

A unified construction for series representations and finite approximations of completely random measures

JUHO LEE¹, XENIA MISOURIDOU² and FRANÇOIS CARON³

¹*Graduate school of AI, KAIST, Daejeon, Republic of Korea*

E-mail: [*juholee@kaist.ac.kr](mailto:juholee@kaist.ac.kr)

²*Department of Mathematics, Imperial College London*

E-mail: [†xeniamisc@gmail.com](mailto:xeniamisc@gmail.com)

³*Department of Statistics, University of Oxford*

E-mail: [‡caron@stats.ox.ac.uk](mailto:caron@stats.ox.ac.uk)

Infinite-activity completely random measures (CRMs) have become important building blocks of complex Bayesian nonparametric models. They have been successfully used in various applications such as clustering, density estimation, latent feature models, survival analysis or network science. Popular infinite-activity CRMs include the (generalised) gamma process and the (stable) beta process. However, except in some specific cases, exact simulation or scalable inference with these models is challenging and finite-dimensional approximations are often considered. In this work, we propose a general and unified framework to derive both series representations and finite-dimensional approximations of CRMs. Our framework can be seen as a generalisation of constructions based on size-biased sampling of Poisson point process [55]. It includes as special cases several known series representations and finite approximations as well as novel ones. In particular, we show that one can get novel series representations for the generalised gamma process and the stable beta process. We show how these construction can be used to derive novel algorithms for posterior inference, including a generalisation of the slice sampler for normalised CRMs mixture models introduced by [29]. We also provide some analysis of the truncation error.

Keywords: Bayesian nonparametrics; Poisson Random Measures; Completely Random Measure; Generalised Gamma Process; Slice sampling

1. Introduction

Probabilistic models with infinite-dimensional parameters arise as building blocks of numerous modern structured statistical models. Of particular interest are models building on infinite-activity completely random measures (CRMs), and more generally functionals of infinite-activity Poisson random measures. Examples include clustering and density estimation [56, 35, 37], spatial statistics [11, 66, 48], latent factor/trait models [30, 53, 69, 13, 5], network modeling [16, 68, 12, 17], recommendation systems [27], prediction, risk management and option pricing of financial assets [18] or survival analysis [31, 51]; see [46] for a review. Popular CRMs include the (generalised) gamma random measure (also known as tempered stable) [32, 11] or the (stable) beta random measure [31, 62]. Some standard random probability measures are also obtained by normalisation (Dirichlet process) or transformation (Pitman-Yor process) of CRMs.

The use of statistical models based on infinite-activity CRMs poses a number of practical challenges regarding posterior inference and estimation. Except in some specific cases where the infinite-

dimensional parameter can be integrated out, most algorithms, either based on Gibbs sampling [33, 68], slice sampling [49, 29, 25], mean-field variational inference [9, 20, 43] or sequential Monte Carlo [6, [1, Section 3.2.], [28], require the use of a finite-dimensional approximation of the CRM. Finite-dimensional approximations can either be obtained by (i) truncating a series representation of the CRM, with almost surely or stochastically decreasing weights, or (ii) by considering a finite measure with n atoms and iid weights, converging in distribution to the CRM as n tends to infinity. Consider for example the beta process with scale parameter $\alpha > 0$ and probability distribution H . Its inverse Lévy series representation is given by [63]

$$G = \sum_{i=1}^{\infty} W_i \delta_{\theta_i} \quad (1.1a)$$

where, for $i = 1, 2, \dots$,

$$W_i = \prod_{j=1}^i \beta_j, \quad \beta_i \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, 1), \quad \theta_i \stackrel{\text{i.i.d.}}{\sim} H. \quad (1.1b)$$

A standard finite-dimensional approximation of the beta process with iid weights is [30]

$$\tilde{G}_n = \sum_{i=1}^n \tilde{W}_{n,i} \delta_{\tilde{\theta}_{n,i}} \quad (1.2a)$$

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

$$\tilde{W}_{n,i} \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(\alpha/n, 1), \quad \tilde{\theta}_{n,i} \stackrel{\text{i.i.d.}}{\sim} H. \quad (1.2b)$$

Both representations are routinely used in Markov chain Monte Carlo and variational Bayes approximate inference algorithms [20, 54, 52]. Series and iid approximations are similarly used for the gamma process [68, 61] and the generalised gamma process [43]. Since the early work of [39] and [24] on the so-called inverse Lévy representation, various generic series representations of Poisson random measures have been proposed [10, 55, 57, 58, 14].

The objective of this article is to present a general framework to obtain both series and iid representations of CRMs. Our construction builds on the definition of a Poisson random measure on an extended space; it generalises the size-biased approach of [55], and admits as special cases both the size-biased and inverse-Lévy representations. We show that under this construction, one can draw connections between existing series and iid representations that appeared unrelated, and it allows to derive new series and iid representations. More precisely, we show how the iid representation of [43] is related to the size-biased construction of [55], derive novel series and iid representations of the generalised gamma and stable beta random measures. In particular, we provide a novel tractable series representation and iid approximation for the generalised gamma process, which admits as limit cases both the size-biased and inverse Lévy representations of this process. Additionally we explain how we can use our framework to (i) develop novel slice sampling algorithms for exact posterior inference, generalising the sampler of [29] and (ii) to derive Gibbs samplers for approximate inference using the iid representation. Finally, we provide an asymptotic analysis of the truncation error for this class of approximations.

This article is organised as follows. In [Section 2](#) we provide background material on completely random measures and some existing series representations for CRMs and describe the objectives. [Section 3](#) describes the general construction for obtaining series and iid representations of CRMs. In [Section 4](#) we describe a number of specific constructions, showing how one recovers some existing constructions

as a particular case of our framework. In [Section 5](#) we provide an analysis of the asymptotic truncation error. In [Section 6](#) we show how we can use the series representation and iid approximation to derive a Markov chain Monte Carlo samplers for posterior inference in normalised CRM models, and provide some numerical illustration in [Section 7](#). Related approaches are discussed in [Section 8](#). Some proofs, additional background material and derivations are provided in the Supplementary Material.

1.0.0.1. Notations. For a measure ν on S and a positive measurable function h on S , write $\nu(h) = \int_S h(x) \nu(dx)$. Let (ξ_1, ξ_2, \dots) be the ordered points of a unit-rate Poisson point process on $(0, \infty)$, that is $\xi_1, \xi_2 - \xi_1, \xi_3 - \xi_2, \dots$ are iid unit-rate exponential random variables. For a random variable X and a probability distribution F , $X \sim F$ means X has distribution F . For two strictly positive functions $f(t)$ and $g(t)$, $f(t) \stackrel{t \rightarrow \infty}{\sim} g(t)$ means $\lim_{t \rightarrow \infty} f(t)/g(t) = 1$. Similarly, for sequences of strictly positive random variables $(R_n)_{n \geq 1}$ and $(S_n)_{n \geq 1}$, $R_n \stackrel{n \rightarrow \infty}{\sim} S_n$ a.s. means $\lim_{n \rightarrow \infty} S_n/R_n = 1$ almost surely. With a slight abuse of notation, we use the same notation for the distribution of a random variable and its pdf. For instance, the probability density function (pdf) of a gamma random variable $X \sim \text{Gamma}(a, b)$, evaluated at x , is written as $\text{Gamma}(x; a, b)$. $\|\cdot\|_1$ denotes the L_1 norm.

2. Background

2.1. Completely random measures

Let (S, \mathcal{S}) be a complete and separable metric space endowed with its Borel σ -field, where $S = (0, \infty) \times \Theta$. For any point $X = (W, \theta) \in S$, we refer to $W > 0$ as the **size** of X . Let N be a Poisson random measure on S with mean measure $\nu(dw, d\theta) = \rho(dw) \mu_w(d\theta)$ where ρ is a Borel measure on $(0, \infty)$, called size measure, satisfying

$$\int_0^\infty (1 - e^{-w}) \rho(dw) < \infty \text{ and } \int_0^\infty \rho(dw) = \infty \quad (2.1)$$

and $\mu_w(\cdot)$ is a Markov probability kernel from $(0, \infty)$ to Θ . The linear functional

$$G(d\theta) = \int_0^\infty w N(dw, d\theta)$$

is an infinite-activity completely random measure [\[41\]](#) on Θ with random weights and random atoms. We write $G \sim \text{CRM}(\rho, \mu_w)$. The conditions [\(2.1\)](#) imply that the atomic random measures N and W have an infinite number of atoms, and $W(\Theta)$ is almost surely finite. If $\mu_w = H$ does not depend on w , the CRM is said to be homogeneous. We assume in the rest of this article that one can easily simulate from μ_w (or H) and/or it admits a tractable density with respect to some reference measure (e.g. Lebesgue).

Two popular examples of CRMs are the generalised gamma process (GGP) [\[32, 11\]](#), also known as (exponentially) tilted stable process, with size measure

$$\rho(dw) = \frac{\alpha}{\Gamma(1 - \sigma)} w^{-1-\sigma} e^{-\tau w} dw \quad (2.2)$$

where $\alpha > 0$, $\sigma \in (0, 1)$ and $\tau \geq 0$, or $\sigma \leq 0$ and $\tau > 0$, and the stable beta process (SBP) [\[31, 62\]](#) with

$$\rho(dw) = \frac{\alpha}{B(1 - \sigma, c + \sigma)} w^{-\sigma-1} (1 - w)^{c+\sigma-1} \mathbb{1}_{\{0 < w < 1\}} dw, \quad (2.3)$$

where $\sigma \in (-\infty, 1)$, $c > -\sigma$, $\alpha > 0$ and $B(\cdot, \cdot)$ is the beta function. When $\sigma \geq 0$, both random measures are infinite-activity. The gamma and stable processes are obtained as special cases when setting $\sigma = 0$.

Remark 2.1. The constructions described in this paper hold more generally when the first condition in Equation (2.1) is not satisfied, but $\int_x^\infty \rho(dw) < \infty$ for all $x > 0$. Note that in this case $W(\Theta) = \infty$ almost surely. An example of this more general case is given in Section 4.4 where $\rho(dw) = w^{-2}dw$.

