Bernoulli **25**(2), 2019, 1256–1288
https://doi.org/10.3150/18-BEJ1020

Truncated random measures

TREVOR CAMPBELL[1,*], JONATHAN H. HUGGINS[1,**], JONATHAN P. HOW[2] and TAMARA BRODERICK[1,†]

[1]*Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: [*] tdjc@mit.edu; [**] jhuggins@mit.edu; [†] tbroderick@csail.mit.edu*
[2]*Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: jhow@mit.edu*

Completely random measures (CRMs) and their normalizations are a rich source of Bayesian nonparametric priors. Examples include the beta, gamma, and Dirichlet processes. In this paper, we detail two major classes of sequential CRM representations—*series representations* and *superposition representations*—within which we organize both novel and existing sequential representations that can be used for simulation and posterior inference. These two classes and their constituent representations subsume existing ones that have previously been developed in an ad hoc manner for specific processes. Since a complete infinite-dimensional CRM cannot be used explicitly for computation, sequential representations are often truncated for tractability. We provide truncation error analyses for each type of sequential representation, as well as their normalized versions, thereby generalizing and improving upon existing truncation error bounds in the literature. We analyze the computational complexity of the sequential representations, which in conjunction with our error bounds allows us to directly compare representations and discuss their relative efficiency. We include numerous applications of our theoretical results to commonly-used (normalized) CRMs, demonstrating that our results enable a straightforward representation and analysis of CRMs that has not previously been available in a Bayesian nonparametric context.

Keywords: Bayesian nonparametrics; completely random measure; normalized completely random measure; Poisson point process; truncation

1. Introduction

In many data sets, we can view the data points as exhibiting a collection of underlying traits. For instance, each document in the *New York Times* might touch on a number of topics or themes, an individual's genetic data might be a product of the populations to which their ancestors belonged, or a user's activity on a social network might be dictated by their varied personal interests. When the traits are not directly observed, a common approach is to model each trait as having some frequency or rate in the broader population [2]. The inferential goal is to learn these rates as well as whether—and to what extent—each data point exhibits each trait. Since the traits are unknown a priori, their cardinality is also typically unknown.

As a data set grows larger, we can reasonably expect the number of traits to increase as well. In the cases above, for example, we expect to uncover more topics as we read more documents, more ancestral populations as we examine more individuals' genetic data, and more unique interests as we observe more individuals on a social network. *Bayesian nonparametric* (BNP) priors provide a flexible, principled approach to creating models in which the number of exhibited traits is random, can grow without bound, and may be learned as part of the inferential procedure. By

1350-7265

generating a countable infinity of potential traits—where any individual data point exhibits only finitely many—these models enable growth in the number of observed traits with the size of the data set.

In practice, however, it is impossible to store a countable infinity of random variables in memory or learn the distribution over a countable infinity of variables in finite time. Conjugate priors and likelihoods have been developed [36] that theoretically circumvent the infinite representation altogether and perform exact Bayesian posterior inference [11]. However, these priors and likelihoods are often just a single piece within a more complex generative model, and ultimately an approximate posterior inference scheme such as Markov Chain Monte Carlo (MCMC) or variational Bayes (VB) is required. These approximation schemes often necessitate a full and explicit representation of the latent variables.

One option is to approximate the infinite-dimensional prior with a related finite-dimensional prior: that is, to replace the infinite collection of random traits by a finite subset of "likely" traits. To do so, first enumerate the countable infinity of traits in the full model and write (ψ_k, θ_k) for each paired trait ψ_k (e.g., a topic in a document) and its rate or frequency θ_k. Then the discrete measure $\Theta := \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$ captures the traits/rates in a sequence indexed by k. The (ψ_k, θ_k) pairs are random in the Bayesian model, so Θ is a random measure. In many cases, the distribution of Θ can be defined by specifying a sequence of simple, familiar distributions for the finite-dimensional ψ_k and θ_k, known as a *sequential representation*. Given a sequential representation of Θ, a natural way to choose a subset of traits is to keep the first $K < \infty$ traits and discard the rest, resulting in an approximate measure Θ_K. This approach is called *truncation*. Note that it is also possible to truncate by removing atoms with weights less than a specified threshold [4,35], though this approach is not as easily incorporated in posterior inference algorithms.

Sequential representations have been shown to exist for *completely random measures* (CRMs) [15,27], a large class of nonparametric priors that includes such popular models as the beta process [18,26] and the gamma process [8,15,23,29,50]. Numerous sequential representations of CRMs have been developed in the literature [7,11,15,24,43,44]. CRM priors are often paired with likelihood processes—such as the Bernoulli process [49], negative binomial process [10,51], and Poisson likelihood process [50]. The likelihood process determines how much each trait is expressed by each data point. Sequential representations also exist for *normalized completely random measures* (NCRMs) (sometimes referred to as normalized random measures with independent increments) [22,25,33,39–42], which provide random distributions over traits, such as the Dirichlet process [14,47]. NCRMs are typically paired with a likelihood that assigns each data point to a single trait using the NCRM as a discrete distribution.

Since (N)CRMs have many possible sequential representations, a method is required for determining which to use for the application at hand and, once a representation is selected, for choosing a truncation level. Our main contributions enable the principled selection of both representation and truncation level using approximation error:

1. We provide a comprehensive characterization of the different types of sequential representations for (N)CRMs, filling in many gaps in the literature of sequential representations along the way. We classify these representations into two major groups: *series representations*, which are constructed by transforming a homogeneous Poisson point process; and *superposition representations*, which are the superposition of infinitely many Poisson point

processes with finite rate measures. We also introduce two novel sequential representations for (N)CRMs.

2. We provide theoretical guarantees on the approximation error induced when truncating these sequential representations. We give the error as a function of the prior process, the likelihood process, and the level of truncation. While truncation error bounds for (N)CRMs have been studied previously, past work has focused on specific combinations of (N)CRM priors and likelihoods—in particular, the Dirichlet-multinomial [6,19,21, 47], beta-Bernoulli [13,37], generalized beta-Bernoulli [45], and gamma-Poisson [46] processes. In the current work, we give much more general results for bounding the truncation error.

Our results fill in large gaps in the analysis of truncation error, which is often measured in terms of the L^1 (a.k.a. total variation) distance between the data distributions induced by the full and truncated priors. We provide the first analysis of truncation error for some sequential representations of the beta process with Bernoulli likelihood [49], for the beta process with negative binomial likelihood [10,51], and for the normalization of the generalized gamma process [8], the σ-stable process, and the generalized inverse gamma [30,33] with discrete likelihood. Moreover, even when truncation results already exist in the literature [13,19,37,46], we improve on those error bounds by a factor of two. The reduction arises from our use of the point process machinery of CRMs, circumventing the total variation bound used originally by [19,20] upon which most modern truncation analyses are built. We obtain our truncation error guarantees by bounding the probability that data drawn from the full model will use a feature that is not available to the truncated model. Thinking in terms of this probability provides a more intuitive interpretation of our bounds that can be communicated to practitioners and used to guide them in their choice of truncation level.

The remainder of this paper is organized as follows. In Section 2, we provide background material on CRMs and establish notation. In our first main theoretical section, Section 3, we describe seven different sequential CRM representations, including four series representations and three superposition representations, two of which are novel. Next, we provide a general theoretical analysis of the truncation error for series and superposition representations in Section 4. We provide analogous theory for the normalized versions of each representation in Section 5 via an infinite extension of the "Gumbel-max trick" [17,34]. We determine the complexity of simulating each representation in Section 6. In Section 7, we summarize our results (Table 1) and provide advice on how to select sequential representations in practice. Proofs for all results developed in this paper are provided in the the supplemental article [12].

2. Background

2.1. CRMs and truncation

Consider a Poisson point process on $\mathbb{R}_+ := [0, \infty)$ with rate measure $\nu(\mathrm{d}\theta)$ such that

$$\nu(\mathbb{R}_+) = \infty \quad \text{and} \quad \int \min(1, \theta)\nu(\mathrm{d}\theta) < \infty. \tag{2.1}$$

Such a process generates a countable infinity of values $(\theta_k)_{k=1}^{\infty}$, $\theta_k \in \mathbb{R}_+$, having an almost surely finite sum $\sum_{k=1}^{\infty} \theta_k < \infty$. In a BNP trait model, we interpret each θ_k as the rate or frequency of the kth trait. Typically, each θ_k is paired with a parameter ψ_k associated with the kth trait (e.g., a topic in a document or a shared interest on a social network). We assume throughout that $\psi_k \in \Psi$ for some space Ψ and $\psi_k \overset{\text{i.i.d.}}{\sim} G$ for some distribution G. Constructing a measure by placing mass θ_k at atom location ψ_k results in a *completely random measure* (CRM) [27]. As shorthand, we will write CRM(ν) for the completely random measure generated as just described:

$$\Theta := \sum_k \theta_k \delta_{\psi_k} \sim \text{CRM}(\nu).$$

The trait distribution G is left implicit in the notation as it has no effect on our results. Further, the possible fixed-location and deterministic components of a CRM [27] are not considered here for brevity; these components can be added (assuming they are purely atomic) and the analysis modified without undue effort. The CRM prior on Θ is typically combined with a likelihood that generates trait counts for each data point. Let $h(\cdot \mid \theta)$ be a proper probability mass function on $\mathbb{N} \cup \{0\}$ for all θ in the support of ν (though the present work may be easily extended to likelihoods with support in $\mathbb{R}$). Then a collection of conditionally independent observations $X_{1:N} := \{X_n\}_{n=1}^N$ given Θ are distributed according to the *likelihood process* LP(h, Θ), that is,

$$X_n := \sum_k x_{nk} \delta_{\psi_k} \overset{\text{i.i.d.}}{\sim} \text{LP}(h, \Theta),$$

if $x_{nk} \sim h(x \mid \theta_k)$ independently across k and i.i.d. across n. The desideratum that each X_n expresses a finite number of traits is encoded by the assumption that

$$\int \big(1 - h(0 \mid \theta)\big) \nu(\mathrm{d}\theta) < \infty.$$

Since the trait counts are typically latent in a full generative model specification, define the observed data $Y_n \mid X_n \overset{\text{indep}}{\sim} f(\cdot \mid X_n)$ for a conditional density f with respect to a measure μ on some space. For instance, if the sequence $(\theta_k)_{k=1}^{\infty}$ represents the topic rates in a document corpus, X_n might capture how many words in document n are generated from each topic and Y_n might be the observed collection of words for that document.

Since the sequence $(\theta_k)_{k=1}^{\infty}$ is countably infinite, it may be difficult to simulate or perform posterior inference in this model. One approximation scheme is to define the *truncation* $\Theta_K := \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$. Since it is finite, the truncation Θ_K can be used for exact simulation or in posterior inference—but some error arises from not using the full CRM Θ. To quantify this error, consider its propagation through the above Bayesian model. Define $Z_{1:N}$ and $W_{1:N}$ for Θ_K analogous to the definitions of $X_{1:N}$ and $Y_{1:N}$ for Θ:

$$Z_n \mid \Theta_K \overset{\text{i.i.d.}}{\sim} \text{LP}(h, \Theta_K), \qquad W_n \mid Z_n \overset{\text{indep}}{\sim} f(\cdot \mid Z_n), \qquad n = 1, \ldots, N.$$

A standard approach to measuring the distance between Θ and Θ_K is to use the L^1 metric between the marginal densities $p_{N,\infty}$ and $p_{N,K}$ (with respect to some measure μ) of the final

observations $Y_{1:N}$ and $W_{1:N}$ [13,19,37]:

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 := \frac{1}{2}\int_{y_{1:N}} \left|p_{N,\infty}(y_{1:N}) - p_{N,K}(y_{1:N})\right| \mu(\mathrm{d}y_{1:N}).$$

All of our bounds on $\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1$ are also bounds on the probability that $X_{1:N}$ contains a feature that is not in the truncation Θ_K (cf. Sections 4 and 5). This interpretation may be easier to digest since it does not depend on the observation model f and is instead framed in terms of the underlying traits the practitioner is trying to estimate.

2.2. The gamma-Poisson process

To illustrate the practical application of the theoretical developments in this work, we provide a number of examples throughout involving the *gamma process* [8], denoted $\Gamma\mathrm{P}(\gamma, \lambda, d)$, with discount parameter $d \in [0, 1)$, scale parameter $\lambda > 0$, mass parameter $\gamma > 0$, and rate measure

$$\nu(\mathrm{d}\theta) = \gamma \frac{\lambda^{1-d}}{\Gamma(1-d)} \theta^{-d-1} e^{-\lambda\theta}\, \mathrm{d}\theta.$$

Setting $d = 0$ yields the undiscounted gamma process [15,29,50]. The gamma process is often paired with a Poisson likelihood,

$$h(x \mid \theta) = \frac{\theta^x}{x!} e^{-\theta}.$$

Throughout the present work, we use the rate parametrization of the gamma distribution (to match the gamma process parametrization), for which the density is given by

$$\mathsf{Gam}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}.$$

Appendix A provides additional example applications of our theoretical results for two other CRMs: the beta process $\mathrm{BP}(\gamma, \alpha, d)$ [9,48] with Bernoulli or negative binomial likelihood, and the beta prime process $\mathrm{BPP}(\gamma, \alpha, d)$ [11] with odds-Bernoulli likelihood.

