Bernoulli **22**(4), 2016, 2301–2324
DOI: 10.3150/15-BEJ729

The combinatorial structure of beta negative binomial processes

CREIGHTON HEAUKULANI[1] and DANIEL M. ROY[2]

[1]*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom. E-mail: ckh28@cam.ac.uk*
[2]*Department of Statistical Sciences, University of Toronto, 100 St. George Street, Toronto, ON, M5S 3G3, Canada. E-mail: droy@utstat.toronto.edu*

We characterize the combinatorial structure of conditionally-i.i.d. sequences of negative binomial processes with a common beta process base measure. In Bayesian nonparametric applications, such processes have served as models for latent *multisets* of features underlying data. Analogously, random *subsets* arise from conditionally-i.i.d. sequences of Bernoulli processes with a common beta process base measure, in which case the combinatorial structure is described by the *Indian buffet process*. Our results give a count analogue of the Indian buffet process, which we call a *negative binomial Indian buffet process*. As an intermediate step toward this goal, we provide a construction for the beta negative binomial process that avoids a representation of the underlying beta process base measure. We describe the key Markov kernels needed to use a NB-IBP representation in a Markov Chain Monte Carlo algorithm targeting a posterior distribution.

Keywords: Bayesian nonparametrics; Indian buffet process; latent feature models; multisets

1. Introduction

The focus of this article is on exchangeable sequences of *multisets*, that is, set-like objects in which repetition is allowed. Let Ω be a complete, separable metric space equipped with its Borel σ-algebra $\mathcal{A}$ and let $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$ denote the non-negative integers. By a *point process* on $(\Omega, \mathcal{A})$, we mean a random measure X on $(\Omega, \mathcal{A})$ such that $X(A)$ is a $\mathbb{Z}_+$-valued random variable for every $A \in \mathcal{A}$. Because $(\Omega, \mathcal{A})$ is Borel, we may write

$$X = \sum_{k \le \kappa} \delta_{\gamma_k} \tag{1.1}$$

for a random element κ in $\overline{\mathbb{Z}}_+ := \mathbb{Z}_+ \cup \{\infty\}$ and some – not necessarily distinct – random elements $\gamma_1, \gamma_2, \ldots$ in Ω. We will take the point process X to represent the *multiset* of its unique atoms γ_k with corresponding multiplicities $X\{\gamma_k\}$. We say X is *simple* when $X\{\gamma_k\} = 1$ for all $k \le \kappa$, in which case X represents a set.

In statistical applications, *latent feature models* associate each data point y_n in a dataset with a latent point process X_n from an exchangeable sequence of simple point processes, which we denote by $(X_n)_{n\in\mathbb{N}} := (X_1, X_2, \ldots)$. The unique atoms among the sequence $(X_n)_{n\in\mathbb{N}}$ are referred to as *features*, and a data point is said to *possess* those features appearing in its associated point process. We can also view these latent feature models as generalizations of mixture models that

allow data points to belong to multiple, potentially overlapping clusters [2,10]. For example, in an object recognition task, a model for a dataset consisting of street camera images could associate each image with a subset of object classes – for example, "trees", "cars", and "houses", etc. – appearing in the images. In a document modeling task, a model for a dataset of news articles could associate each document with a subset of topics – for example, "politics", "Europe", and "economics", etc. – discussed in the documents. Recent work in Bayesian nonparametrics utilizing exchangeable sequences of simple point processes have focused on the Indian buffet process (IBP) [7,10], which characterizes the marginal distribution of the sequence $(X_n)_{n\in\mathbb{N}}$ when they are conditionally-i.i.d. Bernoulli processes, given a common beta process base measure [11,24].

If the point processes $(X_n)_{n\in\mathbb{N}}$ are no longer constrained to be simple, then data points can contain multiple copies of features. For example, in the object recognition task, an image could be associated with two cars, two trees, and one house. In the document modeling task, an article could be associated with 100 words from the politics topic, 200 words from the Europe topic, and 40 words from the economics topic. In this article, we describe a count analogue of the IBP called the *negative binomial Indian buffet processes* (NB-IBP), which characterizes the marginal distribution of $(X_n)_{n\in\mathbb{N}}$ when it is a conditionally i.i.d. sequence of negative binomial processes [3,28], given a common beta process base measure. This characterization allows us to describe new Markov Chain Monte Carlo algorithms for posterior inference that do not require numerical integrations over representations of the underlying beta process.