2.2. Objective

Our objective is to derive general series representations for the Poisson random measure N , or equivalently the CRM G , of the form

$$G = \sum_{i=1}^{\infty} W_i \delta_{\theta_i} \quad (2.4)$$

where the sizes (W_1, W_2, \dots) , are stochastically ordered. That is, for any $w > 0$, $\Pr(W_{i+1} > w) \leq \Pr(W_i > w)$. We write $W_1 \succeq W_2 \succeq \dots$ and $X_i = (W_i, \theta_i)$. Denote by G_n the measure obtained by truncating the above series after n points

$$G_n = \sum_{i=1}^n W_i \delta_{\theta_i} = \sum_{i=1}^n \bar{W}_{n,i} \delta_{\bar{\theta}_{n,i}} \quad (2.5)$$

where $(\bar{X}_{n,1}, \dots, \bar{X}_{n,n})$ is a finitely exchangeable random sequence defined by $\bar{X}_{n,i} = X_{\pi_n(i)}$ where π_n is a random permutation of the set $\{1, \dots, n\}$. We will refer to the sequence (X_1, \dots, X_n) (or (W_1, \dots, W_n)) as the **sequential truncated representation**, and $(\bar{X}_{n,1}, \dots, \bar{X}_{n,n})$ (or $(\bar{W}_{n,1}, \dots, \bar{W}_{n,n})$) as the **exchangeable truncated representation**. In Section 3 we will show that the exchangeable truncated representation can be approximated by a **finite iid representation**, which will be denoted $(\tilde{X}_{n,1}, \dots, \tilde{X}_{n,n})$.

2.3. Existing representations of CRMs

In this section we present some existing series representations which we will show are special cases of the novel general construction introduced in Section 3.

2.3.1. Inverse-Lévy representation

For any $x > 0$, let

$$\bar{\rho}(x) = \int_x^\infty \rho(dw) \quad (2.6)$$

be the tail intensity of the size measure ρ , and denote by $\bar{\rho}^{-1}(y) = \inf\{x > 0 \mid \bar{\rho}(x) \leq y\}$ its generalised inverse. The inverse Lévy representation [39, 24] is given by

$$W_i = \bar{\rho}^{-1}(\xi_i).$$

In this case, the sizes are ordered $W_1 \geq W_2 \geq \dots$ and it therefore leads to the best possible approximation in terms of the sizes. While this representation has been used in many applications [67, 51, 29, 7, 3, 2] its main limitation is that $\bar{\rho}^{-1}$ is in general non-tractable. Two exceptions are the beta random measure, whose inverse Lévy representation is given by Equation (1.1), and the stable random measure (corresponding to the measure (2.2) with $\sigma \in (0, 1)$ and $\tau = 0$) where the inverse Lévy representation is given by $W_i = (\xi_i \sigma \Gamma(1 - \sigma) / \alpha)^{-1/\sigma}$.

2.3.2. Size-biased representation

The size-biased sequential and exchangeable representations (W_1, \dots, W_n) and $(\bar{W}_{n,1}, \dots, \bar{W}_{n,n})$, introduced by [55, Section 4], are given as follows¹. Let $0 < T_1 \leq T_2 \leq \dots$ be defined as

$$T_i = \Psi^{-1}(\xi_i)$$

where Ψ^{-1} is the generalised inverse of the Laplace exponent $\Psi(t) = \int_0^\infty (1 - e^{-wt})\rho(dw)$ and

$$\mathbb{P}(W_i \in dw \mid T_i = t) = \frac{we^{-wt}\rho(dw)}{\psi(t)}$$

where $\psi(t) = \Psi'(t) = \int_0^\infty we^{-wt}\rho(dw)$. Additionally, given $T_{n+1} = t_{n+1}$, we have $\bar{W}_{n,1}, \dots, \bar{W}_{n,n}$ are iid with distribution

$$\mathbb{P}(\bar{W}_{n,i} \in dw \mid T_{n+1} = t_{n+1}) = \frac{(1 - e^{-wt_{n+1}})\rho(dw)}{\Psi(t_{n+1})}.$$

The term size-biased comes from the fact that the atoms are ordered by successively sampling without replacement according to their size W

$$\Pr(0 < T_1 \leq T_2 \leq T_3 \leq \dots \mid G) = \prod_{k=1}^{\infty} \frac{W_k}{\sum_{j \geq k} W_j}.$$

In the case of the gamma random measure, which corresponds to Equation (2.2) with $\sigma = 0$, [55] show that the series representation corresponds to [10]'s representation and is given by $T_i = \Psi^{-1}(\xi_i) = \tau(e^{\xi_i/\alpha} - 1)$ and $W_i \mid T_i = t_i \sim \text{Gamma}(1, \tau + t_i)$.

3. Series representations and finite approximations of CRMs

3.1. Arrival-time augmentation

Let $\lambda_w(dt)$ be some Markov probability kernel from $(0, \infty)$ to $(0, \infty)$ with cdf $\Lambda_w(t) = \int_0^t \lambda_w(du)$ satisfying, for any $t > 0$

$$\int_0^\infty \Lambda_w(t)\rho(dw) < \infty \quad \text{and} \quad \Lambda_{w_2}(t) \leq \Lambda_{w_1}(t) \text{ for all } 0 < w_2 \leq w_1. \quad (3.1)$$

That is, if $T_1 \sim \lambda_{w_1}$ and $T_2 \sim \lambda_{w_2}$, with $0 < w_2 \leq w_1$, then $T_2 \succeq T_1$.

¹Note that this is different from what [14] call a size-biased representation.

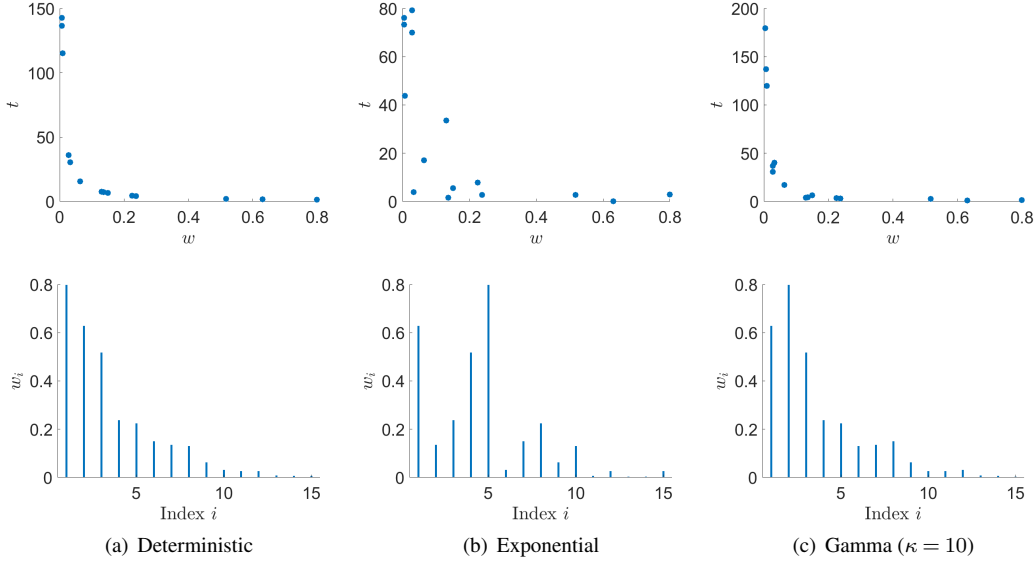


Figure 1. Illustration of the Poisson process construction, for different arrival time cdfs λ_w . The top Figures represent the Poisson point process $(w_i, t_i)_{i \geq 1}$ and the bottom Figures the weights ordered according to their arrival times for (left) deterministic arrival time $\lambda_w(dt) = \delta_{1/w}(dt)$, (middle) Exponential arrival time $\lambda_w(dt) = we^{-wt}dt$ and (right) gamma arrival time $\lambda_w(dt) = \text{Gamma}(t; \kappa, \kappa w)dt$ with $\kappa = 10$.

We consider a Poisson random measure N' on the augmented space $S' = (0, \infty) \times \Theta \times (0, \infty)$ with mean measure $\nu'(dw, d\theta, dt) = \rho(dw)\mu_w(d\theta)\lambda_w(dt)$. For a point $X' = (W, \theta, T) \in S'$, we refer to T as the **arrival time** of the point X' . Indeed, the second condition in Equation (3.1) ensures that points with larger size W are more likely to have a smaller arrival time T . We may therefore consider the following analogy: atoms of the Poisson random measure are enrolled in a race, each atom having a strength W , and stronger atoms are more likely to finish faster and therefore have a smaller T . The first condition in Equation (3.1) ensures that $N'(\mathbb{R}_+, \Theta, (0, t)) < \infty$ for any $t > 0$ hence we can order the arrival times. Let $0 < T_1 \leq T_2 \leq \dots$ denote the sequence of ordered arrival times, and consider the augmented sequential representation

$$N' = \sum_{i=1}^{\infty} \delta_{(W_i, \theta_i, T_i)}$$

where $X_i = (W_i, \theta_i)$, $i \geq 1$ are the associated sizes and locations. By the restriction theorem [40], $N = \sum_{i=1}^{\infty} \delta_{(W_i, \theta_i)}$ is a Poisson random measure with mean $\rho(dw)\mu_w(d\theta)$ and

$$G(d\theta) = \int_0^\infty \int_0^\infty w N'(dw, d\theta, dt) = \int_0^\infty w N(dw, d\theta) \sim \text{CRM}(\rho, \mu_w).$$

An illustration of the augmented Poisson point process construction is given in Figure 1.

We now give the general definitions of the sequential, exchangeable and iid representations of the CRM associated to the arrival time kernel λ_w . For simplicity of presentation, we assume that for any w , λ_w is absolutely continuous with respect to the Lebesgue measure, but one can also consider

discontinuous cdfs Λ_w , see [Section 4.1](#) for an example. We use the same notation $\lambda_w(dt)$ and $\lambda_w(t)$ for the distribution and its density.

