, E., Lijoi, A., and Prünster, I. (2003). "Distributional results for means of normalized random measures with independent increments." *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: https://doi.org/10.1214/aos/1051027881. 1, 4, 6

Seeger, M. (2004). "Gaussian processes for machine learning." *International Journal of Neural Systems*, 14(02): 69–106. 19

Shao, J. (2003). *Mathematical Statistics*. Springer Science & Business Media. 13

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. MR1652247. doi: https://doi.org/10.1017/CBO9780511802256. 13, 14 van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York. MR1385671. doi: https://doi.org/10.1007/978-1-4757-2545-2. 13, 14, 15

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*, volume 2. MIT Press, Cambridge, MA. MR2514435. 19

Zhang, J. and Hu, Y. (2021). "Dirichlet process and Bayesian nonparametric models (in Chinese)." *Scientia Sinica Mathematica*, 51(11): 1895–1932. 2

Zhang, J. and Hu, Y. (2024). "Supplementary Material for "Large sample asymptotic analysis for normalized random measures with independent increments"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1411SUPP. 4

Acknowledgments

We sincerely thank the associate editor and anonymous referees for their constructive and inspiring comments which improved the quality of the article significantly.