1.1. Results

Let $c > 0$, let $\widetilde{B}_0$ be a non-atomic, finite measure on Ω, and let Π be a Poisson (point) process on $\Omega \times (0, 1]$ with intensity

$$(\mathrm{d}s, \mathrm{d}p) \mapsto cp^{-1}(1-p)^{c-1}\,\mathrm{d}p\widetilde{B}_0(\mathrm{d}s). \tag{1.2}$$

As this intensity is non-atomic and merely σ-finite, Π will have an infinite number of atoms almost surely (a.s.), and so we may write $\Pi = \sum_{j=1}^{\infty}\delta_{(\gamma_j,b_j)}$ for some a.s. unique random elements $b_1, b_2, \ldots$ in $(0, 1]$ and $\gamma_1, \gamma_2, \ldots$ in Ω. From Π, construct the random measure

$$B := \sum_{j=1}^{\infty} b_j\delta_{\gamma_j}, \tag{1.3}$$

which is a *beta process* [11]. The construction of B ensures that the random variables $B(A_1), \ldots, B(A_k)$ are independent for every finite, disjoint collection $A_1, \ldots, A_k \in \mathcal{A}$, and B is said to be *completely random* or equivalently, have *independent increments* [14]. We review completely random measures in Section 2.

The conjugacy of the family of beta distributions with various other exponential families carries over to beta processes and randomizations by probability kernels lying in these same exponential families. The beta process is therefore a convenient choice for further randomizations, or in the language of Bayesian nonparametrics, as a prior stochastic process. For example, previous work has focused on the (simple) point process that takes each atom γ_j with probability b_j for

every $j \geq 1$, which is, conditioned on B, called a *Bernoulli process* (with base measure B) [24]. In this article, we study the point process

$$X := \sum_{j=1}^{\infty} \zeta_j \delta_{\gamma_j}, \tag{1.4}$$

where the random variables $\zeta_1, \zeta_2, \ldots$ are conditionally independent given B and

$$\zeta_j | b_j \sim \mathrm{NB}(r, b_j), \qquad j \in \mathbb{N}, \tag{1.5}$$

for some parameter $r > 0$. Here, $\mathrm{NB}(r, p)$ denotes the negative binomial distribution with parameters $r > 0$, $p \in (0, 1]$, whose probability mass function (p.m.f.) is

$$\mathrm{NB}(z; r, p) := \frac{(r)_z}{z!} p^z (1-p)^r, \qquad z \in \mathbb{Z}_+, \tag{1.6}$$

where $(a)_n := a(a+1)\cdots(a+n-1)$ with $(a)_0 := 1$ is the nth rising factorial. Note that, conditioned on B, the point process X is the (fixed component) of a *negative binomial process* [3,28]. Unconditionally, X is the ordinary component of a beta negative binomial process, which we formally define in Section 2.

Conditioned on B, construct a sequence of point processes $(X_n)_{n\in\mathbb{N}}$ that are i.i.d. copies of X. In this case, $(X_n)_{n\in\mathbb{N}}$ is an exchangeable sequence of beta negative binomial processes, and our primary goal is to characterize the (unconditional) distribution of the sequence. This task is non-trivial because the construction of the point process X in equation (1.4) is not *finitary* in the sense that no finite subset of the atoms of B determines X with probability one. In the case of conditionally-i.i.d. Bernoulli processes, the unconditional distributions of the measures remain in the class of Bernoulli processes, and so a finitary construction is straightforwardly obtained with Poisson (point) processes. Then the distribution of the sequence, which Thibaux and Jordan [24] showed is characterized by the IBP, may be derived immediately from the conjugacy between the classes of beta and Bernoulli processes [11,13,24]. While conjugacy also holds between the classes of beta and negative binomial processes [3,28], the unconditional law of the point process X is no longer that of a negative binomial process; instead, it is the law of a beta negative binomial process.