3. Sequential representations

Sequential representations are at the heart of the study of truncated CRMs. They provide an iterative method that can be terminated at any point to yield a finite approximation to the infinite process, where the choice of termination point determines the accuracy of the approximation. Thus, the natural first step in providing a coherent treatment of truncation analysis is to do the same for sequential representations. In past work, two major classes of sequential representation have been used: *series representations* of the form $\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$, and *superposition representations* of the form $\sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}$, where each inner sum of C_k atoms is itself a CRM. This section

examines four series representations [7,15,43,44] and three superposition representations (two of which are novel) [9,11,24]. We show how previously-developed sequential representations for specific CRMs fit into these seven general representations. Finally, we discuss a stochastic mapping procedure that is useful in obtaining new representations from the transformation of others. Proofs for the results in this section may be found in Appendix C.

3.1. Series representations

Series representations arise from the transformation of a homogeneous Poisson point process [44]. They tend to be somewhat difficult to analyze due to the dependence between the atoms but also tend to produce very simple representations with small truncation error (cf. Sections 4 and 7). Throughout the paper we let $\Gamma_k = \sum_{\ell=1}^{k} E_\ell$, $E_\ell \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(1)$, be the ordered jumps of a unit-rate homogeneous Poisson process on $\mathbb{R}_+$, let ν be a measure on $\mathbb{R}_+$ satisfying the basic conditions in Eq. (2.1), and let $\psi_k \overset{\text{i.i.d.}}{\sim} G$.

Inverse-Lévy [15]

Define $\nu^{\leftarrow}(u) := \inf\{x : \nu([x,\infty)) \leq u\}$, the inverse of the tail measure $\nu([x,\infty))$. We say Θ has an *inverse-Lévy* representation and write $\Theta \leftarrow \text{IL-Rep}(\nu)$ if

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \text{with } \theta_k = \nu^{\leftarrow}(\Gamma_k).$$

Ferguson and Klass [15] showed that $\Theta \leftarrow \text{IL-Rep}(\nu)$ implies $\Theta \sim \text{CRM}(\nu)$. The inverse-Lévy representation is analogous to the inverse CDF method for generating an arbitrary random variable from a uniform random variable, with the homogenous Poisson process playing the role of the uniform random variable. It is also the optimal sequential representation in the sense that the sequence θ_k that it generates is non-increasing. While an elegant and general approach, simulating the inverse-Lévy representation is difficult, as inverting the function $\nu([x,\infty))$ is analytically intractable except in a few cases.

Example 3.1 (Gamma process, $\Gamma\text{P}(\gamma,\lambda,0)$). We have $\nu([x,\infty)) = \gamma\lambda E_1(\lambda x)$, where $E_1(x) := \int_x^\infty u^{-1}e^{-u}\,\mathrm{d}u$ is the exponential integral function [1]. The inverse-Lévy representation for $\Gamma\text{P}(\gamma,\lambda,0)$ is thus

$$\Theta = \sum_{k=1}^{\infty} \lambda^{-1} E_1^{-1}\big(\gamma^{-1}\lambda^{-1}\Gamma_k\big)\delta_{\psi_k}.$$

Neither E_1 nor its inverse can be computed in closed form, so one must resort to numerical approximations.

Bondesson [7]

We say Θ has a *Bondesson representation* and write $\Theta \leftarrow \text{B-Rep}(c, g)$ if for $c > 0$ and g a density on $\mathbb{R}_+$,

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \text{with } \theta_k = V_k e^{-\Gamma_k / c},\ V_k \overset{\text{i.i.d.}}{\sim} g.$$

Theorem 3.1 shows that Bondesson representations can be constructed for a large, albeit restricted, class of CRM rate measures. We offer a novel proof of Theorem 3.1 in Appendix C using the induction strategy introduced by [5]. Similar proof ideas are also used to prove truncation error bounds for sequential representations in Section 4. We use a slight abuse of notation for brevity: if $\nu(\mathrm{d}\theta)$ is a measure on $\mathbb{R}_+$ that is absolutely continuous with respect to Lebesgue measure, then $\nu(\theta)$ is the density of $\nu(\mathrm{d}\theta)$ with respect to the Lebesgue measure.

Theorem 3.1 (Bondesson representation [7]). *Let $\nu(\mathrm{d}\theta) = \nu(\theta)\,\mathrm{d}\theta$ be a rate measure satisfying Eq. (2.1). If $\theta\nu(\theta)$ is nonincreasing, $\lim_{\theta\to\infty} \theta\nu(\theta) = 0$, and $c_\nu := \lim_{\theta\to 0} \theta\nu(\theta) < \infty$, then $g_\nu(v) := -c_\nu^{-1} \frac{\mathrm{d}}{\mathrm{d}v}[v\nu(v)]$ is a density on $\mathbb{R}_+$ and*

$$\Theta \leftarrow \text{B-Rep}(c_\nu, g_\nu) \quad \textit{implies} \quad \Theta \sim \text{CRM}(\nu).$$

Example 3.2 (Bondesson representation for $\Gamma\text{P}(\gamma, \lambda, 0)$). The following representation for the gamma process with $d = 0$ was described by [7] and [5]. Since $\theta\nu(\theta) = \gamma\lambda e^{-\lambda\theta}$ is non-increasing and $c_\nu = \lim_{\theta\to 0} \theta\nu(\theta) = \gamma\lambda$, we obtain $g_\nu(v) = \lambda e^{-\lambda v} = \mathsf{Exp}(v; \lambda)$. Thus, it follows from Theorem 3.1 that if $\Theta \leftarrow \text{B-Rep}(\gamma\lambda, \mathsf{Exp}(\lambda))$, then $\Theta \sim \Gamma\text{P}(\gamma, \lambda, 0)$. The condition that $\theta\nu(\theta)$ is non-increasing fails to hold if $d > 0$, so we cannot apply Theorem 3.1 to $\Gamma\text{P}(\gamma, \lambda, d)$ when $d > 0$.

Thinning [43]

Using the nomenclature of [44], we say Θ has a *thinning representation* and write $\Theta \leftarrow \text{T-Rep}(\nu, g)$ if g is a probability measure on $\mathbb{R}_+$ such that ν is absolutely continuous with respect to g, i.e. $\nu \ll g$, and

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \text{with } \theta_k = V_k \mathbb{1}\left(\frac{\mathrm{d}\nu}{\mathrm{d}g}(V_k) \geq \Gamma_k\right),\ V_k \overset{\text{i.i.d.}}{\sim} g.$$

Rosiński [43] showed that $\Theta \leftarrow \text{T-Rep}(\nu, g)$ implies $\Theta \sim \text{CRM}(\nu)$. Note that $\Gamma_k \overset{\text{a.s.}}{\to} \infty$ as $k \to \infty$, so the probability that $\frac{\mathrm{d}\nu}{\mathrm{d}g}(V_k) \geq \Gamma_k$ is decreasing in k. Thus, this representation generates atoms with $\theta_k = 0$ (which have no effect and can be removed) increasingly frequently and becomes inefficient as $k \to \infty$.

Example 3.3 (Thinning representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$). If we let $g = \mathsf{Gam}(1-d, \lambda)$, then the thinning representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$ is

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}(V_k \Gamma_k \le \gamma)\delta_{\psi_k} \qquad \text{with } V_k \overset{\text{i.i.d.}}{\sim} \mathsf{Gam}(1-d, \lambda).$$

Rejection [44]

Using the nomenclature of [44], we say Θ has a *rejection representation* and write $\Theta \leftarrow \text{R-Rep}(\nu, \mu)$ if μ is a measure on $\mathbb{R}_+$ satisfying Eq. (2.1) and $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \le 1$, and

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \text{with } \theta_k = V_k \mathbb{1}\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(V_k) \ge U_k\right), (V_k)_{k\in\mathbb{N}} \sim \text{PoissP}(\mu), U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0,1].$$

Rosiński [44] showed that $\Theta \leftarrow \text{R-Rep}(\nu, \mu)$ implies $\Theta \sim \text{CRM}(\nu)$. This representation is very similar to the thinning representation, except that the sequence $(V_k)_{k\in\mathbb{N}}$ is generated from a Poisson process on $\mathbb{R}_+$ rather than i.i.d. This allows $V_k \overset{\text{a.s.}}{\to} 0$ as $k \to \infty$, causing the frequency of generating ineffective atoms $\theta_k = 0$ to decay as $k \to \infty$, assuming μ is appropriately chosen such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\theta) \to 1$ as $\theta \to 0$. This representation can thus be constructed to be more efficient than the thinning representation. We can calculate the efficiency in terms of the expected number of rejections (that is, the number of θ_k that are identically zero):

Proposition 3.2. *For* $\text{R-Rep}(\nu, \mu)$, *the expected number of rejections is*

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \mathbb{1}(\theta_k = 0)\right] = \int \left(1 - \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x).$$

Remark. If μ and ν can be written as densities with respect to Lebesgue measure, then the integral in Proposition 3.2 can be rewritten as $\int(\mu(x) - \nu(x))\,\mathrm{d}x$.

Example 3.4 (Rejection representation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$). Following [44], consider $\mu(\mathrm{d}\theta) = \gamma\lambda\theta^{-1}(1+\lambda\theta)^{-1}\,\mathrm{d}\theta$. We call $\text{CRM}(\mu)$ the Lomax process, $\text{LomP}(\gamma, \lambda^{-1})$, after the related Lomax distribution. We can use the inverse-Lévy method analytically with μ since $\mu^{\leftarrow}(u) = \frac{1}{\lambda(e^{(\gamma\lambda)^{-1}u} - 1)}$. Thus, the rejection representation of $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ is

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}\big(U_k \le (1+\lambda V_k)e^{-\lambda V_k}\big)\delta_{\psi_k} \qquad \text{with } V_k = \frac{1}{\lambda(e^{(\gamma\lambda)^{-1}\Gamma_k} - 1)}, U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0,1].$$

Unlike in the thinning construction given in Example 3.3, only a finite number of rates will be set to zero almost surely. In particular, the expected number of rejections is $\gamma\lambda c_\gamma$, where c_γ is the Euler–Mascheroni constant.

Example 3.5 (Rejection representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$, $d > 0$). For the case of $d > 0$, we instead use $\mu(\mathrm{d}\theta) = \gamma \frac{\lambda^{1-d}}{\Gamma(1-d)} \theta^{-1-d}\, \mathrm{d}\theta$. We can again use the inverse-Lévy method analytically with μ since $\mu^{\leftarrow}(u) = (\gamma' u^{-1})^{1/d}$, where $\gamma' := \gamma \frac{\lambda^{1-d}}{d\Gamma(1-d)}$. The rejection representation is then

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}\big(U_k \leq e^{-\lambda V_k}\big)\delta_{\psi_k} \qquad \text{with } V_k = \big(\gamma' \Gamma_k^{-1}\big)^{1/d},\ U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0,1].$$

The expected number of rejections is $\gamma \frac{\lambda^{1-d}}{d}$, so the representation is efficient for large d, but extremely inefficient when d is small.

3.2. Superposition representations

Superposition representations arise as an infinite sum of CRMs with finite rate measure. These tend to be easier to analyze than series representations as they decouple atoms between the summed CRMs, but can produce representations with larger truncation error (cf. Sections 4 and 7). Throughout, let ν be a measure on $\mathbb{R}_+$ satisfying the basic conditions in Eq. (2.1), and let $\psi_k \overset{\text{i.i.d.}}{\sim} G$.

Decoupled Bondesson

We say Θ has a *decoupled Bondesson representation* and write $\Theta \leftarrow \text{DB-Rep}(c, g, \xi)$ if for $c > 0, \xi > 0$, and g a density on $\mathbb{R}_+$,

$$\Theta = \sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}} \qquad \text{with}$$
$$C_k \overset{\text{i.i.d.}}{\sim} \mathsf{Poiss}(c/\xi),\ \theta_{ki} = V_{ki} e^{-T_{ki}},\ T_{ki} \overset{\text{indep}}{\sim} \mathsf{Gam}(k, \xi),\ V_{ki} \overset{\text{i.i.d.}}{\sim} g. \tag{3.1}$$

This is a novel superposition representation, though special cases are already known [38,46]. Theorem 3.3 shows that the decoupled Bondesson representation applies to the same class of CRMs as the Bondesson representation from Section 3.1.