3.2. Series and truncated exchangeable constructions

Theorem 3.1. Let λ_w be a parametric distribution on $(0, \infty)$ with parameter $w > 0$ and Λ_w be the associated parametric cumulative distribution function (cdf) satisfying condition (3.1). Consider the conditional distributions

$$\phi_t(dw) = \frac{\lambda_w(t)\rho(dw)}{\psi(t)} \quad \text{and} \quad \varphi_t(dw) = \frac{\Lambda_w(t)\rho(dw)}{\Psi(t)} \quad (3.2)$$

where

$$\psi(t) = \int_0^\infty \lambda_w(t)\rho(dw) \quad \text{and} \quad \Psi(t) = \int_0^t \psi(u)du = \int_0^\infty \Lambda_w(t)\rho(dw). \quad (3.3)$$

The **sequential construction** $G = \sum_{i=1}^\infty W_i \delta_{\theta_i}$ is obtained as follows, for $i \geq 1$

$$T_i = \Psi^{-1}(\xi_i) \quad (3.4a)$$

$$W_i \mid T_i = t_i \sim \phi_{t_i} \quad (3.4b)$$

$$\theta_i \mid W_i = w_i \sim \mu_{w_i}. \quad (3.4c)$$

The **truncated exchangeable construction** $G_n = \sum_{i=1}^n \bar{W}_{n,i} \delta_{\bar{\theta}_{n,i}}$ is obtained, for $i = 1, \dots, n$ by

$$T_{n+1} = \Psi^{-1}(\xi_{n+1}) \quad (3.5a)$$

$$\bar{W}_{n,i} \mid T_{n+1} = t_{n+1} \stackrel{\text{i.i.d.}}{\sim} \varphi_{t_{n+1}} \quad (3.5b)$$

$$\bar{\theta}_{n,i} \mid \bar{W}_{n,i} = \bar{w}_{n,i} \sim \mu_{\bar{w}_{n,i}}. \quad (3.5c)$$

Proof. The proof is an adaptation of the proof for the size-biased construction in [55, Section 4]. The mean measure ν' of the Poisson random measure N' can be expressed as

$$\begin{aligned} \nu'(dw, d\theta, dt) &= \rho(dw) \mu_w(d\theta) \lambda_w(t) dt \\ &= \psi(t) dt \times \frac{\lambda_w(t)\rho(dw)}{\psi(t)} \times \mu_w(d\theta) \\ &= \psi(t) dt \times \phi_t(dw) \times \mu_w(d\theta). \end{aligned}$$

This is the mean measure of a marked Poisson point process, where (T_1, T_2, \dots) are the points of an inhomogeneous Poisson point process with intensity $\psi(t)$, hence admit the representation of Equation (3.4a), and the marks (W_i, θ_i) have conditional distribution $\phi_t(dw) \mu_w(d\theta)$ as shown in Equations (3.4b) and (3.4c). Let $(\bar{X}'_{n,1}, \dots, \bar{X}'_{n,n}) = (X'_{\pi_1}, \dots, X'_{\pi_n})$, where (π_1, \dots, π_n) is a random permutation of $\{1, \dots, n\}$, and $\bar{X}'_{n,i} = (\bar{T}_{n,i}, \bar{W}_{n,i}, \bar{\theta}_{n,i})$. By properties of the Poisson process on the real line, the random variables $\bar{T}_{n,1}, \dots, \bar{T}_{n,n}$ are iid given $T_{n+1} = t_{n+1}$, with pdf

$$\frac{\psi(t) \mathbb{1}_{\{0 < t < t_{n+1}\}}}{\Psi(t_{n+1})}.$$

Hence, given $T_{n+1} = t_{n+1}$, the marks $\bar{W}_{n,i}$ and $\bar{\theta}_{n,i}$ are also iid, with conditional distribution $\varphi_{t_{n+1}}(d\bar{w}_{n,i})\mu_{\bar{w}_{n,i}}(d\bar{\theta}_{n,i})$ where

$$\begin{aligned}\varphi_{t_{n+1}}(d\bar{w}_{n,i}) &= \mathbb{E} \left[\phi_{\bar{T}_{n,i}}(d\bar{w}_{n,i}) \mid T_{n+1} = t_{n+1} \right] \\ &= \int_0^{t_{n+1}} \frac{\lambda_w(t)\rho(d\bar{w}_{n,i})}{\psi(t)} \frac{\psi(t)}{\Psi(t_{n+1})} dt \\ &= \frac{\Lambda_w(t_{n+1})\rho(d\bar{w}_{n,i})}{\Psi(t_{n+1})}.\end{aligned}$$

□

The sequential representation (3.4) requires the evaluation of the inverse Ψ^{-1} of the function Ψ defined in Equation (3.3) for each of the atoms. For some specific choices of Λ_w and ρ , this inverse can be obtained in closed form; see some examples in Section 4. However, in general, there is no closed-form expression, limiting the applicability of this representation.

The truncated exchangeable representation (3.5) also requires the evaluation of the inverse Ψ^{-1} , but only once, irrespective of the truncation level n . Assuming that there is a closed-form expression for the real-valued function Ψ , as it is the case for all the examples covered in Section 4, $\Psi^{-1}(\xi_{n+1})$ can be computed numerically for a negligible computational cost. The exchangeable representation also requires to be able to simulate from the distribution φ_{t_n} defined in Equation (3.2). We provide in Section 4 examples where this distribution has a tractable pdf and/or one can easily simulate from this distribution.

As we will see in Section 4, the sequential and exchangeable representations can be seen as generalisations of both the inverse Lévy and size-biased representations.

3.3. Finite iid construction

Note that ξ_{n+1}/n tends to 1 almost surely as n tends to infinity. This therefore suggests the following **finite iid construction**, as an approximation to the truncated measure G_n

$$\tilde{G}_n = \sum_{i=1}^n \tilde{W}_{n,i} \delta_{\tilde{\theta}_{n,i}} \quad (3.6a)$$

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

$$\tilde{W}_{n,i} \stackrel{\text{i.i.d.}}{\sim} \tilde{\varphi}_n \quad (3.6b)$$

$$\tilde{\theta}_{n,i} \mid \tilde{W}_{n,i} = \tilde{w}_{n,i} \sim \mu_{\tilde{w}_{n,i}} \quad (3.6c)$$

with

$$\tilde{\varphi}_n(dw) = \varphi_{\Psi^{-1}(n)}(dw) = \frac{\Lambda_w(\Psi^{-1}(n))\rho(dw)}{n}. \quad (3.7)$$

Proposition 3.2. Let \tilde{G}_n be the finite iid approximation defined by Equation (3.6). Then \tilde{G}_n converges weakly in distribution to $G \sim \text{CRM}(\rho, \mu_w)$ as $n \rightarrow \infty$.

Proof. The proof is similar to that of [43, Section 3.1]. Let $\mathcal{M}_+(\Theta)$ be the set of non-negative, totally finite measures on Θ , equipped with the topology of weak convergence. Note that, almost surely, $\tilde{G}_n, G \in \mathcal{M}_+(\Theta)$. It suffices to show that $\mathbb{E} \left[e^{-\tilde{G}_n(f)} \right] \rightarrow \mathbb{E} \left[e^{-G(f)} \right]$ for all $f \in C_b^+(\Theta)$, the set of measurable, bounded and continuous functions $f : \Theta \rightarrow [0, \infty)$. For any $f \in C_b^+(\Theta)$, we have

$$\begin{aligned} \mathbb{E} \left[e^{-\tilde{G}_n(f)} \right] &= \mathbb{E} \left[e^{-\sum_{i=1}^n \tilde{W}_{n,i} f(\tilde{\theta}_{n,i})} \right] \\ &= \mathbb{E} \left[e^{-\tilde{W}_{n,1} f(\tilde{\theta}_{n,1})} \right]^n \\ &= \left(\int_S e^{-wf(\theta)} \tilde{\varphi}_n(dw) \mu_w(d\theta) \right)^n \\ &= \left(1 - \int_S (1 - e^{-wf(\theta)}) \tilde{\varphi}_n(dw) \mu_w(d\theta) \right)^n \\ &= \left(1 - \frac{1}{n} \int_S (1 - e^{-wf(\theta)}) \Lambda_w(\Psi^{-1}(n)) \rho(dw) \mu_w(d\theta) \right)^n. \end{aligned}$$

First assume $\int_S (1 - e^{-wf(\theta)}) \rho(dw) \mu_w(d\theta) < \infty$ (the case of an infinite integral is considered below). Note that $\Lambda_w(\Psi^{-1}(n)) \leq 1, \forall n$ and $\Lambda_w(\Psi^{-1}(n)) \rightarrow 1$ as n tends to infinity. By the dominated convergence theorem, we therefore have

$$\int_S (1 - e^{-wf(\theta)}) \Lambda_w(\Psi^{-1}(n)) \rho(dw) \mu_w(d\theta) \rightarrow \int_S (1 - e^{-wf(\theta)}) \rho(dw) \mu_w(d\theta)$$

as $n \rightarrow \infty$. Additionally, for any real sequence $(a_n)_{n \geq 1}$ converging to a we have $(1 - a_n/n)^n \rightarrow e^{-a}$ as $n \rightarrow \infty$. We therefore obtain

$$\mathbb{E} \left[e^{-\tilde{G}_n(f)} \right] \rightarrow \exp \left(- \int_S (1 - e^{-wf(\theta)}) \rho(dw) \mu_w(d\theta) \right),$$

where the right-handside is equal to the Laplace functional $\mathbb{E}[e^{-G(f)}]$ of the CRM $G \sim \text{CRM}(\rho, \mu_w)$ by Campbell's theorem [40].

Consider now that $\int_S (1 - e^{-wf(\theta)}) \rho(dw) \mu_w(d\theta) = \infty$. Note that $\Lambda_w(\Psi^{-1}(m)) < \Lambda_w(\Psi^{-1}(n))$ for any $m < n$, hence

$$\mathbb{E} \left[e^{-\tilde{G}_n(f)} \right] \leq \left(1 - \frac{1}{n} \int_S (1 - e^{-wf(\theta)}) \Lambda_w(\Psi^{-1}(m)) \rho(dw) \mu_w(d\theta) \right)^n.$$

Taking first the limit as $n \rightarrow \infty$, then $m \rightarrow \infty$ gives $\mathbb{E} \left[e^{-\tilde{G}_n(f)} \right] \rightarrow 0$ by sandwiching. \square

For the iid construction, one needs to evaluate $\Psi^{-1}(n)$ only once, and this can be done numerically if there is an analytic form for Ψ . Instead of the distribution $\tilde{\varphi}_n = \varphi_{\Psi^{-1}(n)}$, we can alternatively use more general distributions $\tilde{\varphi}_n = \varphi_{f(n)}$ where f behaves similarly to Ψ^{-1} at infinity. This means that f is an increasing function such that $\Psi(f(n)) \stackrel{n \rightarrow \infty}{\sim} n$. Proposition 3.2 still holds as the proof can be straightforwardly adapted to this case. Note that if $\Psi(t) \stackrel{t \rightarrow \infty}{\sim} ct^\sigma$ for some constant c and $\sigma > 0$, then we can take $f(n) = (n/c)^{1/\sigma}$. Lemma B.1 in the Supplementary Material can be used to

find admissible functions f under generic assumptions on ρ and Λ_w . The exchangeable representation requires to be able to simulate from the distribution $\tilde{\varphi}_n$. We provide in [Section 4](#) examples where this distribution has a tractable pdf and/or one can easily simulate from this distribution.

4. Examples

We first show how the inverse Lévy and size-biased constructions described in [Section 2.3](#) can be recovered as special cases of the general construction introduced in [Section 3](#). We then derive novel constructions within this framework.

4.1. Deterministic arrival times (inverse-Lévy construction)

Assume that the arrival times are deterministic given the size $W = w$, and inversely proportional to it, that is

$$\lambda_w(dt) = \delta_{1/w}(dt). \quad (4.1)$$

The distribution does not admit a density with respect to the Lebesgue measure, but one can still obtain expressions for the different quantities of interest. We obtain

$$\begin{aligned} \Lambda_w(t) &= \mathbb{1}_{\{t \geq 1/w\}}, & \Psi(t) &= \bar{\rho}(1/t), & \Psi^{-1}(\xi) &= 1/\bar{\rho}^{-1}(\xi), \\ \phi_t(dw) &= \delta_{1/t}(dw), & \varphi_t(dw) &= \frac{\rho(dw)\mathbb{1}_{\{w > 1/t\}}}{\bar{\rho}(1/t)}, & \tilde{\varphi}_n(dw) &= \frac{\rho(dw)\mathbb{1}_{\{w > \bar{\rho}^{-1}(n)\}}}{n}. \end{aligned}$$

The sequential construction corresponds to the inverse-Lévy construction described in [Section 2.3](#). The exchangeable representation is similar to the ϵ -truncation of normalised CRMs used in [\[3, 4\]](#), except that the truncation threshold $\epsilon = 1/T_{n+1}$ is treated as a random variable here.