Existing constructions for beta negative binomial processes truncate the number of atoms in the underlying beta process and typically use slice sampling to remove the error introduced by this approximation asymptotically [3,19,23,28]. In this work, we instead provide a construction for the beta negative binomial process directly, avoiding a representation of the underlying beta process. To this end, note that while the beta process B has a countably infinite number of atoms a.s., it can be shown that B is still an a.s. finite measure [11]. It follows as an easy consequence that the point process X is a.s. finite as well and, therefore, has an a.s. finite number of atoms, which we represent with a Poisson process. The atomic masses are then characterized by the *digamma distribution*, introduced by Sibuya [21], which has p.m.f. (for parameters $r, \theta > 0$) given by

$$\mathrm{digamma}(z; r, \theta) := \frac{1}{\psi(r+\theta) - \psi(\theta)} \frac{(r)_z}{(r+\theta)_z} z^{-1}, \qquad z \geq 1, \tag{1.7}$$

where $\psi(a) := \Gamma'(a)/\Gamma(a)$ denotes the digamma function. In Section 3, we prove the following:

Theorem 1.1. *Let Y be a Poisson process on $(\Omega, \mathcal{A})$ with finite intensity*

$$\mathrm{d}s \mapsto c\big[\psi(c+r) - \psi(c)\big]\widetilde{B}_0(\mathrm{d}s), \tag{1.8}$$

that is, $Y = \sum_{k=1}^{\kappa} \delta_{\gamma_k}$ for a Poisson random variable κ with mean $c[\psi(c+r) - \psi(c)]\widetilde{B}_0(\Omega)$ and i.i.d. random variables $(\gamma_k)_{k\in\mathbb{N}}$, independent from κ, each with distribution $\widetilde{B}_0/\widetilde{B}_0(\Omega)$. Let $(\zeta_k)_{k\in\mathbb{N}}$ be an independent collection of i.i.d. digamma(r, c) *random variables. Then*

$$X \overset{d}{=} \sum_{k=1}^{\kappa} \zeta_k \delta_{\gamma_k}, \tag{1.9}$$

where X is the beta negative binomial process defined in equation (1.4).

With this construction and conjugacy (the relevant results are reproduced in Section 4), characterizing the distribution of $(X_n)_{n\in\mathbb{N}}$ is straightforward. However, in applications we are only interested in the *combinatorial structure* of the sequence $(X_n)_{n\in\mathbb{N}}$, that is, the pattern of sharing amongst the atoms while ignoring the locations of the atoms themselves. More precisely, for every $n \in \mathbb{N}$, let $\mathcal{H}_n := \mathbb{Z}_+^n \setminus \{0^n\}$ be the set of all length-n sequences of non-negative integers, excluding the all-zero sequence. Elements in $\mathcal{H}_n$ are called *histories*, and can be thought of as representations of non-empty multisets of $[n] := \{1, \ldots, n\}$. For every $h \in \mathcal{H}_n$, let M_h be the number of elements $s \in \Omega$ such that $X_j\{s\} = h(j)$ for all $j \le n$. By the combinatorial structure of a finite subsequence $X_{[n]} := (X_1, \ldots, X_n)$, we will mean the collection $(M_h)_{h\in\mathcal{H}_n}$ of counts, which together can be understood as representations of multisets of histories. These counts are combinatorial in the following sense: Let $\phi\colon (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ be a Borel automorphism on $(\Omega, \mathcal{A})$, that is, a measurable permutation of Ω whose inverse is also measurable, and define the transformed processes $X_j^{\phi} := X_j \circ \phi^{-1}$, for every $j \le n$, where each atom s is repositioned to $\phi(s)$. The collection $(M_h)_{h\in\mathcal{H}_n}$ is invariant to this transformation, and it is in this sense that they only capture the combinatorial structure. In Section 4, we prove the following.

Theorem 1.2. *The probability mass function of $(M_h)_{h\in\mathcal{H}_n}$ is*

$$\begin{aligned}&\mathbb{P}\{M_h = m_h\colon h \in \mathcal{H}_n\} \\ &\quad= \frac{(cT)^{\sum_{h\in\mathcal{H}_n} m_h}}{\prod_{h\in\mathcal{H}_n} m_h!} \exp\big(-cT\big[\psi(c+nr) - \psi(c)\big]\big) \prod_{h\in\mathcal{H}_n} \left[\frac{\Gamma(S(h))\Gamma(c+nr)}{\Gamma(c+nr+S(h))} \prod_{j=1}^{n} \frac{(r)_{h(j)}}{h(j)!}\right]^{m_h},\end{aligned} \tag{1.10}$$

where $S(h) := \sum_{j\le n} h(j)$, for every $h \in \mathcal{H}_n$, and $T := \widetilde{B}_0(\Omega) > 0$.