Theorem 3.3 (Decoupled Bondesson representation). *Let $\nu(\mathrm{d}\theta) = \nu(\theta)\,\mathrm{d}\theta$, c_ν, and g_ν be as specified in Theorem 3.1. Then for any fixed $\xi > 0$,*

$$\Theta \leftarrow \text{DB-Rep}(c_\nu, g_\nu, \xi) \quad \textit{implies} \quad \Theta \sim \text{CRM}(\nu).$$

The proof of Theorem 3.3 in Appendix C generalizes the arguments from [38] and [46]. The free parameter ξ controls the number of atoms generated for each outer sum index k; its principled selection can be made by trading off computational complexity (cf. Section 6) and truncation error (cf. Section 4).

Example 3.6 (Decoupled Bondesson representation for $\Gamma\mathrm{P}(\gamma,\lambda,0)$). Arguments paralleling those made in Example 3.2 show that the $\Gamma\mathrm{P}(\gamma,\lambda,0)$ representation from [46] follows directly from an application of Theorem 3.3: if $\Theta \leftarrow \text{DB-Rep}(\gamma\lambda, \mathsf{Exp}(\lambda), \xi)$, then $\Theta \sim \Gamma\mathrm{P}(\gamma,\lambda,0)$. As in the Bondesson representation setting, Theorem 3.3 does not apply to $\Gamma\mathrm{P}(\gamma,\lambda,d)$ when $d>0$ because the condition that $\theta\nu(\theta)$ is non-increasing fails to hold.

Size-biased [11,24]

Let $\pi(\theta) := h(0 \mid \theta)$. We say Θ has a *size-biased representation* and write $\Theta \leftarrow \text{SB-Rep}(\nu, h)$ if

$$\begin{aligned}&\Theta = \sum_{k=1}^{\infty}\sum_{i=1}^{C_k}\theta_{ki}\delta_{\psi_{ki}} \qquad \text{with}\\ &C_k \overset{\text{indep}}{\sim} \mathsf{Poiss}(\eta_k), \theta_{ki} \overset{\text{indep}}{\sim} \frac{1}{\eta_k}\pi(\theta)^{k-1}\big(1-\pi(\theta)\big)\nu(\mathrm{d}\theta), \eta_k := \int \pi(\theta)^{k-1}\big(1-\pi(\theta)\big)\nu(\mathrm{d}\theta).\end{aligned} \tag{3.2}$$

Broderick et al. [11] and James [24] showed that $\Theta \leftarrow \text{SB-Rep}(\nu, h)$ implies $\Theta \sim \mathrm{CRM}(\nu)$. If the rate measure ν and the likelihood h are selected to be a conjugate exponential family then, noting that $\sum_{x=1}^{\infty} h(x \mid \theta) = 1 - \pi(\theta)$, the rate θ_{ki} can be sampled from a mixture of exponential family distributions:

$$\begin{aligned}&\theta_{ki} \mid z_{ki} \overset{\text{indep}}{\sim} \frac{1}{\eta_{kz_{ki}}} h(z_{ki} \mid \theta)\pi(\theta)^{k-1}\nu(\mathrm{d}\theta), \qquad z_{ki} \overset{\text{indep}}{\sim} \mathsf{Categorical}\big((\eta_{kx}/\eta_k)_{x=1}^{\infty}\big),\\ &\eta_{kx} := \int h(x \mid \theta)\pi(\theta)^{k-1}\nu(\mathrm{d}\theta).\end{aligned}$$

Example 3.7 (Size-biased representation for $\Gamma\mathrm{P}(\gamma,\lambda,d)$). For the Gamma process, values for η_{kx} and η_k can be found using integration by parts and the standard gamma distribution integral, while $\theta_{ki} \mid z_{ki}$ is sampled from a gamma distribution by inspection:

$$\eta_{kx} = \frac{\gamma\lambda^{1-d}\Gamma(x-d)}{x!\Gamma(1-d)(\lambda+k)^{x-d}}, \qquad \eta_k = \begin{cases}\dfrac{\gamma\lambda^{1-d}}{d}\big((\lambda+k)^d - (\lambda+k-1)^d\big), & d>0,\\ \gamma\lambda\big(\log(\lambda+k) - \log(\lambda+k-1)\big), & d=0,\end{cases}$$

$$\theta_{ki} \mid z_{ki} \overset{\text{indep}}{\sim} \mathsf{Gam}(x-d, \lambda+k).$$

Power-law

We say Θ has a *power-law representation* and write $\Theta \leftarrow \text{PL-Rep}(\gamma,\alpha,d,g)$ if for $\gamma>0$, $0 \le d < 1$, $\alpha > -d$, and g a density on $\mathbb{R}_+$,

$$\begin{aligned}&\Theta = \sum_{k=1}^{\infty}\sum_{i=1}^{C_k}\theta_{ki}\delta_{\psi_{ki}} \qquad \text{with}\\ &C_k \overset{\text{i.i.d.}}{\sim} \mathsf{Poiss}(\gamma), \theta_{ki} = V_{ki}U_{kik}\prod_{j=1}^{k-1}(1-U_{kij}), V_{ki} \overset{\text{i.i.d.}}{\sim} g, U_{kij} \overset{\text{indep}}{\sim} \mathsf{Beta}(1-d, \alpha+jd).\end{aligned} \tag{3.3}$$

This is a novel superposition representation, although it was previously developed in the special case of the beta process (where $g(v) = \delta_1$) [9]. The name of this representation arises from the fact that it exhibits Types I and II power-law behavior [9] under mild conditions when $d > 0$, as we show in Theorem C.1 in the Appendix (note, however, that it will not exhibit power-law behavior when $d = 0$). Theorem 3.5 below shows the conditions under which $\Theta \leftarrow \text{PL-Rep}(\gamma, \alpha, d, g)$ implies $\Theta \sim \text{CRM}(\nu)$. Its proof in Appendix C relies on the notion of *stochastic mapping* (Lemma 3.4), a powerful technique for transforming one CRM into another. Note that in Lemma 3.4, the case where u is a deterministic function of θ via the mapping $u = \tau(\theta)$ may be recovered by setting $\kappa(\theta, \mathrm{d}u) = \delta_{\tau(\theta)}$.

Lemma 3.4 (CRM stochastic mapping). *Let $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \sim \text{CRM}(\nu)$. Then for any probability kernel $\kappa(\theta, \mathrm{d}u)$, we have $\kappa(\Theta) \sim \text{CRM}(\nu_\kappa)$, where*

$$\kappa(\Theta) := \sum_{k=1}^{\infty} u_k \delta_{\psi_k}, \qquad u_k \mid \theta_k \sim \kappa(\theta_k, \cdot) \text{ and } \nu_\kappa(\mathrm{d}u) := \int \kappa(\theta, \mathrm{d}u)\nu(\mathrm{d}\theta).$$

Theorem 3.5 (Power-law representation). *Let $\nu(\mathrm{d}\theta) = \nu(\theta)\,\mathrm{d}\theta$ be a rate measure satisfying Eq. (2.1), and let g_ν be a density on $\mathbb{R}_+$ such that*

$$\nu(u) = \int \theta^{-1} g_\nu\big(u\theta^{-1}\big)\nu_{\text{BP}}(\mathrm{d}\theta),$$

where

$$\nu_{\text{BP}}(\mathrm{d}\theta) = \gamma \frac{\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)} \mathbb{1}[\theta \le 1]\theta^{-1-d}(1-\theta)^{\alpha+d-1}\,\mathrm{d}\theta$$

is the rate measure for the beta process $\text{BP}(\gamma, \alpha, d)$ from Eq. (A.1). Then

$$\Theta \leftarrow \text{PL-Rep}(\gamma, \alpha, d, g_\nu) \quad \textit{implies} \quad \Theta \sim \text{CRM}(\nu).$$

Example 3.8 (Power-law representation for $\Gamma\text{P}(\gamma, \lambda, d)$). If we choose $g_\nu = \mathsf{Gam}(\lambda, \lambda)$, then using the change of variable $w = u(\theta^{-1} - 1)$,

$$\begin{aligned}
&\int \theta^{-1} g_\nu\big(u\theta^{-1}\big)\nu_{\text{BP}}(\mathrm{d}\theta) \\
&= \gamma\lambda \frac{\lambda^\lambda}{\Gamma(1-d)\Gamma(\lambda+d)} u^{\lambda-1} \int \theta^{-\lambda-d-1} e^{-\lambda u\theta^{-1}} (1-\theta)^{\lambda+d-1}\,\mathrm{d}\theta\,\mathrm{d}u \\
&= \gamma\lambda \frac{\lambda^\lambda}{\Gamma(1-d)\Gamma(\lambda+d)} u^{-1-d} e^{-\lambda u} \int w^{\lambda+d-1} e^{-\lambda w}\,\mathrm{d}w\,\mathrm{d}u \\
&= \gamma \frac{\lambda^{1-d}}{\Gamma(1-d)} u^{-1-d} e^{-\lambda u}\,\mathrm{d}u.
\end{aligned}$$

It follows immediately from Theorem 3.5 that if $\Theta \leftarrow \text{PL-Rep}(\gamma, \lambda, d, \mathsf{Gam}(\lambda, \lambda))$, then $\Theta \sim \Gamma\text{P}(\gamma, \lambda, d)$. To the best knowledge of the authors, this power-law representation for the gamma process is novel.

4. Truncation analysis

Each of the sequential representations developed in Section 3 shares a common structural element—an outer infinite sum—which is responsible for generating a countably infinite number of atoms in the CRM. In this section, we terminate these outer sums at a finite *truncation level* $K \in \mathbb{N}$, resulting in a *truncated CRM* Θ_K possessing a finite number of atoms. We develop upper bounds on the error induced by this truncation procedure. All of the truncated CRM error bounds in this section rely on Lemma 4.1, which is a tightening (by a factor of two) of the bound in [19,20] (for its generalization to arbitrary discrete random measures, see Lemma D.1).

Lemma 4.1 (CRM protobound). *Let* $\Theta \sim \mathrm{CRM}(\nu)$. *For any truncation* Θ_K, *if*

$$X_n \mid \Theta \overset{i.i.d.}{\sim} \mathrm{LP}(h, \Theta), \qquad Z_n \mid \Theta_K \overset{i.i.d.}{\sim} \mathrm{LP}(h, \Theta_K),$$

$$Y_n \mid X_n \overset{indep}{\sim} f(\cdot \mid X_n), \qquad W_n \mid Z_n \overset{indep}{\sim} f(\cdot \mid Z_n),$$

then, with $p_{N,\infty}$ *and* $p_{N,K}$ *denoting the marginal densities of* $Y_{1:N}$ *and* $W_{1:N}$, *respectively,*

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \mathbb{P}\big(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)\big).$$

The proof of all results in this section (including Lemma 4.1) can be found in Appendix D. All of the provided truncation results use the generative model in Lemma 4.1, and are summarized in Table 1 in Section 7. Throughout this section, for a given likelihood model $h(x \mid \theta)$ we define $\pi(\theta) := h(0 \mid \theta)$ for notational brevity. The asymptotic behavior of truncation error bounds is specified with tilde notation:

$$a(K) \sim b(K), \qquad K \to \infty \quad \Longleftrightarrow \quad \lim_{K\to\infty} \frac{a(K)}{b(K)} = 1.$$

4.1. Series representations

Each of the series representations can be viewed a functional of a standard Poisson point process and a sequence of i.i.d. random variables with some distribution g on $\mathbb{R}_+$. In particular, we may write each in the form

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \text{with } \theta_k = \tau(V_k, \Gamma_k),\ V_k \overset{\text{i.i.d.}}{\sim} g, \tag{4.1}$$

where Γ_k are the jumps of a unit-rate homogeneous Poisson point process on $\mathbb{R}_+$, and $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a non-negative measurable function such that $\lim_{u\to\infty} \tau(v, u) = 0$ for g-almost every v. The truncated CRM then takes the form

$$\Theta_K := \sum_{k=1}^{K} \theta_k \delta_{\psi_k}.$$

Theorem 4.2 provides a general truncation error bound for series representations of the form Eq. (4.1), specifies its range, and guarantees that the bound decays to 0 as $K \to \infty$.