4.2. Exponential arrival times (size-biased construction)

Consider an exponential arrival time distribution for t with parameter w ; $\lambda_w(dt) = we^{-wt}dt$ and $\Lambda_w(t) = 1 - e^{-wt}$. This leads to [\[55\]](#)'s size-biased sequential and exchangeable representations described in [Section 2.3](#). While this construction is not novel, it appears that it provides a novel series representation for the generalised gamma random measure. We also show that the iid representation associated to this arrival time distribution corresponds to the finite approximation proposed by [\[43\]](#).

Generalised gamma process. In the case of the size measure [\(2.2\)](#) with $\alpha > 0$, $\sigma \in (0, 1)$ and $\tau \geq 0$, we obtain the following novel sequential construction for the GGP

$$W_i \mid \xi_i \sim \text{Gamma}\left(1 - \sigma, (\sigma\xi_i/\alpha + \tau^\sigma)^{\frac{1}{\sigma}}\right). \quad (4.2)$$

Details are given in [Section E.1](#) in the Supplementary Material, as well as a comparison to Rosinski's series representation for the GGP [\[59, 60\]](#). The conditional distribution for the exchangeable and iid constructions is given by

$$\varphi_t(dw) = \frac{\sigma w^{-1-\sigma} e^{-\tau w} (1 - e^{-tw}) dw}{\Gamma(1 - \sigma) ((t + \tau)^\sigma - \tau^\sigma)}. \quad (4.3)$$

This is the distribution of an exponentially-tilted BFRY distribution [8, 19, 43], written as $\text{etBFRY}(\sigma, t, \tau)$. One can easily simulate from this distribution via inversion, see Section D.2 in the Supplementary Material. Note that

$$\Psi(t) \stackrel{t \rightarrow \infty}{\sim} \alpha t^\sigma / \sigma,$$

hence we can consider the iid distribution

$$\tilde{\varphi}_n(dw) = \varphi_{(n\sigma/\alpha)^{1/\sigma}}(dw).$$

This corresponds precisely to the finite-dimensional approximation introduced by [43] for the GGP, which can therefore be seen as a particular case of our approach. See Section E.1 in the Supplementary Material for more details.

4.3. Gamma arrival times

As a generalisation of the exponential arrival times, consider now a gamma arrival distribution

$$\lambda_w(dt) = \frac{t^{\kappa-1} e^{-\kappa w t} (\kappa w)^\kappa}{\Gamma(\kappa)}, \quad \Lambda_w(t) = \frac{\gamma(\kappa, \kappa w t)}{\Gamma(\kappa)}, \quad (4.4)$$

where $\kappa \geq 1$ is a tuning parameter and $\gamma(\kappa, t) = \int_0^t x^{\kappa-1} e^{-x} dx$ is the lower incomplete gamma function. Since $\mathbb{E}[T|w] \rightarrow 1/w$ and $\text{Var}(T|w) \rightarrow 0$ as $\kappa \rightarrow \infty$, T converges in probability hence in distribution to $1/w$, and therefore $\Lambda_w(t) \rightarrow \mathbb{1}_{\{t \geq 1/w\}}$ as κ tends to infinity, which corresponds to the arrival time cdf of the inverse Lévy representation. Hence, the construction based on the gamma arrival times bridges between the size-biased ($\kappa = 1$) and inverse-Lévy ($\kappa \rightarrow \infty$) constructions.

Generalised gamma process. Consider the generalised gamma process with size measure (2.2) and parameters $\alpha > 0$, $\sigma \in (0, 1)$ and $\tau \geq 0$. We have

$$\psi(t) = \eta \frac{t^{\kappa-1}}{(t + \tau/\kappa)^{\kappa-\sigma}} \quad \text{and} \quad \Psi(t) = \begin{cases} \eta \left(\frac{t}{\kappa}\right)^\sigma B_{\frac{\kappa t}{\kappa t + \tau}}(\kappa, -\sigma) & \text{if } \tau > 0 \\ \frac{\eta}{\sigma} t^\sigma & \text{if } \tau = 0 \end{cases}$$

where $B_x(\cdot, \cdot)$ is the incomplete beta function and $\eta = \frac{\alpha \kappa^\sigma \Gamma(\kappa - \sigma)}{\Gamma(\kappa) \Gamma(1 - \sigma)}$. For the sequential and exchangeable constructions, we get

$$\phi_t(dw) = \text{Gamma}(w; \kappa - \sigma, \kappa t + \tau) dw \quad \text{and} \quad \varphi_t(dw) \propto w^{-\sigma-1} e^{-\tau w} \gamma(\kappa, \kappa t w) dw.$$

Details of the derivations are given in Section E.2 in the Supplementary Material. For the iid construction, we can use Equation (3.7) and estimate $\Psi^{-1}(n)$ numerically or, using Lemma B.1 and Table 1, we can alternatively use $\tilde{\varphi}_n(dw) = \varphi_{(\sigma n/\eta)^{1/\sigma}}(dw)$. The normalising constant of φ_t (and therefore $\tilde{\varphi}_n$) has an analytic expression via standard functions. We call the random variable having distribution φ_t an *exponentially-tilted generalised BFRY* random variable, due to the form of the pdf obtained by exponentially tilting the pdf of generalised BFRY. This distribution has a number of remarkable properties, summarised in Section D.4 of the Supplementary Material. In Section 7, we provide a simulation study for posterior inference in normalised generalised gamma process models using this iid approximation.

4.4. Inverse gamma arrival times

Consider now an inverse gamma arrival distribution

$$\lambda_w(dt) = \text{iGamma}(t; \kappa, \kappa/w) dt$$

where $\kappa \geq 1$ is a tuning parameter and

$$\text{iGamma}(t; a, b) = \frac{b^a}{\Gamma(a)} t^{-a-1} e^{-b/t}$$

is the pdf of an inverse gamma random variable with shape parameter $a > 0$ and scale parameter $b > 0$. By a similar argument as for the gamma arrival times, we have $\Lambda_w(t) = \frac{\Gamma(\kappa, \kappa/(wt))}{\Gamma(\kappa)} \rightarrow \mathbb{1}_{\{t \geq 1/w\}}$ as $\kappa \rightarrow \infty$ hence it also admits the inverse Lévy construction as a limiting case. The case $\kappa = 1$ is of particular interest, as it leads to a tractable novel representation for the GGP (see Section E.3 in the Supplementary Material), and provide a novel way of interpreting the classical iid approximation of the beta process.

Beta process. Consider the one-parameter beta process with size measure (2.3) with $\sigma = 0$ and $c = 1$. The bijective transformation $u = -(\alpha \log(w))^{-1}$ gives the measure $\rho(du) = u^{-2} du$ on $(0, \infty)$. Note that $\rho(du)$ is not a Lévy measure, but we can nonetheless use our construction as the tail Lévy intensity is finite. Using the inverse gamma kernel with $\kappa = 1$, we obtain $\Psi(t) = t$ and the iid distribution $\tilde{\varphi}_n(du) = \text{iGamma}(u; 1, 1/n) du$. Applying the inverse transformation $\tilde{W}_i = e^{-1/(\alpha \tilde{U}_i)}$, we obtain

$$\tilde{W}_i \sim \text{Beta}(\alpha/n, 1),$$

which corresponds to the classical iid approximation for the beta process, described in Equation (1.2). The iid construction for the beta process can alternatively be recovered using the arrival time cdf $\Lambda_w(t) = w^{\frac{\alpha}{t}}$ directly with the measure (2.3), without change of variable, as shown in Section E.3 in the Supplementary Material.

4.5. Generalised Pareto arrival time

Consider the arrival time distribution $\lambda_w(dt) = \frac{cw}{(tw+1)^{c+1}} dt$ where $c > 0$.

Stable beta process. Consider the stable beta process with measure (2.3) with $\sigma > 0$. We have $\Psi(t) = \frac{\alpha c}{\sigma}((t+1)^\sigma - 1)$ and, as shown in Section E.4 in the Supplementary Material,

$$\begin{aligned} \phi_t(dw) &= \frac{w^{-\sigma}(1-w)^{c+\sigma-1}(tw+1)^{-c-1} \mathbb{1}_{\{0 < w < 1\}} dw}{B(1-\sigma, c+\sigma)(t+1)^{\sigma-1}} \\ \varphi_t(dw) &= \frac{\sigma w^{-1-\sigma}(1-w)^{c+\sigma-1}}{cB(1-\sigma, c+\sigma)((t+1)^\sigma - 1)} \left(1 - \frac{1}{(tw+1)^c}\right) \mathbb{1}_{\{0 < w < 1\}} dw. \end{aligned}$$

These distributions admit the same conjugacy properties as the beta distribution, and one can sample exactly from these distributions as detailed in Section E.4 in the Supplementary Material.

5. Truncation error analysis

In this section, we discuss the asymptotic error when using the truncated series representation. We consider both the case of a normalised and unnormalised CRM. We present in [Section 5.1](#) results on the asymptotic error between the normalised/unnormalised CRM and its truncation. In [Section 5.2](#), we consider the approximation error induced when using truncated priors in hierarchical Bayesian models; as has been done in earlier work [\[33, 14\]](#), we present asymptotic bounds on the error on the marginal likelihood under the truncated prior. Although we focus on general asymptotic bounds here, it is also possible to obtain non-asymptotic bounds for some specific CRM and arrival time distributions. To illustrate this we provide some simple non-asymptotic bounds for the generalised gamma process with exponential arrival time in [Section I](#) in the Supplementary material.

5.1. Asymptotic error of the truncated unnormalised and normalised measures

Denote

$$R_n = \|G - G_n\|_1 = \sum_{i>n} W_i \quad (5.1)$$

the L_1 error between the CRM and its finite approximation. Similarly, denote

$$\bar{R}_n = \left\| \frac{G}{G(\Theta)} - \frac{G_n}{G_n(\Theta)} \right\|_1 = \sum_{i \leq n} W_i \left(\frac{1}{G_n(\Theta)} - \frac{1}{G(\Theta)} \right) + \sum_{i>n} \frac{W_i}{G(\Theta)} \quad (5.2)$$

$$= 2 \frac{R_n}{G(\Theta)}. \quad (5.3)$$

the L_1 error between the normalised CRM and its finite approximation; such metric is routinely used to compare normalised/unnormalised measures [\[33, 14\]](#).