As one would expect, equation (1.10) is reminiscent of the p.m.f. for the IBP, and indeed the collection $(M_h)_{h\in\mathcal{H}_n}$ is characterized by what we call the *negative binomial Indian buffet process*, or NB-IBP. Let beta-NB(r, α, β) denote the *beta negative binomial distribution (with parameters $r, \alpha, \beta > 0$)*, that is, we write $Z \sim$ beta-NB(r, α, β) if there exists a beta random variable $p \sim \mathrm{beta}(\alpha, \beta)$ such that $Z|p \sim \mathrm{NB}(r, p)$. In the NB-IBP, a sequence of customers enters an Indian buffet restaurant:

- The first customer
  - selects Poisson$(c\gamma[\psi(c+r)-\psi(c)])$ distinct dishes, taking digamma(r,c) servings of each dish, independently.
- For $n \geq 1$, the $(n+1)$st customer
  - takes beta-NB$(r, S_{n,k}, c+nr)$ servings of each previously sampled dish k; where $S_{n,k}$ is the total number of servings taken of dish k by the first n customers;
  - selects Poisson$(c\gamma[\psi(c+(n+1)r)-\psi(c+nr)])$ new dishes to taste, taking digamma$(r, c+nr)$ servings of each dish, independently.

The interpretation here is that, for every $h \in \mathcal{H}_n$, the count M_h is the number of dishes k such that, for every $j \leq n$, customer j took $h(j)$ servings of dish k. Then the sum $S(h)$ in equation (1.10) is the total number of servings taken of dish k by the first n customers. Because the NB-IBP is the combinatorial structure of a conditionally i.i.d. process, its distribution, given in Theorem 1.2, must be invariant to every permutation of the customers. We can state this property formally as follows.

Theorem 1.3 (Exchangeability). *Let π be a permutation of $[n] := \{1, \ldots, n\}$, and, for $h \in \mathcal{H}_n$, note that the composition $h \circ \pi \in \mathcal{H}_n$ is given by $(h \circ \pi)(j) = h(\pi(j))$, for every $j \leq n$. Then*

$$(M_h)_{h\in\mathcal{H}_n} \overset{d}{=} (M_{h\circ\pi})_{h\in\mathcal{H}_n}. \tag{1.11}$$

The exchangeability of the combinatorial structure and its p.m.f. in equation (1.10) allows us to develop Gibbs sampling techniques analogous to those originally developed for the IBP [7, 17]. In particular, because the NB-IBP avoids a representation of the beta process underlying the exchangeable sequence $(X_n)_{n\in\mathbb{N}}$, these posterior inference algorithms do not require numerical integration over representations of the beta process. We discuss some of these techniques in Section 5.

2. Preliminaries

Here, we review *completely random measures* and formally define the negative binomial and beta negative binomial processes. We provide characterizations via Laplace functionals and conclude the section with a discussion of related work.

2.1. Completely random measures

Let $\mathcal{M}(\Omega, \mathcal{A})$ denote the space of σ-finite measures on $(\Omega, \mathcal{A})$ equipped with the σ-algebra generated by the projection maps $\mu \mapsto \mu(A)$ for all $A \in \mathcal{A}$. A random measure ξ on $(\Omega, \mathcal{A})$ is a random element in $\mathcal{M}(\Omega, \mathcal{A})$, and we say that ξ is *completely random* or *has independent increments* when, for every finite collection of disjoint, measurable sets $A_1, \ldots, A_n \in \mathcal{A}$, the random variables $\xi(A_1), \ldots, \xi(A_n)$ are independent. Here, we briefly review completely random measures; for a thorough treatment, the reader should consult Kallenberg [12], Chapter 12, or the

classic text by Kingman [14]. Every completely random measure ξ can be written as a sum of three independent parts

$$\xi = \bar{\xi} + \sum_{s \in \mathcal{A}} \vartheta_s \delta_s + \sum_{(s,p) \in \eta} p \delta_s \qquad \text{a.s.}, \tag{2.1}$$

called the *diffuse*, *fixed*, and *ordinary* components, respectively, where:

1. $\bar{\xi}$ is a non-random, non-atomic measure;
2. $\mathcal{A} \subseteq \Omega$ is a non-random countable set whose elements are referred to as the *fixed atoms* and whose masses $\vartheta_1, \vartheta_2, \ldots$ are independent random variables in $\mathbb{R}_+$ (the non-negative real numbers);
3. η is a Poisson process on $\Omega \times (0, \infty)$ whose intensity $\mathbb{E}\eta$ is σ-finite and has diffuse projections onto Ω, that is, the measure $(\mathbb{E}\eta)(\cdot \times (0, \infty))$ on Ω is non-atomic.