Theorem 4.2 (Series representation truncation error). *The error in approximating a series representation of Θ with its truncation Θ_K satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - e^{-B_{N,K}} \leq 1,$$

where

$$B_{N,K} := \int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(\tau(V, u + G_K)\right)^N\right]\right) \mathrm{d}u, \tag{4.2}$$

$G_0 := 0$, $G_K \overset{indep}{\sim} \mathsf{Gam}(K, 1)$ *for* $K \geq 1$, *and* $V \overset{indep}{\sim} g$. *Furthermore,*

$$\forall N \in \mathbb{N}, \qquad \lim_{K\to\infty} B_{N,K} = 0.$$

Remark. An alternate form of $B_{N,K}$ that is sometimes easier to use in practice can be found by applying the standard geometric series formula to Eq. (4.2), which yields

$$B_{N,K} = \sum_{n=1}^{N} \int_0^\infty \mathbb{E}\left[\pi\left(\tau(V, u + G_K)\right)^{n-1}\left(1 - \pi\left(\tau(V, u + G_K)\right)\right)\right] \mathrm{d}u.$$

A simplified upper bound on $B_{N,K}$ can be derived by noting that $\pi(\theta) \leq 1$, so

$$B_{N,K} \leq N \int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(\tau(V, u + G_K)\right)\right]\right) \mathrm{d}u. \tag{4.3}$$

This bound usually gives the same asymptotics in K as Eq. (4.2).

The main task in using Theorem 4.2 to develop a truncation error bound for a series representation is evaluating the integrand in the definition of $B_{N,K}$. Thus, we next evaluate the integrand and provide expressions of the truncation error bound for the four series representations outlined in Section 3.1. Throughout the remainder of this section, G_K is defined as in Theorem 4.2, $F_0 \equiv 1$, and F_K is the CDF of G_K.

Inverse-Lévy representation

For this representation, we have

$$\tau(v, u) = \nu^{\leftarrow}(u) := \inf\left\{y : \nu\left([y, \infty)\right) \leq u\right\}.$$

To evaluate the bound in Eq. (4.3), we use the transformation of variables $x = \nu^{\leftarrow}(u + G_K)$ and the fact that for $a, b \geq 0$, $\nu^{\leftarrow}(a) \geq b \iff a \leq \nu([b, \infty))$ to conclude that

$$B_{N,K} \leq N \int_0^\infty F_K\left(\nu[x, \infty)\right)\left(1 - \pi(x)\right)\nu(\mathrm{d}x). \tag{4.4}$$

Recent work on the inverse-Lévy representation has developed Monte Carlo estimates of the error of the truncated random measure moments for those $\nu([x,\infty))$ with known inverse $\nu^{\leftarrow}$ [3]. In contrast, the result above provides an explicit bound on the L^1 truncation error. Our bound does not require knowing $\nu^{\leftarrow}$, which is often the most challenging aspect of applying the inverse-Lévy representation.

Example 4.1 (IL-Rep **truncation for** $\mathrm{LomP}(\gamma,\lambda^{-1})$ **with Poisson likelihood).** Recall from Example 3.4 that the Lomax process $\mathrm{LomP}(\gamma,\lambda^{-1})$ is the CRM with rate measure $\nu(\mathrm{d}\theta) = \gamma\lambda\theta^{-1}(1+\lambda\theta)^{-1}\,\mathrm{d}\theta$, so $\nu[x,\infty) = \gamma\lambda\log\{1+(\lambda x)^{-1}\}$. Using Eq. (4.4), we have

$$B_{N,K} \le N\gamma\lambda\int_0^\infty F_K\big(\gamma\lambda\log\big\{1+(\lambda x)^{-1}\big\}\big)\big(1-e^{-x}\big)x^{-1}(1+\lambda x)^{-1}\,\mathrm{d}x.$$

Since $F_K(t) \le t^K/K! \le (3t/K)^K$, for any $a>0$ the integral is upper bounded by

$$\begin{aligned}
&\int_0^a (1+\lambda x)^{-1}\,\mathrm{d}x + F_K\left(\gamma\lambda\log\left\{1+\frac{1}{\lambda a}\right\}\right)\int_a^\infty x^{-1}(1+\lambda x)^{-1}\,\mathrm{d}x \\
&\quad\le a + F_K\left(\gamma\lambda\log\left\{1+\frac{1}{\lambda a}\right\}\right)\log\left\{1+\frac{1}{\lambda a}\right\} \qquad (4.5)\\
&\quad\le \lambda^{-1}\big(e^b-1\big)^{-1} + b(3\gamma\lambda b/K)^K \qquad \text{where } b := \log\big\{1+(\lambda a)^{-1}\big\}.
\end{aligned}$$

Replacing $(e^b-1)^{-1}$ with the approximation e^{-b} and then setting the two terms in Eq. (4.5) equal, we obtain $b = K\,W_0(\{3\gamma\lambda^{\frac{K+2}{K+1}}(K+1)^{\frac{1}{K+1}}\}^{-1})$, where W_0 is the product logarithm function, that is,

$$W_0(y) = x \quad\Longleftrightarrow\quad xe^x = y. \qquad (4.6)$$

Thus, using the fact that $e^{-t} \le (e^t-1)^{-1}$ and $\lambda^{\frac{K+2}{K+1}}(K+1)^{\frac{1}{K+1}}$ reaches its maximum at $K = \max(0, e\lambda^{-1}-1)$, we conclude that

$$\begin{aligned}
B_{N,K} &\le \frac{2N\gamma[1+(3\gamma\lambda)^{-1}]}{\exp(K\,W_0(\{3\gamma\lambda\max(\lambda,e)\}^{-1}))-1}\\
&\sim 2N\gamma\big[1+(3\gamma\lambda)^{-1}\big]e^{-K\,W_0(\{3\gamma\lambda\max(\lambda,e)\}^{-1})}, \qquad K\to\infty.
\end{aligned}$$

Bondesson representation

For this representation, we have

$$\tau(v,u) = ve^{-u/c}, \qquad g(\mathrm{d}v) = -c^{-1}\frac{\mathrm{d}}{\mathrm{d}v}\big(v\nu(v)\big)\,\mathrm{d}v.$$

Writing the expectation over V explicitly as an integral with measure $g(v)\,\mathrm{d}v$, using the transformation of variables $u = -c\log x/v$ (so $x = ve^{-u/c}$), and given the definition of $g(v) =$

$-c^{-1}\frac{\mathrm{d}}{\mathrm{d}v}(v\nu(v))$ for the Bondesson representation, we have

$$B_{N,K} \leq N \int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(ve^{-G_K/c}\right)\right]\right)\nu(\mathrm{d}v).$$

Example 4.2 (Truncation of the Bondesson representation for $\Gamma\mathrm{P}(\gamma,\lambda,0)$). Let $\tilde{G}_K \overset{\mathcal{D}}{=} G_K/(\gamma\lambda)$. Since $\pi(\theta) = e^{-\theta}$ and $c = \gamma\lambda$, we have

$$\begin{aligned}\int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(ve^{-G_K/c}\right)\right]\right)\nu(\mathrm{d}v) &= \gamma\lambda\mathbb{E}\left[\int_0^\infty \left(1 - e^{-ve^{-\tilde{G}_K}}\right)v^{-1}e^{-\lambda v}\,\mathrm{d}v\right] \\ &= \gamma\lambda\mathbb{E}\left[\log\left(1 + e^{-\tilde{G}_K}/\lambda\right)\right] \\ &\leq \gamma\mathbb{E}\left[e^{-\tilde{G}_K}\right] = \gamma\left(\frac{\gamma\lambda}{1+\gamma\lambda}\right)^K.\end{aligned}$$

The second equality follows by using the power series for the exponential integral [1], Chapter 5. Thus,

$$B_{N,K} \leq N\gamma\left(\frac{\gamma\lambda}{1+\gamma\lambda}\right)^K.$$

Thinning representation

For this representation, we have

$$\tau(v,u) = v\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) \geq u\right], \qquad g \text{ any distribution on } \mathbb{R}_+ \text{ s.t. } \nu \ll g.$$

Since $\pi(0) = 1$ by Lemma B.3, we have that $1 - \pi(v\mathbb{1}(A)) = (1 - \pi(v))\mathbb{1}(A)$ for any event A. Using this fact, we have

$$B_{N,K} \leq N \int_0^\infty \left(1 - \pi(v)\right) \int_0^{\frac{\mathrm{d}\nu}{\mathrm{d}g}(v)} F_K(u)\,\mathrm{d}u g(\mathrm{d}v). \tag{4.7}$$

Analytic bounds for the thinning representation of specific processes tend to be opaque and notationally cumbersome, so we simply compare its truncation error in Section 7 to the other representations by numerical approximation of Eq. (4.7).

Rejection representation

Assume that we can use the inverse-Lévy representation to simulate $\mathrm{PoissP}(\mu)$. Then for the rejection representation we have

$$\tau(v,u) = \mu^{\leftarrow}(u)\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\left(\mu^{\leftarrow}(u)\right) \geq v\right], \qquad g(\mathrm{d}v) = \mathbb{1}[0 \leq v \leq 1]\,\mathrm{d}v,$$

where μ satisfies $\nu \ll \mu$, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \leq 1$, and $\mu^{\leftarrow}(u) := \inf\{x : \mu([x,\infty)) \leq u\}$. Using the same techniques as for the thinning and inverse-Lévy representations, we have that

$$B_{N,K} \leq N \int_0^\infty F_K\big(\mu[x,\infty)\big)\big(1-\pi(x)\big)\nu(\mathrm{d}x). \tag{4.8}$$

Example 4.3 (R-Rep **truncation for** $\Gamma\mathrm{P}(\gamma,\lambda,0)$ **with Poisson likelihood).** Using Eq. (4.8) and the fact that $1-e^{-x} \leq x$, we have

$$B_{N,K} \leq N\gamma\lambda \int_0^\infty F_K\big(\gamma\lambda \log\big\{1+(\lambda x)^{-1}\big\}\big)e^{-\lambda x}\,\mathrm{d}x. \tag{4.9}$$

Arguing as in Example 4.1, we see that the integral in Eq. (4.9) is upper bounded by

$$\begin{aligned}
&\int_0^a e^{-\lambda x}\,\mathrm{d}x + F_K\left(\gamma\lambda\log\left\{1+\frac{1}{\lambda a}\right\}\right)\int_a^\infty e^{-\lambda x}\,\mathrm{d}x \\
&\quad\leq a + \lambda^{-1}F_K\left(\gamma\lambda\log\left\{1+\frac{1}{\lambda a}\right\}\right) \\
&\quad= \lambda^{-1}\big(\big(e^b-1\big)^{-1} + (3\gamma\lambda b/K)^K\big),
\end{aligned} \tag{4.10}$$

where $b := \log\{1+(\lambda a)^{-1}\}$. Replacing $(e^b-1)^{-1}$ with the approximation e^{-b} and then setting the two terms in Eq. (4.10) equal to each other, we obtain $b = K\,W_0(\{3\gamma\lambda\}^{-1})$ (where W_0 is defined in Eq. (4.6)) and conclude that

$$B_{N,K} \leq \frac{2N\gamma}{e^{K\,W_0(\{3\gamma\lambda\}^{-1})}-1} \sim 2N\gamma e^{-K\,W_0(\{3\gamma\lambda\}^{-1})}, \qquad K\to\infty.$$

Example 4.4 (R-Rep **truncation for** $\Gamma\mathrm{P}(\gamma,\lambda,d)$ **with Poisson likelihood,** $d>0$**).** We have

$$B_{N,K} \leq N\frac{\gamma\lambda^{1-d}}{\Gamma(1-d)}\int_0^\infty F_K\big(\gamma' x^{-d}\big)\big(1-e^{-x}\big)x^{-1-d}e^{-\lambda x}\,\mathrm{d}x.$$

The integral can be upper bounded as

$$\begin{aligned}
&\int_0^a x^{-d}\,\mathrm{d}x + F_K\big(\gamma' a^{-d}\big)\int_a^\infty \big(1+e^{-x}\big)x^{-1-d}e^{-\lambda x}\,\mathrm{d}x \\
&\quad\leq (1-d)^{-1}a^{1-d} + \Gamma(-d)\big(\lambda^d-(1+\lambda)^d\big)\big(3\gamma' K^{-1}a^{-d}\big)^K.
\end{aligned}$$

Setting the two terms equal and solving for a, we obtain

$$\begin{aligned}
B_{N,K} &\leq 2N\frac{\gamma\lambda^{1-d}}{\Gamma(2-d)}\big[(1-d)\Gamma(-d)\big]^{\frac{d(1-d)}{d(1-d)+K}}\left[\frac{3\gamma\lambda^{1-d}}{d\Gamma(1-d)K}\right]^{\frac{Kd(1-d)}{d(1-d)+K}} \\
&\sim 2N\frac{\gamma\lambda^{1-d}}{\Gamma(2-d)}\left[\frac{3\gamma\lambda^{1-d}}{d\Gamma(1-d)}\right]^{d(1-d)}K^{-d(1-d)}, \qquad K\to\infty.
\end{aligned}$$

4.2. Superposition representations

For superposition representations, the truncated CRM takes the form

$$\Theta_K := \sum_{k=1}^{K} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}.$$

Let $\Theta_K^+ := \Theta - \Theta_K$ denote the *tail measure*. By the superposition property of Poisson point processes [28], the tail measure is itself a CRM with some rate measure ν_K^+ and is independent of Θ_K:

$$\Theta_K^+ = \sum_{k=K+1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}} \sim \mathrm{CRM}\big(\nu_K^+\big), \qquad \Theta_K^+ \perp\!\!\!\perp \Theta_K, \qquad \Theta = \Theta_K + \Theta_K^+. \tag{4.11}$$

The following result provides a general truncation error bound for superposition representations, specifies its range, and guarantees that the bound decays to 0 as $K \to \infty$.