Proposition 5.1. For $\xi \sim \text{Gamma}(n+1, 1)$, R_n has the following moment generating function

$$\mathbb{E}[e^{-\lambda R_n}] = \mathbb{E}_\xi \left[\exp \left(- \int_S (1 - e^{-\lambda w}) (1 - \Lambda_w(\Psi^{-1}(\xi))) \rho(dw) \mu_w(d\theta) \right) \right]. \quad (5.4)$$

Proof. By the marking theorem for Poisson point processes [\[40, Chapter 5\]](#), given $T_{n+1} = t_{n+1}$, the random measure $\sum_{i|T_i \geq t_{n+1}} \delta_{X_i}$ is a Poisson random measure with mean measure $(1 - \Lambda_w(t_{n+1})) \rho(dw) \mu_w(d\theta)$. The result follows from Campbell's theorem and the fact that $T_{n+1} = \Psi^{-1}(\xi_{n+1})$. \square

The truncation error analysis relies on the following two assumptions on the Lévy measure ρ , and the arrival time distribution Λ_w .

Assumption 1. Assume that the mean measure $\rho(dw)$ is absolutely continuous with respect to the Lebesgue measure with density function $\varrho(w)$ such that

$$\varrho(w) \stackrel{w \rightarrow 0}{\sim} \zeta_0 w^{-1-\sigma} \quad (5.5)$$

where $\sigma \in (0, 1)$ and $\zeta_0 > 0$.

Assumption 2. Assume additionally that

$$\Lambda_w(t) = 1 - k(wt)$$

where k is a positive function on $(0, \infty)$ such that its Mellin transform

$$\check{k}(z) = \int_0^\infty t^{-z-1} k(t) dt$$

is defined at $\sigma - 1 + \epsilon$ for some $\epsilon > 0$. Assume additionally that either (i) k is differentiable with derivative k' and that the Mellin transform \check{k}' of k' is defined in some open interval containing $\sigma - 1$, or (ii) that $k(x) = \mathbb{1}_{\{x \leq 1\}}$.

[Assumption 1](#) states that the Lévy measure ρ has a regularly varying density function with index $\sigma \in (0, 1)$. This is satisfied for the generalised gamma and stable beta measures with $\sigma \in (0, 1)$. The deterministic, exponential, gamma, inverse-gamma (for $\kappa > 2 - \sigma$) and generalised Pareto (for $c > 2 - \sigma$) arrival time distributions discussed in [Section 4](#) all satisfy [Assumption 2](#) for any $\sigma < 1$. Background on regular variation and Mellin transforms is given in [Section A](#) in the Supplementary Material.

The next proposition provides an asymptotic expression for the error term, giving insights on how the error relates to the choice of the arrival time distribution $\lambda_w(t)$. Its proof is given in [Section B.2](#) in the Supplementary Material.

Proposition 5.2. Assume that [Assumptions 1](#) and [2](#) hold for some kernel k , $\sigma \in (0, 1)$ and $\zeta_0 > 0$. Then we have

$$R_n \stackrel{n \rightarrow \infty}{\sim} C_1(\sigma) \zeta_0^{1/\sigma} \sigma^{1-1/\sigma} n^{1-1/\sigma} \quad \text{almost surely} \quad (5.6)$$

where the constant $C_1(\sigma)$ is given by $C_1(\sigma) = (1 - \sigma)^{-1}$ if $k(x) = \mathbb{1}_{\{x \leq 1\}}$ and $C_1(\sigma) = \frac{\check{k}(\sigma-1)}{(-\check{k}'(\sigma-1))^{1-1/\sigma}}$ if k is differentiable, and only depends on the arrival time distribution Λ_w and σ . Additionally, if $\int_0^\infty w \rho(dw) < \infty$, we also have

$$\mathbb{E}[R_n] \stackrel{n \rightarrow \infty}{\sim} C_1(\sigma) \zeta_0^{1/\sigma} \sigma^{1-1/\sigma} n^{1-1/\sigma}. \quad (5.7)$$

In the Proposition above, \check{k}' denotes the Mellin transform of the derivative of k . The associated kernels, Mellin transforms and constants $C_1(\sigma)$ are given in [Table 1](#).

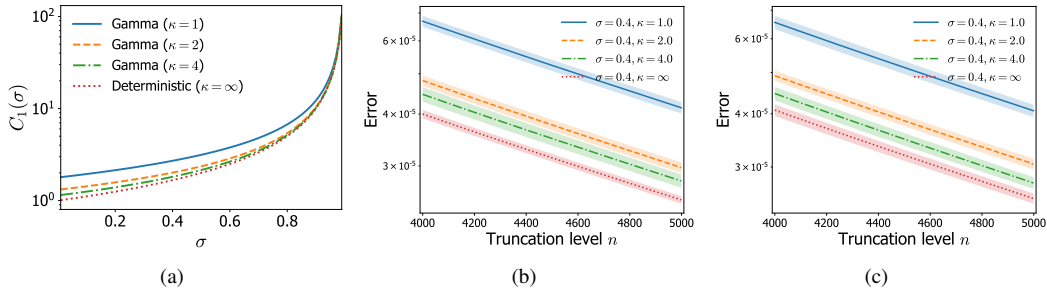
As is apparent from [Proposition 5.2](#), under [Assumptions 1](#) and [2](#), the rate of the truncation error only depends on the parameter σ , and is of order $1/n^{1/\sigma-1}$. The smaller σ , the faster the decrease of the truncation error as n increases. The arrival time distribution tunes the constant $C_1(\sigma)$. The smallest value is obtained for the deterministic arrival time: in this case, the weights are almost surely ordered.

[Figure 2\(a\)](#) shows the value of the constant $C_1(\sigma)$ for the deterministic and gamma arrival times, with different values of the parameter κ . As indicated in [Section 4.3](#), the approximation gets closer to the deterministic/inverse Lévy construction as κ increases. Both the GGP and the SBP with $\sigma > 0$ satisfy [Equation \(5.5\)](#), with $\zeta_0 = \frac{\alpha}{\Gamma(1-\sigma)}$ for the GGP and $\zeta_0 = \frac{\alpha}{B(1-\sigma, c+\sigma)}$ for the SBP.

We ran a simulation study in order to investigate the finite- n properties of the proposed approximations. We report in [Figure 2\(b-c\)](#) the mean and variance of R_n for gamma arrival times, for the stable process and the GGP with $\sigma = 0.4$. For the stable process, we also compare to the inverse-Lévy approximation, as it has an analytic form. As expected, the approximation gets better as the parameter value κ increases. Additional simulations for other arrival time distributions are given in [Section G](#) in the Supplementary Material.

Table 1. Kernels, Mellin transforms and asymptotic constants for different arrival time distributions

	Determin.	Exponential	Gamma	Inv. gamma	Gen. Pareto
$\Lambda_w(t)$	$\mathbb{1}_{\{t \geq 1/w\}}$	$1 - e^{-wt}$	$1 - \frac{\Gamma(\kappa, \kappa wt)}{\Gamma(\kappa)}$	$\frac{\Gamma(\kappa, \kappa/(wt))}{\Gamma(\kappa)}$	$1 - (1 + wt)^{-c}$
$k(x)$	$\mathbb{1}_{\{x \leq 1\}}$	e^{-x}	$\frac{\Gamma(\kappa, x\kappa)}{\Gamma(\kappa)}$	$\frac{\gamma(\kappa, \kappa/x)}{\Gamma(\kappa)}$	$(1 + x)^{-c}$
$\check{k}(-z)$	z^{-1}	$\Gamma(z)$	$\frac{\Gamma(z + \kappa)}{z\Gamma(\kappa)\kappa^z}$	$\frac{\Gamma(\kappa - z)}{z\Gamma(\kappa)\kappa^{-z}}$	$B(z, c - z)$
$k'(x)$	—	$-e^{-x}$	$\frac{-\kappa^\kappa x^{\kappa-1} e^{-\kappa x}}{\Gamma(\kappa)}$	$\frac{-\kappa^\kappa e^{-\kappa/x}}{x^{\kappa+1}\Gamma(\kappa)}$	$-c(1 + x)^{-c-1}$
$-\check{k}'(-z)$	—	$\Gamma(z)$	$\frac{\kappa^{1-z}\Gamma(\kappa + z - 1)}{\Gamma(\kappa)}$	$\frac{\kappa^{z-1}\Gamma(-z + \kappa + 1)}{\Gamma(\kappa)}$	$cB(z, c + 1 - z)$
$C_1(\sigma)$	$(1 - \sigma)^{-1}$	$\Gamma(1 - \sigma)^{1/\sigma}$	$\frac{(\kappa - \sigma)\Gamma(\kappa - \sigma)^{1/\sigma}}{(1 - \sigma)\Gamma(\kappa)^{1/\sigma}}$	$\frac{(\Gamma(\kappa + \sigma)/\Gamma(\kappa))^{1/\sigma}}{(1 - \sigma)(\kappa + \sigma - 1)}$	$\frac{B(1 - \sigma, c + \sigma - 1)}{(cB(1 - \sigma, c + \sigma))^{1-1/\sigma}}$

**Figure 2.** (a) Constant $C_1(\sigma)$ for deterministic and gamma arrival times; (b-c) Simulated error R_n with gamma arrival times for (b) stable process and (c) GGP.

Combining Equation (5.3) and Proposition 5.2 yields the following expression for the asymptotic error on the normalised CRM.