In this article, we will only study purely-atomic completely random measures, which therefore have no diffuse component. It follows that we may characterize the law of ξ by (1) the distributions of the atomic masses in the fixed component, and (2) the intensity of the Poisson process underlying the ordinary component.

2.2. Definitions

By a *base measure* on $(\Omega, \mathcal{A})$, we mean a σ-finite measure B on $(\Omega, \mathcal{A})$ such that $B\{s\} \leq 1$ for all $s \in \Omega$. For the remainder of the article, fix a base measure B_0. We may write

$$B_0 = \widetilde{B}_0 + \sum_{s \in \mathcal{A}} \bar{b}_s \delta_s \tag{2.2}$$

for some non-atomic measure $\widetilde{B}_0$; a countable set $\mathcal{A} \subseteq \Omega$; and constants $\bar{b}_1, \bar{b}_2, \ldots$ in $(0, 1]$.[1] As discussed in the Introduction, a convenient model for random base measures are *beta processes*, a class of completely random measures introduced by Hjort [11]. For the remainder of the article, let $c\colon \Omega \to \mathbb{R}_+$ be a measurable function, which we call a *concentration function* (or *parameter* when it is constant).

Definition 2.1 (Beta process). *A random measure B on $(\Omega, \mathcal{A})$ is a* beta process with concentration function c and base measure B_0, *written $B \sim \mathrm{BP}_{\mathcal{L}}(c, B_0)$, when it is purely atomic and completely random, with a fixed component*

$$\sum_{s \in \mathcal{A}} \vartheta_s \delta_s, \qquad \vartheta_s \overset{\text{ind}}{\sim} \mathrm{beta}\big(c(s)\bar{b}_s, c(s)(1 - \bar{b}_s)\big), \tag{2.3}$$

and an ordinary component with intensity measure

$$(\mathrm{d}s, \mathrm{d}p) \mapsto c(s)\, p^{-1}(1-p)^{c(s)-1}\, \mathrm{d}p\, \widetilde{B}_0(\mathrm{d}s). \tag{2.4}$$

[1] Note that we have relaxed the condition on $\widetilde{B}_0$ (in the Introduction) to be merely σ-finite.

It is straightforward to show that a beta process is itself a base measure with probability one. This definition of the beta process generalizes the version given in the introduction to a non-homogeneous process with a fixed component. Likewise, we generalize our earlier definition of a *negative binomial process* to include an ordinary component.

Definition 2.2 (Negative binomial process). *A point process X on $(\Omega, \mathcal{A})$ is a* negative binomial process with parameter $r > 0$ and base measure B_0, *written $X \sim \mathrm{NBP}(r, B_0)$, when it is purely atomic and completely random, with a fixed component*

$$\sum_{s \in \mathcal{A}} \vartheta_s \delta_s, \qquad \vartheta_s \overset{\text{ind}}{\sim} \mathrm{NB}(r, \bar{b}_s), \tag{2.5}$$

and an ordinary component with intensity measure

$$(\mathrm{d}s, \mathrm{d}p) \mapsto r\delta_1(\mathrm{d}p)\widetilde{B}_0(\mathrm{d}s). \tag{2.6}$$

The fixed component in this definition was given by Broderick *et al.* [3] and Zhou *et al.* [28] (and by Thibaux [25] for the case $r = 1$). Here, we have additionally defined an ordinary component, following intuitions from Roy [20].