Theorem 4.3 (Superposition representation truncation error). *The error in approximating a superposition representation of* $\Theta \sim \mathrm{CRM}(\nu)$ *with its truncation* Θ_K *satisfies*

$$0 \leq \frac{1}{2} \| p_{N,\infty} - p_{N,K} \|_1 \leq 1 - e^{-B_{N,K}} \leq 1,$$

where

$$B_{N,K} := \int \big(1 - \pi(\theta)^N\big) \nu_K^+(\mathrm{d}\theta). \tag{4.12}$$

Furthermore, $\forall N \in \mathbb{N}, \lim_{K \to \infty} B_{N,K} = 0$.

Remark. As for series representations, an alternate form of $B_{N,K}$ that is sometimes easier to use can be found by applying the standard geometric series formula to Eq. (4.12):

$$B_{N,K} = \sum_{n=1}^{N} \int \pi(\theta)^{n-1} \big(1 - \pi(\theta)\big) \nu_K^+(\mathrm{d}\theta).$$

A simplified upper bound on $B_{N,K}$ can be derived by noting that $\pi(\theta) \leq 1$, so

$$B_{N,K} \leq N \int_0^{\infty} \big(1 - \pi(\theta)\big) \nu_K^+(\mathrm{d}\theta).$$

This bound usually gives the same asymptotics in K as Eq. (4.12).

The main task in using Theorem 4.3 to develop a truncation error bound for a superposition representation is determining its tail measure ν_K^+. In the following, we provide the tail measure for the three superposition representations outlined in Section 3.2.

Decoupled Bondesson representation

For each point process in the superposition, an average of c/ξ atoms are generated with independent weights of the form Ve^{-T_k} where $V \overset{\text{indep}}{\sim} g$ and $T_k \overset{\text{indep}}{\sim} \mathsf{Gam}(k,\xi)$. Therefore, the tail measure is

$$\nu_K^+(\mathrm{d}\theta) = \frac{c}{\xi} \sum_{k=K+1}^{\infty} \tilde{g}_{k,\xi}(\theta)\,\mathrm{d}\theta,$$

where $\tilde{g}_{k,\xi}$ is the density of Ve^{-T_k}. The bound for the decoupled Bondesson representation can therefore be expressed as

$$B_{N,K} \leq N\frac{c}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}\big[1-\pi\big(Ve^{-T_k}\big)\big].$$

Example 4.5 (Decoupled Bondesson representation truncation for $\Gamma\mathrm{P}(\gamma,\lambda,0)$). Using the fact that $1-e^{-\theta} \leq \theta$, we have

$$\begin{aligned} B_{N,K} &= \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}\big[1-\pi\big(V_{k1}e^{-T_{k1}}\big)\big] \leq \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}\big[V_{k1}e^{-T_{k1}}\big] \\ &= \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \frac{1}{\lambda}\left(\frac{\xi}{1+\xi}\right)^k = N\gamma\left(\frac{\xi}{1+\xi}\right)^K, \end{aligned}$$

which is equivalent (up to a factor of 2) to the bound in [46].

Size-biased representation

The constructive derivation of the size-biased representation [11], proof of Theorem 5.1, immediately yields

$$\nu_K^+(\mathrm{d}\theta) = \pi(\theta)^K \nu(\mathrm{d}\theta).$$

Therefore, the size-biased representation truncation error bound can be expressed using the formula for η_k from Eq. (3.2) as

$$B_{N,K} = \sum_{n=1}^{N} \int \pi(\theta)^{K+n-1}\big(1-\pi(\theta)\big)\nu(\mathrm{d}\theta) = \sum_{n=1}^{N} \eta_{K+n}. \tag{4.13}$$

Example 4.6 (Size-biased representation truncation for $\Gamma\mathrm{P}(\gamma,\lambda,d)$). For $d>0$, the standard gamma integral yields

$$\eta_k = \int \pi(\theta)^{k-1}\big(1-\pi(\theta)\big)\nu(\mathrm{d}\theta) = \frac{\gamma\lambda^{1-d}}{d}\big((\lambda+k)^d - (\lambda+k-1)^d\big).$$

The sum from Eq. (4.13) is telescoping, so canceling terms,

$$B_{N,K} \leq \frac{\gamma\lambda^{1-d}}{d}\big((\lambda+K+N)^d-(\lambda+K)^d\big) \sim \gamma N\lambda^{1-d}K^{d-1}, \qquad K\to\infty,$$

where the asymptotic result follows from Lemma B.4. To analyze the $d=0$ case, we use L'Hospital's rule to take the limit of the integral:

$$\lim_{d\to 0}\int \pi(\theta)^{k-1}\big(1-\pi(\theta)\big)\nu(\mathrm{d}\theta) = \gamma\lambda\big(\log(\lambda+k)-\log(\lambda+k-1)\big).$$

Canceling terms in the telescopic sum yields

$$B_{N,K} \leq \gamma\lambda\big(\log(\lambda+K+N)-\log(\lambda+K)\big) \sim \gamma\lambda N K^{-1}, \qquad K\to\infty,$$

where the asymptotic result follows from an application of Lemma B.4.

Power-law representation

For each point process in the superposition, an average of γ atoms are generated with independent weights of the form $VU_k\prod_{\ell=1}^{k-1}(1-U_\ell)$, where $V \overset{\text{indep}}{\sim} g$ and $U_\ell \overset{\text{indep}}{\sim} \mathsf{Beta}(1-d,\alpha+\ell d)$. Therefore, the tail measure is

$$\nu_K^+(\mathrm{d}\theta) = \gamma\sum_{k=K+1}^{\infty}\tilde{g}_k(\theta)\,\mathrm{d}\theta,$$

where $\tilde{g}_k$ is the density of the random variable $VU_k\prod_{\ell=1}^{k-1}(1-U_\ell)$. The truncation error bound may be expressed as

$$B_{N,K} \leq N\gamma\sum_{k=K+1}^{\infty}\mathbb{E}\left[1-\pi\left(VU_k\prod_{\ell=1}^{k-1}(1-U_\ell)\right)\right].$$

Example 4.7 (Power-law representation truncation for $\Gamma\mathrm{P}(\gamma,\lambda,d)$). Let β_k be a random variable with density $\tilde{g}_k$ (with λ in the place of α). Using $1-e^{-\theta}\leq\theta$, we have

$$\sum_{k=K+1}^{\infty}\mathbb{E}\big[1-\pi(\beta_k)\big] \leq \sum_{k=K+1}^{\infty}\mathbb{E}[\beta_k] = \mathbb{E}\left[\sum_{k=K+1}^{\infty}\beta_k\right] = \prod_{k=1}^{K}\frac{\lambda+kd}{\lambda+kd-d+1},$$

where the final equality follows from [19], Theorem 1. Thus,

$$B_{N,K} \leq \gamma N \prod_{k=1}^{K} \frac{\lambda + kd}{\lambda + kd - d + 1} \sim \gamma N \begin{cases} \left(\frac{\lambda}{\lambda+1}\right)^K, & d = 0, \\ \frac{\Gamma(\frac{\lambda+1}{d})}{\Gamma(\frac{\lambda+d}{d})} K^{1-d^{-1}}, & 0 < d < 1, \end{cases} \quad K \to \infty, \tag{4.14}$$

where the $0 < d < 1$ case in Eq. (4.14) follows by Lemma B.5 applied to

$$\prod_{k=1}^{K} \frac{\lambda + kd}{\lambda + kd - d + 1} = \frac{\Gamma((\lambda+1)/d)}{\Gamma((\lambda+d)/d)} \frac{\Gamma(\lambda/d + K + 1)}{\Gamma(\lambda/d + K + d^{-1})}.$$

4.3. Stochastic mapping

We now show how truncation bounds developed elsewhere in this paper can be applied to CRM representations that have been transformed using Lemma 3.4. For $\Theta \sim \mathrm{CRM}(\nu)$, we denote its transformation by $\tilde{\Theta} = \kappa(\Theta)$. For any object defined with respect to Θ, the corresponding object for $\tilde{\Theta}$ is denoted with a tilde. For example, in place of N and $X_{1:N}$ (for Θ), we use $\tilde{N}$ and $\tilde{X}_{1:\tilde{N}}$ (for $\tilde{\Theta}$). We make $B_{N,K}$ a function of $\pi(\theta)$ in the notation of Proposition 4.4; when one applies stochastic mapping to a CRM, one usually also wants to change the likelihood $h(x \mid \theta)$, and thus also changes $\pi(\theta) = h(0 \mid \theta)$. The proof of Proposition 4.4 may be found in Appendix D.

Proposition 4.4 (Truncation error under a stochastic mapping). *Consider a representation for $\Theta \sim \mathrm{CRM}(\nu)$ with truncation error bound $B_{N,K}(\pi)$. Then for any likelihood $\tilde{h}(x \mid u)$, if $\tilde{\Theta}$ is a stochastic mapping of Θ under the probability kernel $\kappa(\theta, \mathrm{d}u)$, its truncation error bound is $B_{1,K}(\pi_{\kappa,\tilde{N}})$, where $\pi_{\kappa,\tilde{N}}(\theta) := \int \tilde{h}(0 \mid u)^{\tilde{N}} \kappa(\theta, \mathrm{d}u)$.*

4.4. Hyperpriors

In practice, prior distributions are often placed on the hyperparameters of the CRM rate measure (i.e., γ, α, λ, d, etc.). We conclude our investigation of CRM truncation error by showing how bounds developed in this section can be modified to account for the use of hyperpriors. Note that we make the dependence of $B_{N,K}$ on the hyperparameters Φ explicit in the notation of Proposition 4.5.

Proposition 4.5 (CRM truncation error with a hyperprior). *Given hyperparameters Φ, consider a representation for $\Theta \mid \Phi \sim \mathrm{CRM}(\nu)$, and let $B_{N,K}(\Phi)$ be given by Eq. (4.2) (for a series*

representation) *or Eq.* (4.12) (*for a superposition representation*). *The error of approximating* Θ *with its truncation* Θ_K *satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - e^{-\mathbb{E}[B_{N,K}(\Phi)]} \leq 1.$$

Example 4.8 (Decoupled Bondesson representation truncation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$). A standard choice of hyperprior for the mass γ is a gamma distribution, i.e. $\gamma \sim \mathsf{Gam}(a, b)$. Combining Proposition 4.5 and Example 4.5, we have that

$$\mathbb{E}\big[B_{N,K}(\Phi)\big] \leq N\frac{a}{b}\left(\frac{\xi}{\xi+1}\right)^K.$$

5. Normalized truncation analysis

In this section, we provide truncation error bounds for *normalized CRMs* (NCRMs). Examples include the Dirichlet process [14], the normalized gamma process [8,22,31–33,41], and the normalized σ-stable process [29,33]. Given a CRM Θ on Ψ, we define the corresponding NCRM Ξ via $\Xi(S) := \Theta(S)/\Theta(\Psi)$ for each measurable subset $S \subseteq \Psi$. Likewise, given a truncated CRM Θ_K, we define its normalization Ξ_K via $\Xi_K(S) := \Theta_K(S)/\Theta_K(\Psi)$. Note that any simulation algorithm for Θ_K can be used for Ξ_K by simply normalizing the result. This does not depend on the particular representation of the CRM, and thus applies equally to all the representations in Section 3.

The first step in the analysis of NCRM truncations is to define their approximation error in a manner similar to that of CRM truncations. Since Ξ and Ξ_K are both normalized, they are distributions on Ψ; thus, observations $X_{1:N}$ are generated i.i.d. from Ξ, and $Z_{1:N}$ are generated i.i.d. from Ξ_K. $Y_{1:N}$ and $W_{1:N}$ have the same definition as for CRMs. As in the developments of Section 4, the theoretical results of this section rely on a general upper bound, provided by Lemma 5.1.