Corollary 5.3. Assume that Assumptions 1 and 2 hold for some kernel k , $\sigma \in (0, 1)$ and $\zeta_0 > 0$. Then we have

$$\overline{R}_n \stackrel{n \rightarrow \infty}{\sim} \frac{2}{Z} C_1(\sigma) \zeta_0^{1/\sigma} \sigma^{1-1/\sigma} n^{1-1/\sigma} \quad \text{almost surely} \quad (5.8)$$

where $C_1(\sigma)$ is as defined in Proposition 5.2 and Z is a random variable with Laplace transform

$$\mathbb{E}[e^{-tZ}] = e^{-\int_0^\infty (1 - e^{-wt}) \rho(dw) \mu_w(d\theta)}.$$

5.2. Truncation error analysis for posterior inference in hierarchical models

5.2.1. Non-normalised CRM hierarchical model

In this section we discuss the L_1 error on the marginal likelihood when truncated CRMs are used for hierarchical Bayesian models under the framework described in [13]. Let $G \sim \text{CRM}(\rho, \mu_w)$, and G_n be its finite approximation with n atoms. Let $H(\cdot|w)$ be a probability distribution on $\mathbb{N} \cup \{0\}$ for all w ,

and denote $\pi(w) := H(0|w)$. Assume

$$\int_0^\infty (1 - \pi(w))\rho(dw) < \infty. \quad (5.9)$$

Consider a hierarchical Bayesian model for m observations $Y_{1:m} = Y_1, \dots, Y_m$ in some space \mathcal{Y} . For $j = 1, \dots, m$ and $i \geq 1$, let

$$Z_{ji}|W_i = w_i \sim H(\cdot|w_i)$$

and write

$$Z_j = \sum_{i \geq 1} Z_{ji} \delta_{\theta_i}.$$

Under the conditions (5.9) and (2.1), Z_j is, almost surely, a discrete, totally finite integer-valued measure. As in [13], we write simply $Z_j | G \sim \text{LP}(H, G)$ where LP stands for likelihood process. Assume finally that $Y_j|Z_j \sim f(\cdot | Z_j)$ for a conditional density f with respect to some measure on \mathcal{Y} . The full hierarchical model takes the form

$$G \sim \text{CRM}(\rho, \mu_w) \quad (5.10a)$$

for $j = 1, \dots, m$

$$Z_j | G \sim \text{LP}(H, G) \quad (5.10b)$$

$$Y_j | Z_j \sim f(\cdot | Z_j) \quad (5.10c)$$

Such CRM-based models have been widely used in machine learning applications; examples include the linear-Gaussian beta-Bernoulli process [30], the linear-Gaussian gamma-Poisson process [64] and the Poisson beta negative binomial process [69]; see [36, 13] for further discussion and examples of such constructions.

Denote $\mathbf{p}_{m,\infty}(Y_{1:m})$ the marginal likelihood for the hierarchical model defined by Equation (5.10). Similarly, denote by $\mathbf{p}_{m,n}(Y_{1:m})$ the marginal likelihood of the model with the same generative process as in Equation (5.10), but where G is replaced by its finite approximation G_n . Following [33, 34, 14], we analyse the quality of approximation by looking at the L_1 distance between $\mathbf{p}_{m,\infty}$ and $\mathbf{p}_{m,n}$.

Proposition 5.4. We have the following bound on the L_1 distance between the marginal likelihoods

$$0 \leq \frac{1}{2} \|\mathbf{p}_{m,\infty} - \mathbf{p}_{m,n}\|_1 \leq 1 - e^{-B_{m,n}},$$

where

$$B_{m,n} \leq m \int_0^\infty \int_0^\infty (1 - \pi(w))(1 - \Lambda_w(\Psi^{-1}(\xi))\rho(dw) \text{Gamma}(\xi; n+1, 1) d\xi.$$

Assume that Assumptions 1 and 2 hold for some kernel k , $\sigma \in (0, 1)$ and $\zeta_0 > 0$. Additionally, assume that $1 - \pi(w) \stackrel{w \rightarrow 0}{\sim} \phi w$ for some $\phi > 0$. Then

$$\lim_{n \rightarrow \infty} \frac{\|\mathbf{p}_{m,\infty} - \mathbf{p}_{m,n}\|_1}{m \times n^{1-1/\sigma}} \leq 2\phi C_1(\sigma) \zeta_0^{1/\sigma} \sigma^{1-1/\sigma}$$

where $C_1(\sigma)$, which depends on the kernel k , is as defined in Proposition 5.2.

5.2.2. Normalised CRM hierarchical model

Consider a hierarchical normalised CRM model [56, 44, 45]

$$G \sim \text{CRM}(\rho, \mu_w) \quad (5.11a)$$

for $j = 1, \dots, m$,

$$\vartheta_j \mid G \sim \frac{G}{G(\Theta)} \quad (5.11b)$$

$$Y_j \mid \vartheta_j \sim \mathcal{K}(\cdot \mid \vartheta_j) \quad (5.11c)$$

where $\mathcal{K}(\cdot \mid \theta)$ is a continuous probability density function on \mathcal{Y} for any $\theta \in \Theta$ and ρ satisfies the conditions (2.1). As in the previous section, denote $\mathbf{p}_{m,\infty}(Y_{1:m})$ the marginal likelihood for the hierarchical model defined by Equation (5.11). Similarly, denote by $\mathbf{p}_{m,n}(Y_{1:m})$ the marginal likelihood of the model with the same generative process as in Equation (5.11), but where G is replaced by its finite approximation G_n .

Proposition 5.5. We have the following bound on the L_1 distance between the marginal likelihoods

$$0 \leq \frac{1}{2} \|\mathbf{p}_{m,\infty} - \mathbf{p}_{m,n}\|_1 \leq m \mathbb{E} \left[\frac{R_n}{G(\Theta)} \right].$$

Assume that Assumptions 1 and 2 hold for some kernel k , $\sigma \in (0, 1)$ and $\zeta_0 > 0$. Additionally, assume that $\int_0^\infty w^2 \rho(dw) < \infty$. Then

$$\lim_{n \rightarrow \infty} \frac{\|\mathbf{p}_{m,\infty} - \mathbf{p}_{m,n}\|_1}{m \times n^{1-1/\sigma}} \leq 2C_1(\sigma) \zeta_0^{1/\sigma} \sigma^{1-1/\sigma} \mathbb{E}[1/G(\Theta)^2]^{1/2}$$

where $C_1(\sigma)$, which depends on the kernel k , is as defined in Proposition 5.2, and $\mathbb{E}[1/G(\Theta)^2] < \infty$.

6. Posterior inference in normalised completely random measure mixture models

We consider here an application to posterior inference in mixture models [47] where the mixing distribution is a normalised CRM (NCRM) [56]. Such models have been developed by [44, 45] and [37]; see also the overview by [46]. An analysis of the asymptotic error on the marginal likelihood under this model has been given in Section 5.2.2. We describe here how to use the series and iid representations to perform posterior inference with this class of models. In Section 6.1 we describe a slice sampler to obtain samples asymptotically distributed from the true posterior distribution, using the series representation. In Section 6.2, we describe a Gibbs sampler to produce samples asymptotically distributed from the approximated posterior distribution, under the iid approximation.

Assume that the data $Y_1, \dots, Y_m \in \mathcal{Y}$ are drawn from a continuous probability density function $f_G(y)$ admitting the following mixture form

$$f_G(y) = \frac{1}{G(\Theta)} \int_{\Theta} \mathcal{K}(y \mid \theta) dG(d\theta) = \frac{1}{\sum_{i=1}^{\infty} W_i} \sum_{i=1}^{\infty} W_i \mathcal{K}(y \mid \theta_i) \quad (6.1)$$

where $\mathcal{K}(\cdot | \theta)$ is a continuous probability density function on \mathcal{Y} for any $\theta \in \Theta$ and $G \sim \text{CRM}(\rho, H)$ is an infinite-activity CRM, where ρ satisfies conditions (2.1). These conditions ensure that $0 < G(\Theta) < \infty$ almost surely. Introducing latent variables $(\vartheta_1, \dots, \vartheta_m)$, the model admits the hierarchical construction (5.11), with $\mu_w = H$. We also introduce allocation variables $s_j \in \mathbb{N}$ such that $\vartheta_j = \theta_{s_j}$, indicating from which component of the CRM originates the latent variable ϑ_j .

6.1. Slice Sampler

Slice sampling for infinite mixture models has been originally proposed by [65] for the mixing distribution being a Dirichlet Process, and extended by [38] to stick-breaking priors. Slice samplers have also been developed for mixture models with a Poisson-Kingman prior [23, 21].

We consider here an extension of the slice sampling algorithm proposed by [29] for NCRM mixture models (see also [22]). [29] proposed two slice samplers, Slice 1 and Slice 2. We focus here on extending the Slice 1 algorithm; a generalised version of Slice 2 can be developed along the same lines.

Let $N' = \sum_{i=1}^{\infty} \delta_{(W_i, \theta_i, T_i)}$ be a Poisson random measure on $S' = (0, \infty) \times \Theta \times (0, \infty)$ with mean measure $\nu'(dw, d\theta, dt) = \rho(dw)H(d\theta)\lambda_w(dt)$, as introduced in Section 3. Then $G = \int_0^\infty \int_0^\infty w N'(dw, \cdot, dt) \sim \text{CRM}(\rho, H)$ and the mixture model of Equation (6.1) can be equivalently written as

$$f_{N'}(y) = \frac{1}{G(\Theta)} \int_{S'} w \mathcal{K}(y | \theta) dN'(dw, d\theta, dt). \quad (6.2)$$

For each observation Y , we introduce a slice variable U which has joint density with Y given by

$$f_{N'}(u, y) = \frac{1}{G(\Theta)} \int_{S'} \mathbb{1}_{\{u < 1/t\}} w t \mathcal{K}(y | \theta) dN'(dw, d\theta, dt) \quad (6.3)$$

$$= \frac{1}{\sum_{i=1}^{\infty} W_i} \sum_{i=1}^{\infty} \mathbb{1}_{\{u < 1/T_i\}} W_i T_i \mathcal{K}(y | \theta_i). \quad (6.4)$$

This slice data augmentation generalises the approach of [29], which is obtained as a special case. Indeed, under the deterministic arrival time distribution (4.1), we have $T_i = 1/W_i$ and Equation (6.4) reduces to

$$\frac{1}{\sum_{i=1}^{\infty} W_i} \sum_{i=1}^{\infty} \mathbb{1}_{\{u < W_i\}} \mathcal{K}(y | \theta_i)$$

which is the classical slice sampling augmentation for mixture models, used by [29]. The slice sampler of [29] can be rather straightforwardly adapted to the more general slice variables. As the derivations are quite lengthy, and require the introduction of a number of additional latent variables, details are given in Section F of the Supplementary Material. We show there that, by carefully choosing the arrival time distribution, one can obtain a Gibbs sampler with simpler and more tractable updates than the one based on the deterministic arrival times.

We note that, the practicality and usefulness of the slice sampler lies primarily in the regime of small σ values. Indeed, for large values of σ and whatever the choice of the series representation, the slice sampler becomes impractical as the number of latent variables to be sampled at each iteration increases quickly with the number m of observations, as noted by [23] and [15].

6.2. Inference for the approximate iid model

An alternative to performing exact posterior inference for the original model is to perform inference on the finite-dimensional iid approximation. We approximate the infinite dimensional measure G with a finite measure \tilde{G}_n . That is, for posterior inference, we assume that the data $Y_1, \dots, Y_m \in \mathcal{Y}$ are drawn from a continuous probability density function $f_{\tilde{G}_n}(y)$ admitting the following finite mixture form

$$f_{\tilde{G}_n}(y) = \frac{1}{\tilde{G}_n(\Theta)} \int_{\Theta} \mathcal{K}(y | \theta) d\tilde{G}_n(d\theta) = \frac{1}{\sum_{k=1}^n \tilde{W}_{n,k}} \sum_{i=1}^n \tilde{W}_{n,i} \mathcal{K}(y | \tilde{\theta}_{n,i}) \quad (6.5)$$

where the weights $\tilde{W}_{n,i}$ are iid from $\tilde{\varphi}_n$, defined in Equation (3.7), and the locations $\tilde{\theta}_{n,i}$ are iid from H , with density h .