Lemma 5.1 (NCRM protobound). *Let* $\Theta \sim \mathrm{CRM}(\nu)$, *and let its truncation be* Θ_K. *Let their normalizations be* Ξ *and* Ξ_K *respectively. If*

$$X_n \mid \Xi \overset{i.i.d.}{\sim} \Xi, \qquad Z_n \mid \Xi_K \overset{i.i.d.}{\sim} \Xi_K,$$

$$Y_n \mid X_n \overset{indep}{\sim} f(\cdot \mid X_n), \qquad W_n \mid Z_n \overset{indep}{\sim} f(\cdot \mid Z_n),$$

then

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \mathbb{P}\big(X_{1:N} \subseteq \mathrm{supp}(\Xi_K)\big),$$

where $p_{N,\infty}$, $p_{N,K}$ *are the marginal densities of* $Y_{1:N}$ *and* $W_{1:N}$, *respectively.*

The analysis of CRMs in Section 4 relied heavily on the Poisson process structure of the rates in Θ and $X_{1:N}$; unfortunately, the rates in Ξ do not possess the same structure and thus lack many

useful independence properties (the rates must sum to one). Likewise, sampling X_n for each n does not depend on the atoms of Ξ independently (X_n randomly selects a single atom based on their rates). Rather than using the basic definitions of the above random quantities to derive an error bound, we decouple the atoms of Ξ and $X_{1:N}$ using a technique from extreme value theory. A *Gumbel* random variable T with location $\mu \in \mathbb{R}$ and scale $\sigma > 0$, denoted $T \sim \mathsf{Gumbel}(\mu, \sigma)$, is defined by the cumulative distribution function and corresponding density

$$\mathbb{P}(T \leq t) = e^{-e^{-\frac{t-\mu}{\sigma}}} \quad \text{and} \quad \frac{1}{\sigma} e^{-(\frac{t-\mu}{\sigma})-e^{-(\frac{t-\mu}{\sigma})}}.$$

An interesting property of the Gumbel distribution is that if one perturbs the log-probabilities of a finite discrete distribution by i.i.d. $\mathsf{Gumbel}(0, 1)$ random variables, the arg max of the resulting set is a sample from the discrete distribution [17,34]. This technique is invariant to normalization, as the arg max is invariant to the corresponding constant shift in the log-transformed space. For present purposes, we develop the infinite extension of this result.

Lemma 5.2 (Infinite Gumbel-max sampling). *Let $(p_i)_{i=1}^{\infty}$ be a collection of positive numbers such that $\sum_i p_i < \infty$ and let $\bar{p}_j := \frac{p_j}{\sum_i p_i}$. If $(T_i)_{i=1}^{\infty}$ are i.i.d. $\mathsf{Gumbel}(0, 1)$ random variables, then $\arg\max_{i \in \mathbb{N}} T_i + \log p_i$ exists, is unique a.s., and has distribution*

$$\arg\max_{i \in \mathbb{N}} T_i + \log p_i \sim \mathsf{Categorical}\big((\bar{p}_j)_{j=1}^{\infty}\big).$$

The proof of this result, along with the others in this section, may be found in Appendix E. The utility of Lemma 5.2 is that it allows the construction of Ξ and $X_{1:N}$ without the problematic coupling of the underlying CRM atoms due to normalization; rather than dealing directly with Ξ, we log-transform the rates of Θ, perturb them by i.i.d. $\mathsf{Gumbel}(0, 1)$ random variables, and characterize the distribution of the maximum rate in this process. The combination of this distribution with Lemma 5.2 yields the key proof technique used to develop the truncation bounds in Theorems 5.3 and 5.4. The results presented in this section are summarized in Table 1 in Section 7.

5.1. Series representations

The following result provides a general truncation error bound for normalized series representations, specifies its range, and guarantees that it decays to 0 as $K \to \infty$. We again use the general series representation notation from Eq. (4.1), where g is a distribution on $\mathbb{R}_+$, and $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function such that $\lim_{u \to \infty} \tau(v, u) = 0$ for g-almost every v.

Theorem 5.3 (Normalized series representation truncation error bound). *The error of approximating a series representation of $\Xi \sim \mathrm{NCRM}(\nu)$ with its truncation Ξ_K satisfies*

$$0 \leq \frac{1}{2} \| p_{N,\infty} - p_{N,K} \|_1 \leq 1 - (1 - B_K)^N \leq 1,$$

where

$$B_K := \mathbb{E}\left[\int_0^\infty J(\Gamma_K, t)\left(\int_0^1 J(\Gamma_K u, t)\,\mathrm{d}u\right)^{K-1}\left(-\frac{\mathrm{d}}{\mathrm{d}t}e^{\int_0^\infty (J(u+\Gamma_K,t)-1)\,\mathrm{d}u}\right)\mathrm{d}t\right], \quad (5.1)$$

$$J(u,t) = \mathbb{E}\left[e^{-t\cdot\tau(V,u)}\right], \qquad V \sim g, \quad \textit{and} \quad \Gamma_K \sim \mathsf{Gam}(K,1).$$

Furthermore, $\lim_{K\to\infty} B_K = 0$.

Example 5.1 (Dirichlet process, DP(γ)**,** B-Rep**).** The Dirichlet process with concentration $\gamma > 0$ is a normalized gamma process NΓP($\gamma, 1, 0$). From Example 3.2, we have $c_v = \gamma$ and $g_v = \mathsf{Exp}(1)$, and from Section 4.1 we have $\tau(v,u) = ve^{-u/c_v}$. Therefore J and its antiderivative are

$$J(u,t) = \mathbb{E}\left[e^{-tVe^{-u/\gamma}}\right] = \left(1 + te^{-u/\gamma}\right)^{-1} \quad \text{and} \quad \int J(u,t)\,\mathrm{d}u = \gamma\log\left(e^{u/\gamma}+t\right).$$

Using the antiderivative to evaluate the integrals in the formula for B_K, writing the expectation over $\Gamma_K \sim \mathsf{Gam}(K,1)$ explicitly, and making a change of variables we have

$$B_K = \frac{\gamma^{K+1}}{\Gamma(K)}\int_0^\infty\int_1^\infty\left(\log\left(\frac{s+t}{1+t}\right)\right)^{K-1}(s+t)^{-(\gamma+2)}\,\mathrm{d}s\,\mathrm{d}t = \left(\frac{\gamma}{1+\gamma}\right)^K,$$

where the last equality is found by multiplying and dividing the integrand by $(1+t)^{-(\gamma+2)}$, and making the change of variables from s to $x = \log\frac{s+t}{1+t}$. Therefore, the truncation error can be bounded by

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \le 1 - \left(1 - \left(\frac{\gamma}{\gamma+1}\right)^K\right)^N \sim N\left(\frac{\gamma}{\gamma+1}\right)^K, \qquad K\to\infty.$$

The bound in Example 5.2 has exponential decay, and reproduces earlier DP truncation error bound rates due to [19] and [21]. However, the techniques used in past work do not generalize beyond the Dirichlet process, while those developed here apply to any NCRM.

5.2. Superposition representations

The following result provides a general truncation error bound for normalized superposition representations, specifies its range, and guarantees that it decays to 0 as $K\to\infty$. We once again rely on the property that the truncation Θ_K and tail Θ_K^+ are mutually independent CRMs, as expressed in Eq. (4.11), with the tail measure denoted ν_K^+.

Theorem 5.4 (Truncation error bound for normalized superposition representations). *The error of approximating a superposition representation of* $\Xi \sim \mathrm{NCRM}(\nu)$ *with its truncation* Ξ_K *satisfies*

$$0 \le \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \le 1 - (1-B_K)^N \le 1,$$

where

$$B_K := \int_0^\infty \left(\int \theta e^{-\theta t} \nu_K^+(\mathrm{d}\theta) \right) e^{\int (e^{-\theta t}-1)\nu(\mathrm{d}\theta)} \, \mathrm{d}t. \tag{5.2}$$

Furthermore, $\lim_{K\to\infty} B_K = 0$.

This bound can be applied by using the tail measures derived earlier in Section 4.2.

Example 5.2 (Dirichlet process, DP(γ), DB-Rep). As in Example 5.1, we view the Dirichlet process with concentration $\gamma > 0$ as a normalized gamma process NΓP$(\gamma, 1, 0)$. First, by Lemma B.8, the integral in the exponential is

$$\exp\left(\int \left(e^{-t\theta} - 1\right) \nu(\mathrm{d}\theta) \right) = \exp\left(\gamma \int_0^\infty \left(e^{-t\theta} - 1\right) \theta^{-1} e^{-\theta} \, \mathrm{d}\theta \right) = (t+1)^{-\gamma}.$$

Example 3.2 shows $c_\nu = \gamma$ and $g_\nu(v) = e^{-v}$, and Eq. (C.1) provides the tail measure ν_K^+ for the decoupled Bondesson representation,

$$\nu_K^+(\mathrm{d}\theta) = \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \left(\int_0^1 (-\log x)^{k-1} x^{\xi-2} e^{-\theta x^{-1}} \, \mathrm{d}x \right) \mathrm{d}\theta.$$

Substituting this result, using Fubini's theorem to swap the order of integration and summation, evaluating the integral over θ, and making the substitution $x = e^{-s}$ yields

$$B_K = \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \iint_{s,t\geq 0} \frac{s^{k-1} e^{-(\xi-1)s} (t+1)^{-\gamma}}{(e^s + t)^2} \, \mathrm{d}s \, \mathrm{d}t.$$

Noting that $\forall s \geq 0, e^s \geq 1$, we have for any $a \in (0,1] \cap (0,\gamma)$,

$$\begin{aligned} B_K &\leq \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \int_0^\infty s^{k-1} e^{-(\xi+a)s} \, \mathrm{d}s \int_0^\infty (t+1)^{-(\gamma+1-a)} \, \mathrm{d}t \\ &= \frac{\gamma}{(\gamma-a)\xi} \sum_{k=K+1}^\infty \left(\frac{\xi}{\xi+a} \right)^k = \frac{\gamma}{a(\gamma-a)} \left(\frac{\xi}{\xi+a} \right)^K. \end{aligned}$$

Therefore, for any $a \in (0,1] \cap (0,\gamma)$,

$$\begin{aligned} \frac{1}{2} \| p_{N,\infty} - p_{N,K} \|_1 &\leq 1 - \left(1 - \frac{\gamma}{a(\gamma-a)} \left(\frac{\xi}{\xi+a} \right)^K \right)^N \\ &\sim \frac{N\gamma}{a(\gamma-a)} \left(\frac{\xi}{\xi+a} \right)^K, \qquad K \to \infty. \end{aligned}$$

To find the tightest bound, one can minimize with respect to a given γ, ξ, K.

Example 5.3 (Normalized gamma process, $\mathrm{N\Gamma P}(\gamma, \lambda, d)$, SB-Rep). By Lemma B.8, the integral in the exponential is

$$\exp\left(\int \left(e^{-\theta t} - 1\right)\nu(\mathrm{d}\theta)\right) = \begin{cases} \exp\left(-\gamma\lambda^{1-d}d^{-1}\left((t+\lambda)^d - \lambda^d\right)\right), & d > 0, \\ \left(\frac{\lambda}{t+\lambda}\right)^{\gamma\lambda}, & d = 0, \end{cases} \tag{5.3}$$

and the standard gamma integral yields

$$\int \theta e^{-\theta t}\nu_K^+(\mathrm{d}\theta) = \gamma\frac{\lambda^{1-d}}{\Gamma(1-d)}\int \theta^{-d}e^{-(K+t+\lambda)\theta}\,\mathrm{d}\theta = \gamma\lambda^{1-d}(K+t+\lambda)^{d-1}. \tag{5.4}$$

When $d > 0$, multiplying the previous two displays and integrating over $t \geq 0$ yields

$$B_K = \gamma\lambda^{1-d}e^{\gamma\lambda/d}\int_\lambda^\infty (K+t)^{d-1}e^{-\gamma\lambda^{1-d}t^d/d}\,\mathrm{d}t \leq C_{\gamma,\lambda,d}(K+\lambda)^{d-1},$$

where we have used $(K+t)^{d-1} \leq (K+\lambda)^{d-1}$ for $t \geq \lambda$ and the change of variables $u = \gamma\lambda^{1-d}d^{-1}t^d$ to find that $C_{\gamma,\lambda,d} = e^\sigma\sigma^{1-d^{-1}}\lambda^{1-d}\Gamma(d^{-1},\sigma)$, where $\sigma = \gamma\lambda d^{-1}$ and $\Gamma(a,x) := \int_x^\infty \theta^{a-1}e^{-\theta}\,\mathrm{d}\theta$ is the upper incomplete gamma function. Therefore,

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \left(1 - C_{\gamma,\lambda,d}(K+\lambda)^{d-1}\right)^N \sim NC_{\gamma,\lambda,d}K^{d-1}, \qquad K \to \infty.$$

When $d = 0$, multiplying Eqs. (5.3) and (5.4) and integrating over $t \geq 0$ yields

$$\frac{B_K}{\gamma\lambda^{1+\gamma\lambda}} = \int_\lambda^\infty (K+t)^{-1}t^{-\gamma\lambda}\,\mathrm{d}t \leq \begin{cases} (K+\lambda)^{-1}\left(\frac{\frac{1}{\gamma\lambda}(K+\lambda)^{1-\gamma\lambda} - \lambda^{1-\gamma\lambda}}{1-\gamma\lambda}\right), & \gamma\lambda \neq 1, \\ K^{-1}\log\left(\frac{K+\lambda}{\lambda}\right), & \gamma\lambda = 1, \end{cases}$$

where we obtain the bound for $\gamma\lambda \neq 1$ by splitting the integral into the intervals $[\lambda, K+\lambda]$ and $[K+\lambda, \infty)$ and bounding each section separately, and we obtain the bound for $\gamma\lambda = 1$ via the transformation $u = t/(K+t)$. Therefore, asymptotically

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \lesssim N\begin{cases} C_{\gamma,\lambda}K^{-\min(1,\gamma\lambda)}, & \gamma\lambda \neq 1, \\ \lambda K^{-1}\log K, & \gamma\lambda = 1, \end{cases} \qquad K \to \infty,$$

where $C_{\gamma,\lambda} := \max(\frac{\lambda^{\gamma\lambda}}{1-\gamma\lambda}, \frac{\gamma\lambda^2}{\gamma\lambda-1})$.