Adding a latent variable $V \sim \text{Gamma}(n, \sum_{i=1}^n \tilde{w}_{n,i})$ and the allocation variable $s_{1:m}$, the joint posterior density of the finite iid mixture model is proportional to

$$\pi(\tilde{w}_{n,1:n}, v, s_{1:m}, \tilde{\theta}_{n,1:n} | y_{1:m}) \propto v^{n-1} e^{-v \sum_i \tilde{w}_{n,i}} \prod_{i=1}^n \tilde{w}_{n,i}^{m_i} \tilde{\varphi}_n(\tilde{w}_{n,i}) \left[\prod_{j:s_j=i} \mathcal{K}(y_j | \tilde{\theta}_{n,i}) \right] h(\tilde{\theta}_{n,i}) \quad (6.6)$$

where $m_i = \sum_{j=1}^m \mathbb{1}_{\{s_j=i\}}$ is the number of observations assigned to cluster i . One can then implement a Gibbs sampler or Metropolis-Hastings within Gibbs sampler to obtain samples asymptotically distributed from the joint posterior distribution. A numerical illustration is given in the next section for the generalised gamma process with gamma arrival times. In this case, one can sample exactly from all conditional distributions of the Gibbs sampler.

7. Simulation study

We provide here a simulation study for approximate posterior inference in normalised CRM mixture models. The CRM we consider is the generalised gamma process (see Equation (2.2)) with parameters (α, σ, τ) . The base measure is

$$H(d\theta) = \mathcal{N}(\mu; \mu_0, (r\Sigma)^{-1}) i\mathcal{W}(\Sigma; \nu, \Sigma_0) d\mu d\Sigma, \quad (7.1)$$

where $\theta = (\mu, \Sigma)$. \mathcal{N} and $i\mathcal{W}$ respectively denote the normal and inverse Wishart pdf. We set $\alpha = 1$, $\tau = 1$, $r = 0.005$, $\nu = 12$. For μ_0 and Σ_0 , we randomly draw $X := (X_1, \dots, X_m) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$ and set (μ_0, Σ_0) as the empirical mean and covariance of X . We consider a Gaussian mixed pdf $\mathcal{K}(y | \theta) = \mathcal{N}(y; \mu, \Sigma)$. Then we generated 90 datasets in total with varying σ values. Specifically, we chose $\sigma \in \{0.1, 0.3, 0.5\}$ and generated 30 datasets per each σ value. The GGP G was simulated with the method described in [58, 60].

We consider a finite iid approximation using gamma arrival times with $\kappa \geq 1$, as described in Section 4.3. This includes the iid representation of [43] as a special case when $\kappa = 1$. We obtain

$$\tilde{\varphi}_n(dw) = \text{etgBFRY} \left(w; \kappa, \sigma, \left(\frac{\sigma \Gamma(\kappa) \Gamma(1 - \sigma) n}{\alpha \Gamma(\kappa - \sigma)} \right)^{\frac{1}{\sigma}}, \tau \right) dw. \quad (7.2)$$

where the pdf etgBFRY is defined in Equation (D.7) in the Supplementary Material. The etgBFRY is a conjugate prior for a gamma likelihood, and one can sample exactly from this distribution, as discussed in Section D.4 in the Supplementary Material. It is therefore straightforward to derive a Gibbs sampler for the finite iid model; details of the Gibbs sampler are given in Section H in the Supplementary Material. Note that the finite Gibbs sampler is an (asymptotically) exact sampler targeting an approximate posterior distribution.

To assess the usefulness of the finite iid approximations, we compare the marginal Gibbs sampler [22], targeting the true posterior distribution of the infinite-dimensional model, to the Gibbs sampler targeting the posterior of the finite iid model. We collected the posterior samples for the number K of non-empty clusters from both samplers and measure the Kolmogorov-Sminorov (KS) statistic between them. The posterior samples obtained from the Gibbs sampler for the infinite-dimensional model are considered to be drawn from the true posterior distribution.

For each dataset, we ran the samplers for the infinite and finite iid models for 30,000 iterations with 15,000 burn-in iterations. We ran three independent chains for each dataset and collected every 10th sample after burn-in to construct empirical distributions of the posteriors. We tested the parameters $\kappa = \{1, 2, 4\}$ for the finite iid model where $\kappa = 1$ corresponds to the exponential arrival time [43]. The average \hat{R} statistics [26] over all datasets is 1.0019 ± 0.0019 for the infinite model and 1.0040 ± 0.0062 for the finite iid models, suggesting that all samplers have converged.

Figure 3 compares the approximate posteriors of the finite iid models with $\kappa \in \{1, 2, 4\}$ in terms of KS statistics. As one can see, the approximate posterior distributions using gamma arrival times ($\kappa \in \{2, 4\}$) are closer to the true posterior when compared to the one with exponential arrival times ($\kappa = 1$). As the truncation level n grows, the approximate posteriors get closer to true posterior (Figure 3, right), and get worse as σ grows. The differences in behaviour between the exponential ($\kappa = 1$) and gamma ($\kappa \in \{2, 4\}$) become more apparent for larger values of σ .

Figure 4 compares the actual posterior distributions on K simulations using $\sigma = 0.5$. As can be seen from the figure, both samplers have converged, but the posterior distributions obtained using $\kappa = 4$ are closer to the one obtained from the marginal Gibbs sampler.

8. Discussion

As mentioned in the introduction, there exists wide interest and large literature on approximations of CRMs regarding series representations. [14] provided a recent survey of existing and novel (nested) series representations together with a truncation analysis. Our series construction can be seen as a special case of Rosinski's shot-noise series representation [58] (which includes as special cases most series constructions, see [58]), using the disintegration

$$\rho(dw) = \int_0^\infty \nu(\xi, dw) d\xi$$

where

$$\nu(\xi, dw) = \lambda_w(\Psi^{-1}(\xi)) \rho(dw) / \psi(\Psi^{-1}(\xi))$$

is a Markov kernel (noting that $\int_0^\infty \lambda_w(\Psi^{-1}(\xi)) / \psi(\Psi^{-1}(\xi)) d\xi = 1$). Nested series representations have also been recently proposed for some specific CRMs [54, 52, 61]. Regarding finite iid approximations, the literature is not as rich. Finite iid representations can be obtained using the infinite divisibility properties of CRMs [42] but as noted by [43], it generally does not lead to tractable representations, except for the gamma process case. Subsequent to our work, other ways of obtaining iid constructions,

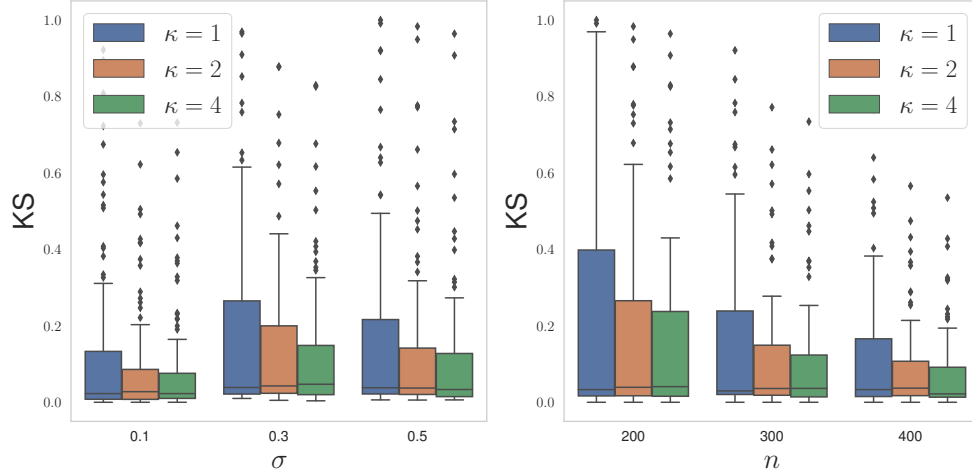


Figure 3. KS statistics between samples from the true posterior and samples from the approximate posteriors of the iid models, for different values of the parameter κ . Lower is better. (Left) KS statistics w.r.t. the change of σ . (Right) KS statistics w.r.t. the change of truncation level n .

which does not rely on a latent Poisson construction are described in [50] for some family of CRMs; see this article for a discussion of the differences with the approach presented here.

All of the above work consider separately series or iid representations, without drawing connections between the two. A key contribution of our work is to consider both series and iid approximation under the same unified framework.

Note that our construction, while very general, and capturing a great range of cases, does not cover all standard iid approximations. For example, the standard iid approximation for the Gamma process $\widetilde{W}_{n,i} \sim \text{Gamma}(\alpha/n, \tau)$ cannot be recovered with our construction. Indeed, from Equation (3.7), we have under our iid construction for the gamma process

$$\tilde{\varphi}_n(dw) = \frac{g_n(w)\alpha w^{-1}e^{-\tau w}}{n}dw$$

where g_n is a bounded function. To obtain a $\text{Gamma}(\alpha/n, \tau)$ distribution would require to set

$$g_n(w) = \frac{nw^{\alpha/n}\tau^{\alpha/n}}{\alpha\Gamma(\alpha/n)}$$

which does not fit into our framework as this function is not bounded.

Beyond giving the iid representations, we also quantify the error and provide in Section 5 generic asymptotic results on the approximation error; these results apply under mild assumptions on the CRM and on the arrival time distribution. Finally, it is of interest and can be seen as a further extensions to this work, to obtain non asymptotic error bounds for some specific CRMs and specific arrival time distributions.

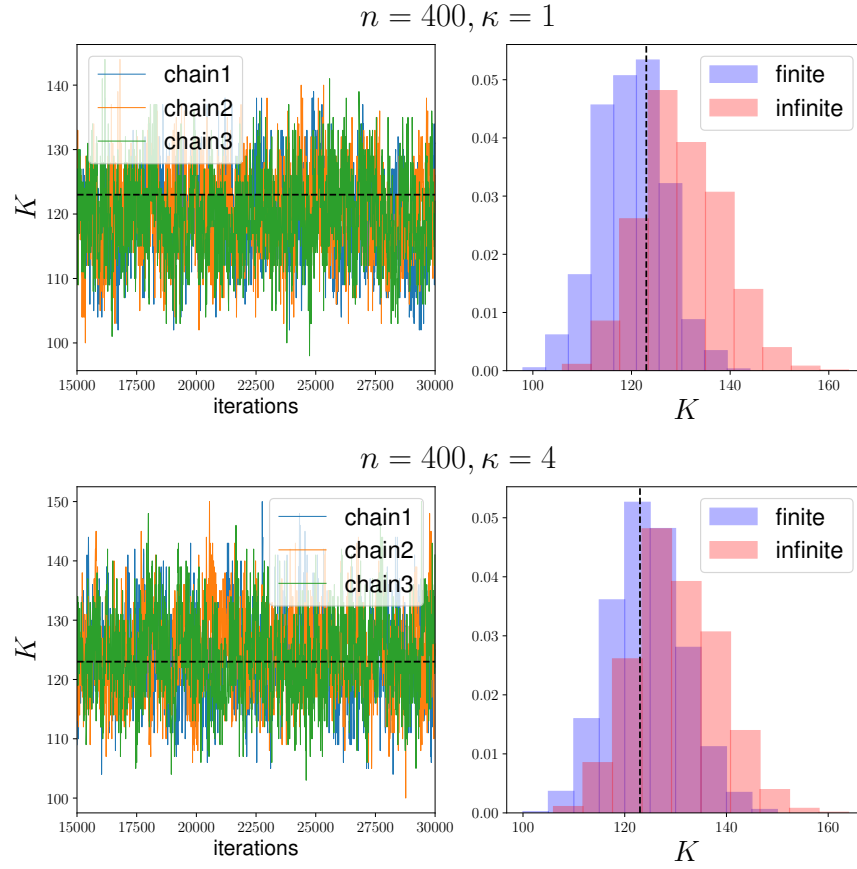


Figure 4. Posterior samples of K from the finite iid model with $n = 400$, $\kappa = 1$ (top) and $\kappa = 4$ (bottom). The dataset was generated with $\sigma = 0.5$. The black dotted lines denote the true number of clusters that were actually used to generate the dataset.

References

- [1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 269–342.
- [2] ARBEL, J. and PRÜNSTER, I. (2017). A moment-matching Ferguson & Klass algorithm. *Statistics and Computing* **27** 3–17.
- [3] ARGIENTO, R., BIANCHINI, I. and GUGLIELMI, A. (2016). A blocked Gibbs sampler for NGG-mixture models via a priori truncation. *Statistics and Computing* **26** 641–661.
- [4] ARGIENTO, R., BIANCHINI, I. and GUGLIELMI, A. s. (2016). Posterior sampling from ε -approximation of normalized completely random measure mixtures. *Electronic Journal of Statistics* **10** 3516–3547.
- [5] AYED, F. and CARON, F. (2021). Nonnegative Bayesian nonparametric factor models with completely random measures for community detection. *Statistics and Computing* **31** 1–24.

- [6] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** 167–241.
- [7] BARRIOS, E., LIJOI, A., NIETO-BARAJAS, L. E. and PRÜNSTER, I. (2013). Modeling with normalized random measure mixture models. *Statistical Science* **28** 313–334.
- [8] BERTOIN, J., FUJITA, T., ROYNETTE, B. and YOR, M. (2006). On a particular class of self-decomposable random variables : the durations of Bessel excursions straddling independent exponential times. *Probability and Mathematical Statistics* **26** 315–366.
- [9] BLEI, D. M. and JORDAN, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian analysis* **1** 121–143.
- [10] BONDESSON, L. (1982). On simulation from infinitely divisible distributions. *Advances in Applied Probability* **14** 855–869.
- [11] BRIX, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability* **31** 929–953.
- [12] CAI, D., CAMPBELL, T. and BRODERICK, T. (2016). Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 4249–4257. Curran Associates, Inc.
- [13] CAMPBELL, T., CAI, D. and BRODERICK, T. (2018). Exchangeable trait allocations. *Electronic Journal of Statistics* **12** 2290–2322.
- [14] CAMPBELL, T., HUGGINS, J. H., HOW, J. P. and BRODERICK, T. (2019). Truncated Random Measures. *Bernoulli*.
- [15] CANALE, A., CORRADIN, R. and NIPOTI, B. (2022). Importance conditional sampling for Pitman-Yor mixtures. *Statistics and Computing* **32** 1–18.
- [16] CARON, F. (2012). Bayesian nonparametric models for bipartite graphs. In *Advances in Neural Information Processing Systems* 25.
- [17] CARON, F. and FOX, E. (2017). Sparse Graphs using Exchangeable Random Measures. *Journal of the Royal Statistical Society B* **79** 1295–1366. Part 5.
- [18] CONT, R. and TANKOV, P. (2004). *Financial modelling with jump processes* **2**. CRC press.
- [19] DEVROYE, L. and JAMES, L. F. (2014). On simulation and properties of the stable law. *Statistical Methods and Applications* **23** 307–343.
- [20] DOSHI, F., MILLER, K., VAN GAEL, J. and TEH, Y. W. (2009). Variational inference for the Indian buffet process. In *Artificial Intelligence and Statistics* 137–144.
- [21] FAVARO, S., LOMELI, M., BERNARDO, N. and TEH, Y. W. (2014). On the stick-breaking representation of σ -stable Poisson-Kingman models. *Electronic Journal of Statistics* **8** 1063–1085.
- [22] FAVARO, S. and TEH, Y. W. (2013). MCMC for Normalized Random Measure Mixture Models. *Statist. Sci.* **28** 335–359.
- [23] FAVARO, S. and WALKER, S. G. (2013). Slice sampling σ -stable Poisson-Kingman mixture models. *Journal of Computational and Graphical Statistics* **22** 830–847.
- [24] FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without gaussian components. *The Annals of Mathematical Statistics* **43** 1634–1643.
- [25] FOTI, N. and WILLIAMSON, S. (2012). Slice sampling normalized kernel-weighted completely random measure mixture models. In *Advances in Neural Information Processing Systems* 2240–2248.
- [26] GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7** 457–511.
- [27] GOPALAN, P., RUIZ, F. J., RANGANATH, R. and BLEI, D. M. (2014). Bayesian Nonparametric Poisson Factorization for Recommendation Systems. In *AISTATS* 275–283.
- [28] GRIFFIN, J. E. (2017). Sequential Monte Carlo methods for mixtures with normalized random measures with independent increments priors. *Statistics and Computing* **27** 131–145.

- [29] GRIFFIN, J. E. and WALKER, S. G. (2011). Posterior simulation of normalized random measure mixtures. *Journal of Computational and Graphical Statistics* **20** 241–259.
- [30] GRIFFITHS, T. L. and GHARAMANI, Z. (2006). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems 18 (NIPS 2005)*.
- [31] HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics* **18** 1259–1294.
- [32] HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73** 387–396.
- [33] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96** 161–173.
- [34] ISHWARAN, H. and ZAREPOUR, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics* **30** 269–283.
- [35] JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics* **33** 1771–1799.
- [36] JAMES, L. F. (2017). Bayesian Poisson calculus for latent feature modeling via generalized Indian buffet process priors. *The Annals of Statistics* **45** 2016–2045.
- [37] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics* **36** 76–97.
- [38] KALLI, M., GRIFFIN, J. E. and WALKER, S. G. (2011). Slice sampling mixture models. *Statistics and computing* **21** 93–105.
- [39] KHINTCHINE, A. (1937). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Rec. Math. Moscou, n. Ser* **2** 79–117.
- [40] KINGMAN, J. F. C. (1993). *Poisson processes* **3**. Oxford University Press, USA.
- [41] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics* **21** 59–78.
- [42] KINGMAN, J. F. C. (1975). Random discrete distributions. *Journal of the Royal Statistical Society: Series B (Methodological)* **37** 1–15.
- [43] LEE, J., JAMES, L. F. and CHOI, S. (2016). Finite-dimensional BFRY priors and variational Bayesian inference for power-law models. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*.
- [44] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association* **100** 1278–1291.
- [45] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 715–740.
- [46] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Non-parametrics* (N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds.) Cambridge University Press.
- [47] LO, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *Ann. Statist.* **12** 351–357.
- [48] MOLLER, J. and WAAGEPETERSEN, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC.
- [49] NEAL, R. (2003). Slice sampling. *The Annals of Statistics* **31** 705–767.
- [50] NGUYEN, T. D., HUGGINS, J., MASOERO, L., MACKAY, L. and BRODERICK, T. (2020). Independent finite approximations for Bayesian nonparametric inference. *arXiv preprint arXiv:2009.10780*.
- [51] NIETO-BARAJAS, L. E., PRÜNSTER, I. and WALKER, S. G. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* **32** 2343–2360.

- [52] PAISLEY, J., BLEI, D. and JORDAN, M. L. (2012). Stick-breaking beta processes and the Poisson process. In *Artificial Intelligence and Statistics* 850–858.
- [53] PAISLEY, J. and CARIN, L. (2009). Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th Annual International Conference on Machine Learning* 777–784. ACM.
- [54] PAISLEY, J., CARIN, L. and BLEI, D. (2011). Variational Inference for Stick-Breaking Beta Process Priors. In *ICML* 889–896. Citeseer.
- [55] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields* **92** 21–39.
- [56] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics* **31** 560–585.
- [57] ROSIŃSKI, J. (1990). On series representations of infinitely divisible random vectors. *The Annals of Probability* 405–430.
- [58] ROSIŃSKI, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes - Theory and Applications* (O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, eds.) 401–415. Birkhäuser, Boston.
- [59] ROSIŃSKI, J. (2001). Contribution to the discussion of the paper “Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics” by O. Barndorff-Nielsen and N. Shephard. *Journal of the Royal Statistical Society, Series B* **63** 230–231.
- [60] ROSIŃSKI, J. (2007). Tempering stable processes. *Stochastic processes and their applications* **117** 677–707.
- [61] ROYCHOWDHURY, A. and KULIS, B. (2015). Gamma processes, stick-breaking, and variational inference. In *Artificial Intelligence and Statistics* 800–808.
- [62] TEH, Y. W. and GORUR, D. (2009). Indian buffet processes with power-law behavior. In *Advances in neural information processing systems* 1838–1846.
- [63] TEH, Y. W., GÖRÜR, D. and GHAHRAMANI, Z. (2007). Stick-breaking construction for the Indian buffet process. In *Artificial Intelligence and Statistics* 556–563.
- [64] TITSIAS, M. K. (2008). The infinite gamma-Poisson feature model. In *Advances in Neural Information Processing Systems* 1513–1520.
- [65] WALKER, S. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation* **36** 45–54.
- [66] WOLPERT, R. L. and ICKSTADT, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267.
- [67] WOLPERT, R. L. and ICKSTADT, K. (1998). Simulation of Lévy random fields. In *Practical nonparametric and semiparametric Bayesian statistics* 227–242. Springer.
- [68] ZHOU, M. (2015). Infinite edge partition models for overlapping community detection and link prediction. In *Artificial Intelligence and Statistics* 1135–1143.
- [69] ZHOU, M., HANNAH, L., DUNSON, D. and CARIN, L. (2012). Beta-negative binomial process and Poisson factor analysis. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (N. D. LAWRENCE and M. GIROLAMI, eds.). *Proceedings of Machine Learning Research* **22** 1462–1471. PMLR, La Palma, Canary Islands.