Truncation of the $\mathrm{N\Gamma P}(\gamma, \lambda, d)$ has been studied previously: [4] threshold the weights of the unnormalized CRM to be beyond a fixed level $\epsilon > 0$ prior to normalization, and develop error bounds for that method of truncation. These results are not directly comparable to those of the present work due to the different methods of truncation (i.e., sequential representation termination versus weight thresholding).

5.3. Hyperpriors

As in the CRM case, we can place priors on the hyperparameters of the NCRM rate measure (i.e., γ, α, λ, d, etc.). We conclude our investigation of NCRM truncation error by showing how bounds developed in this section can be modified to account for hyperpriors. Note that we make the dependence of B_K on the hyperparameters Φ explicit in the notation of Proposition 5.5.

Proposition 5.5 (NCRM truncation error with a hyperprior). *Given hyperparameters* Φ, *consider a representation for* $\Theta \mid \Phi \sim \mathrm{CRM}(\nu)$, *let* $\Xi \mid \Phi$ *be its normalization, and let* $B_K(\Phi)$ *be given by Eq.* (5.1) (*for a series representation*) *or Eq.* (5.2) (*for a superposition representation*). *The error of approximating* Ξ *with its truncation* Ξ_K *satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \left(1 - \mathbb{E}\left[B_K(\Phi)\right]\right)^N \leq 1.$$

Example 5.4 (Dirichlet process, DP(γ), B-Rep). If we place a Lomax prior on γ, that is, $\gamma \sim \mathrm{LomP}(a, 1)$, then combining Proposition 5.5 and Example 5.1 yields

$$\begin{aligned}\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 &\leq 1 - \left(1 - \frac{\Gamma(a+1)\Gamma(K+1)}{\Gamma(a+K+1)}\right)^N \\ &\sim N\Gamma(a+1)(K+1)^{-a}, \qquad K \to \infty.\end{aligned}$$

6. Simulation and computational complexity

The sequential representations in Section 3 are each generated from a different finite sequence of distributions, resulting in a different expected computational cost for the same truncation level. Thus, the truncation level itself is not an appropriate parameter with which to compare the error bounds for different representations and we require a characterization of the computational cost. We investigate the mean complexity $\mathbb{E}[R]$ of each representation, where R is the number of random variables sampled, as a function of the truncation level for each of the representations in Section 3.

We begin with the series representations. For each value of $k = 1, \ldots, K$, each series representation generates a single trait $\psi_k \sim G$ and a rate θ_k composed of some transformation of random variables. Thus, all of the series representations in this work satisfy $\mathbb{E}[R] = rK$ for some constant r: by inspection, the inverse-Lévy representation has $r = 2$, and all the remaining series representations have $r = 3$.

The superposition representations, on the other hand, generate a Poisson random variable to determine the number of atoms at each value of $k = 1, \ldots, K$, and then generate those atoms. Therefore, the mean simulation complexity takes the form $\mathbb{E}[R] = \sum_{k=1}^{K} 1 + r_k \mathbb{E}[C_k]$ for some constants r_k that might depend on the value of k. For the decoupled Bondesson representation, $r_k = 3$ since each atom requires generating three values (ψ_{ki}, V_{ki}, and T_{ki}), and $\mathbb{E}[C_k] = c/\xi$, so $\mathbb{E}[R] = (\frac{3c}{\xi} + 1)K$. For the size-biased representation, $r_k = 3$ since each atom requires generating three values (ψ_{ki}, z_{ki}, and θ_{ki}), and $\mathbb{E}[C_k] = \eta_k$, so $\mathbb{E}[R] = K + 3\sum_{k=1}^{K} \eta_k$. Note that here

$\mathbb{E}[R] \sim K$, for $K \to \infty$ since η_k is a decreasing sequence. For the power-law representation, $r_k = k + 2$, since each atom requires generating ψ_{ki}, V_{ki}, and k beta random variables, and therefore $\mathbb{E}[R] = (1 + \frac{5\gamma}{2})K + \frac{\gamma}{2}K^2$.

7. Summary of results

Table 1 summarizes our truncation and simulation cost results as applied to the beta, (normalized) gamma, and beta prime processes. Results for the Bondesson representation of $\mathrm{BP}(\gamma, 1, 0)$ as well as the decoupled Bondesson representations of $\mathrm{BP}(\gamma, \lambda, 0)$ and $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ were previously

Table 1. Asymptotic error bounds and simulation cost summary. Error bounds are presented up to a constant that varies between models. Be = Bernoulli, OBe = odds Bernoulli, Poi = Poisson

Rep.	Random measure	h	Asymptotic error bound	Complexity
IL	$\mathrm{LomP}(\gamma, \lambda^{-1})$	Poi	$Ne^{-KW_0(\{3\gamma\lambda\max(\lambda,e)\}^{-1})}$	$2K$
B	$\mathrm{BP}(\gamma, \lambda \geq 1, 0)$ $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ $\mathrm{BPP}(\gamma, \lambda, 0)$	Be Poi OBe	$N\gamma(\frac{\gamma\lambda}{\gamma\lambda+1})^K$	$3K$
	$\mathrm{DP}(\gamma)$	–	$N(\frac{\gamma}{\gamma+1})^K$	
T	–	–	See Eq. (4.7)	$3K$
R	$\mathrm{BP}(\gamma, \lambda, 0)$ $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ $\mathrm{BPP}(\gamma, \lambda, 0)$	Be Poi OBe	$N\begin{cases} e^{-KW_0(\{3\gamma\lambda\}^{-1})}, & d = 0, \\ K^{-d(1-d)}, & d > 0(\Gamma\mathrm{P}), \\ K^{-1/d}, & d > 0(\mathrm{BP}, \mathrm{BPP}) \end{cases}$	$3K$
DB	$\mathrm{BP}(\gamma, \lambda \geq 1, 0)$ $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ $\mathrm{BPP}(\gamma, \lambda > 1, 0)$	Be Poi OBe	$N(\frac{\xi}{\xi+1})^K$	$(\frac{3c}{\xi}+1)K$
	$\mathrm{DP}(\gamma)$	–	$\frac{N\gamma}{a(\gamma-a)}(\frac{\xi}{\xi+a})^K, a \in (0,1] \cap (0,\gamma)$	
SB	$\mathrm{BP}(\gamma, \lambda, d)$ $\Gamma\mathrm{P}(\gamma, \lambda, d)$ $\mathrm{BPP}(\gamma, \lambda, d)$	Be Poi OBe	NK^{d-1}	K
	$\mathrm{N}\Gamma\mathrm{P}(\gamma, \lambda, d)$	–	$N\begin{cases} K^{-1}\log K, & d = 0, \gamma\lambda = 1, \\ K^{-\min(1,\gamma\lambda)}, & d = 0, \gamma\lambda \neq 1, \\ K^{d-1}, & d > 0 \end{cases}$	
PL	$\mathrm{BP}(\gamma, \lambda, d)$ $\Gamma\mathrm{P}(\gamma, \lambda, d)$ $\mathrm{BPP}(\gamma, \lambda > 1, d)$	Be Poi OBe	$N\begin{cases} (\frac{\lambda}{\lambda+1})^K, & d = 0(\mathrm{BP}, \Gamma\mathrm{P}), \\ 2^{-K}, & d = 0(\mathrm{BPP}), \\ K^{1-1/d}, & d > 0 \end{cases}$	$\frac{\gamma}{2}K^2$

known, and are reproduced by our results. All other results in the table are novel to the best of the authors' knowledge. It is interesting to note that the bounds and expected costs within each of the representation classes often have the same form, aside from some constants. Across classes, however, they vary significantly, indicating that the chosen sequential representation of a process has more of an influence on the truncation error than the process itself.

Figure 1 shows a comparison of how the truncation error bounds vary with the expected computational cost $\mathbb{E}[R]$ of simulation for the (normalized) gamma process and Poisson likelihood with $N = 5$ observations. Results shown for the thinning, rejection, and inverse-Lévy representations are computed by Monte-Carlo approximation of the formula for $B_{N,K}$ in Eq. (4.3), while

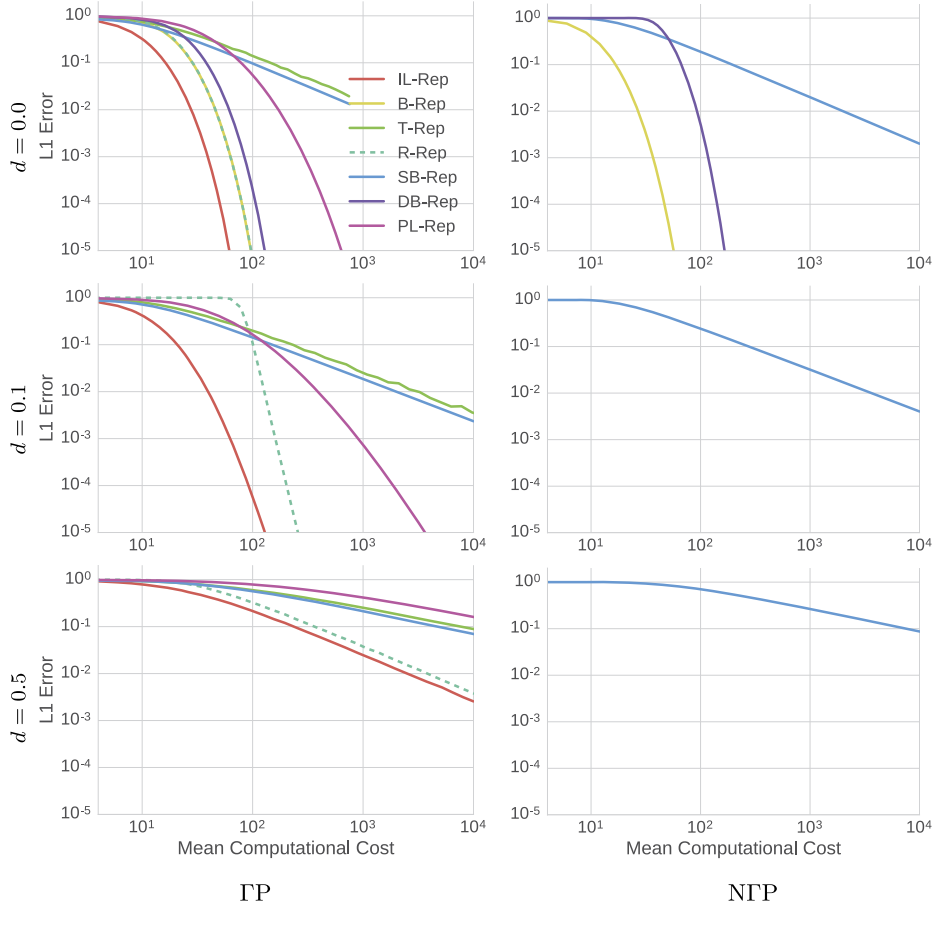

Figure 1. Truncation error bounds for representations of the (normalized) gamma-Poisson process, with $\gamma = 1$, $\lambda = 2$, and $\xi = \gamma\lambda$. The left column is for the unnormalized process, while the right column is for the normalized process. Each row displays results for a different value of the discount parameter $d \in \{0, 0.1, 0.5\}$.

all others use closed-form expressions from the examples in Sections 4 and 5. Note that the Bondesson and decoupled Bondesson representations do not exist when $d > 0$. Further, only those representations for which we provide closed-form bounds in the examples are shown for the normalized gamma process; we leave the numerical approximation of the results from Theorems 5.3 and 5.4 as an open problem. Similar figures for other processes (in particular, the beta-Bernoulli and beta prime-odds Bernoulli) are provided in Appendix A. Note that all bounds presented are improved by a factor of two versus comparable past results in the literature, due to the reliance on Lemmas 4.1 and 5.1 rather than the earlier bound found in [19].

In Figure 1, the top row shows results for the light-tailed process ($\gamma = 1$, $\lambda = 1$, $d = 0$, and $\xi = c = \gamma\lambda$). All representations except for thinning and size-biased capture its exponential truncation error decay. This is due to the fact that the thinning representation generates increasingly many atoms of weight 0 as $K \to \infty$, and the expected number of atoms at each outer index for the size-biased representation decays as $K \to \infty$. The inverse-Lévy representation has the lowest truncation error as expected, as it is the only representation that generates a nonincreasing sequence of weights (and so must be the most efficient [3]). Based on this figure and those in Appendix A for other processes, it appears that the Bondesson representation typically provides the best tradeoff between simplicity and efficiency, and should be used whenever its conditions in Theorem 3.1 are satisfied. When the technical conditions are not satisfied, the rejection representation is a good alternative. If ease of theoretical analysis is a concern, the decoupled Bondesson representation provides comparable efficiency with the analytical simplicity of a superposition representation.

The bottom two rows of Figure 1 show results for the heavy-tailed process ($\gamma = 1$, $\lambda = 2$, and $d \in \{0.1, 0.5\}$). The representation options are more limited, as the technical conditions of the Bondesson and decoupled Bondesson representations are not satisfied. Here the rejection representation is often the best choice due to its simplicity and competitive performance with the inverse-Lévy representation. However, one must take care to check its efficiency beforehand using Proposition 3.2 given a particular choice of $\mu(\mathrm{d}\theta)$. For example, the choice of $\mu(\mathrm{d}\theta) \propto \theta^{-1-d}\,\mathrm{d}\theta$ in the present work makes the rejection representation very inefficient when $d \ll 1$ for both the gamma-Poisson (Figure 1) and beta prime-odds Bernoulli (Figure A.2) processes, but efficient for the beta-Bernoulli process (Figure A.1). If no $\mu(\mathrm{d}\theta)$ yields reasonable results, the power-law representation is a good choice for $d \ll 1$ as its truncation bound approaches the exponential decay of the light-tailed process. For larger $d > 0$, the size-biased representation is a good alternative.

Based on the results in Figure 1, it appears that there is no single dominant representation for all situations (provided the inverse-Lévy representation is intractable, as it most often is). However, as a guideline, the rejection and Bondesson representations tend to be good choices for light-tailed processes, while the rejection, size-biased, and power-law representations are good choices for heavy-tailed processes.

8. Discussion

We have investigated sequential representations, truncation error bounds, and simulation algorithms for (normalized) completely random measures. In past work, the development and analysis

of these tools has occurred only on an ad hoc basis. The results in the present paper, in contrast, provide a comprehensive characterization and analysis of the different types of sequential (N)CRM representations available to the practitioner. However, there are a number of remaining open questions and limitations.

First, this work does not consider the influence of observed data: all analyses assume an *a priori* perspective, as truncation is typically performed before data are incorporated via posterior inference (e.g., in variational inference for the DP mixture [6] and BP latent feature model [13]). However, analysis of *a posteriori* truncation has been studied in past work as well [16,19,21]. In the language of CRMs, observations introduce a fixed-location component in the posterior process, while the unobserved traits are drawn from the (possibly normalized) ordinary component of a CRM [11,21]. We anticipate that this property makes observations reasonably simple to include: the truncation tools provided in the present paper can be used directly on the unobserved ordinary component, while the fixed-location component may be treated exactly.

In addition, there are important open questions regarding the sequential representations developed in this work. It is unknown whether generalized versions of the Bondesson and decoupled Bondesson representations can be developed for larger classes of rate measures. The power-law representation does provide a partial answer in the decoupled Bondesson case. Regarding size-biased representations, one might expect that the use of conjugate exponential family CRMs [11] would yield a closed-form expression for the truncation bound. In all of the cases provided in this paper, this was indeed the case; the integrals were evaluated exactly and a closed-form expression was found. However, we were unable to identify a general expression applicable to all conjugate exponential family CRMs. Based on the examples provided, we conjecture that such an expression exists. Finally, fundamental connections between some of the representations were left largely unexplored in this work. This is an open area of research, although progress has been made by connecting decoupled Bondesson and size-biased representations for (hierarchies of) generalized beta processes [45], Section 6.4.

A final remark is that one of the primary uses of sequential representations in past work has been in the development of posterior inference procedures [6,13,38]. The present work provides no guidance on which truncated representations are best paired with which inference methods. We leave this as an open direction for future research, which will require both theoretical and empirical investigation.

Acknowledgments

First authorship is shared jointly by T. Campbell and J.H. Huggins. The authors thank the anonymous reviewers for their thoughtful comments, which led to substantial improvements in both the presentation and content of the paper. The authors also thank Tin Nguyen for pointing out a flaw in the proof of a result appearing in a preprint draft of the paper. All authors are supported by the Office of Naval Research under MURI grant N000141110688. J. Huggins is supported by the U.S. Government under FA9550-11-C-0028 and awarded by the DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. T. Campbell and T. Broderick are supported by DARPA award FA8750-17-2-0019.

Supplementary Material

Supplement to "Truncated random measures" (DOI: 10.3150/18-BEJ1020SUPP; .pdf). Proofs for all results developed in this paper along with some additional example applications and simulations.

References

[1] Abramowitz, M. and Stegun, I.A., eds. (1964). *Handbook of Mathematical Functions*, *with Formulas*, *Graphs*, *and Mathematical Tables*. New York: Dover Publications, Inc. MR0208797

[2] Airoldi, E.M., Blei, D., Erosheva, E.A. and Fienberg, S.E. (2014). *Handbook of Mixed Membership Models and Their Applications*. Boca Raton, FL: CRC Press.

[3] Arbel, J. and Prünster, I. (2017). A moment-matching Ferguson & Klass algorithm. *Stat*. *Comput*. **27** 3–17. MR3598905

[4] Argiento, R., Bianchini, I. and Guglielmi, A. (2016). A blocked Gibbs sampler for NGG-mixture models via a priori truncation. *Stat*. *Comput*. **26** 641–661. MR3489862

[5] Banjevic, D., Ishwaran, H. and Zarepour, M. (2002). A recursive method for functionals of Poisson processes. *Bernoulli* **8** 295–311. MR1913109

[6] Blei, D.M. and Jordan, M.I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Anal*. **1** 121–143. MR2227367

[7] Bondesson, L. (1982). On simulation from infinitely divisible distributions. *Adv. in Appl*. *Probab.* **14** 855–869. MR0677560

[8] Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl*. *Probab.* **31** 929–953. MR1747450

[9] Broderick, T., Jordan, M.I. and Pitman, J. (2012). Beta processes, stick-breaking and power laws. *Bayesian Anal*. **7** 439–475. MR2934958

[10] Broderick, T., Mackey, L., Paisley, J. and Jordan, M.I. (2015). Combinatorial clustering and the beta negative binomial process. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **37** 290–306.

[11] Broderick, T., Wilson, A.C. and Jordan, M.I. (2018). Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli* **24** 3181–3221. MR3788171

[12] Campbell, T., Huggins, J.H., How, J.P. and Broderick, T. (2019). Supplement to "Truncated random measures". DOI:10.3150/18-BEJ1020SUPP.

[13] Doshi-Velez, F., Miller, K.T., Van Gael, J. and Teh, Y.W. (2009). Variational inference for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*.

[14] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann*. *Statist*. **1** 209–230. MR0350949

[15] Ferguson, T.S. and Klass, M.J. (1972). A representation of independent increment processes without Gaussian components. *Ann*. *Math*. *Stat*. **43** 1634–1643. MR0373022

[16] Gelfand, A.E. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J*. *Comput*. *Graph*. *Statist*. **11** 289–305. MR1938136

[17] Gumbel, E.J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications*. *A Series of Lectures*. *National Bureau of Standards Applied Mathematics Series* **33**. Washington, DC: U.S. Government Printing Office. MR0061342

[18] Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann*. *Statist*. **18** 1259–1294. MR1062708

[19] Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J*. *Amer. Statist*. *Assoc*. **96** 161–173. MR1952729

[20] Ishwaran, H. and James, L.F. (2002). Approximate Dirichlet process computing in finite normal mixtures: Smoothing and prior information. *J. Comput. Graph. Statist.* **11** 508–532. MR1938445
[21] Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canad. J. Statist.* **30** 269–283. MR1926065
[22] James, L.F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Preprint. Available at arXiv:0205093.
[23] James, L.F. (2013). Stick-breaking PG(α, ζ)-generalized gamma processes. Preprint. Available at arXiv:1308.6570.
[24] James, L.F. (2014). Poisson latent feature calculus for generalized Indian buffet processes. Preprint. Available at arXiv:1411.2936v3.
[25] James, L.F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. MR2508332
[26] Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27** 562–588. MR1714717
[27] Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185
[28] Kingman, J.F.C. (1993). *Poisson Processes. Oxford Studies in Probability* **3**. New York: The Clarendon Press, Oxford Univ. Press. MR1207584
[29] Kingman, J.F.C., Taylor, S.J., Hawkes, A.G., Walker, A.M., Cox, D.R., Smith, A.F.M., Hill, B.M., Burville, P.J. and Leonard, T. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** 1–22. With a discussion by S.J. Taylor, A.G. Hawkes, A.M. Walker, D.R. Cox, A.F.M. Smith, B.M. Hill, P.J. Burville, T. Leonard and a reply by the author. MR0368264
[30] Lijoi, A., Mena, R.H. and Prünster, I. (2005). Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Stat. Inference Stoch. Process.* **8** 283–309. MR2177315
[31] Lijoi, A., Mena, R.H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69** 715–740. MR2370077
[32] Lijoi, A. and Prünster, I. (2003). On a normalized random measure with independent increments relevant to Bayesian nonparametric inference. In *Proceedings of the* 13*th European Young Statisticians Meeting* 123–124.
[33] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N.L. Hjort, C. Holmes, P. Müller and S. Walker, eds.). *Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge: Cambridge Univ. Press. MR2730661
[34] Maddison, C., Tarlow, D. and Minka, T.P. (2014). A* sampling. In *Advances in Neural Information Processing Systems*.
[35] Muliere, P. and Tardella, L. (1998). Approximating distributions of random functionals of Ferguson–Dirichlet priors. *Canad. J. Statist.* **26** 283–297. MR1648431
[36] Orbanz, P. (2010). Conjugate projective limits. Preprint. Available at arXiv:1012.0363.
[37] Paisley, J.W., Blei, D.M. and Jordan, M.I. (2012). Stick-breaking beta processes and the Poisson process. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research* **22** 850–858.
[38] Paisley, J.W., Zaas, A.K., Woods, C.W., Ginsburg, G.S. and Carin, L. (2010). A stick-breaking construction of the beta process. In *International Conference on Machine Learning*.
[39] Perman, M. (1993). Order statistics for jumps of normalised subordinators. *Stochastic Process. Appl.* **46** 267–281. MR1226412
[40] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448
[41] Pitman, J. (2003). Poisson–Kingman partitions. In *Statistics and Science*: *A Festschrift for Terry Speed. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. Beachwood, OH: IMS. MR2004330

[42] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31** 560–585. Dedicated to the memory of Herbert E. Robbins. MR1983542

[43] Rosiński, J. (1990). On series representations of infinitely divisible random vectors. *Ann. Probab.* **18** 405–430. MR1043955

[44] Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes* (O. Barndorff-Nielson, S. Resnick and T. Mikosch, eds.) 401–415. Boston, MA: Birkhäuser. MR1833707

[45] Roy, D. (2014). The continuum-of-urns scheme, generalized beta and Indian buffer processes, and hierarchies thereof. Preprint. Available at arXiv:1501.00208.

[46] Roychowdhury, A. and Kulis, B. (2015). Gamma processes, stick-breaking, and variational inference. In *International Conference on Artificial Intelligence and Statistics*.

[47] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

[48] Teh, Y.W. and Görür, D. (2009). Indian buffet processes with power-law behavior. In *Advances in Neural Information Processing Systems*.

[49] Thibaux, R. and Jordan, M.I. (2007). Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*.

[50] Titsias, M. (2008). The infinite gamma-Poisson feature model. In *Advances in Neural Information Processing Systems*.

[51] 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*. *Proceedings of Machine Learning Research* **22** 1462–1471.

Received February 2017 and revised July 2017