The law of a random measure is completely characterized by its Laplace functional, and this representation is often simpler to manipulate: From Campbell's theorem, or a version of the Lévy–Khinchin formula for Borel spaces, one can show that the Laplace functional of X is

$$f \mapsto \mathbb{E}\big[e^{-X(f)}\big] = \exp\left[-\int \big(1 - e^{-f(s)}\big) r \widetilde{B}_0(\mathrm{d}s)\right] \prod_{s \in \mathcal{A}} \left[\frac{1 - \bar{b}_s}{1 - \bar{b}_s e^{-f(s)}}\right]^r, \tag{2.7}$$

where f ranges over non-negative measurable functions and $X(f) := \int f(s) X(\mathrm{d}s)$.

Finally, we define *beta negative binomial processes* via their conditional law.

Definition 2.3 (Beta negative binomial process). *A random measure X on $(\Omega, \mathcal{A})$ is a* beta negative binomial process with parameter $r > 0$, concentration function c, and base measure B_0, *written*

$$X \sim \mathrm{BNBP}(r, c, B_0),$$

if there exists a beta process $B \sim \mathrm{BP}_{\mathcal{L}}(c, B_0)$ such that

$$X | B \sim \mathrm{NBP}(r, B). \tag{2.8}$$

This characterization was given by Broderick *et al.* [3] and can be seen to match a special case of the model in Zhou *et al.* [28] (see the discussion of related work in Section 2.3). It is

straightforward to show that a beta negative binomial process is also completely random, and that its Laplace functional is given by

$$\begin{aligned}\mathbb{E}\big[e^{-X(f)}\big] = \exp\bigg[-\int\bigg[1-\bigg(\frac{1-p}{1-pe^{-f(s)}}\bigg)^r\bigg]c(s)p^{-1}(1-p)^{c(s)-1}\,\mathrm{d}p\,\widetilde{B}_0(\mathrm{d}s)\bigg] \\ \times\prod_{s\in\mathcal{A}}\int\bigg(\frac{1-p}{1-pe^{-f(s)}}\bigg)^r \mathrm{beta}\big(p; c(s)\bar{b}_s, c(s)(1-\bar{b}_s)\big)\,\mathrm{d}p,\end{aligned} \tag{2.9}$$

for $f\colon\Omega\to\mathbb{R}_+$ measurable, where we note that the factors in the product term take the form of the Laplace transform of the beta negative binomial distribution.

2.3. Related work

The term "negative binomial process" has historically been reserved for processes with negative binomial increments – a class into which the process we study here does not fal – and these processes have been long-studied in probability and statistics. We direct the reader to Kozubowski and Podgórski [15] for references.

One way to construct a process with negative binomial increments is to rely upon the fact that a negative binomial distribution is a gamma mixture of Poisson distributions. In particular, similarly to the construction by Lo [16], consider a Cox process X directed by a gamma process G with finite non-atomic intensity. So constructed, X has independent increments with negative binomial distributions. Like the beta process (with a finite intensity underlying its ordinary component), the gamma process has, with probability one, a countably infinite number of atoms but a finite total mass, and so the Cox process X is a.s. finite as well. Despite similarities, a comparison of Laplace functionals shows that the law of X is not that of a beta negative binomial process. Using an approach directly analogous to the derivation of the IBP in [10], Titsias [26] characterizes the combinatorial structure of a sequence of point processes that, conditioned on G, are independent and identically distributed to the Cox process X. See Section 4 for comments. This was the first count analogue of the IBP; the possibility of a count analogue arising from beta negative binomial processes was first raised by Zhou *et al.* [28], who described the distribution of the number of new dishes sampled by each customer. Recent work by Zhou, Madrid and Scott [29], independent of our own and proceeding along different lines, describes a combinatorial process related to the NB-IBP (following a re-scaling of the beta process intensity).

Finally, we note that another negative binomial process without negative binomial increments was defined on Euclidean space by Barndorff-Nielsen and Yeo [1] and extended to general spaces by Grégoire [9] and Wolpert and Ickstadt [27]. These measures are generally Cox processes on $(\Omega,\mathcal{A})$ directed by random measures of the form

$$\mathrm{d}s\mapsto\int_{\mathbb{R}_+}\nu(t,\mathrm{d}s)G(\mathrm{d}t),$$

where G is again a gamma process, this time on $\mathbb{R}_+$, and ν is a probability kernel from Ω to $\mathbb{R}_+$, for example, the Gaussian kernel.

3. Constructing beta negative binomial processes

Before providing a finitary construction for the beta negative binomial process, we make a few remarks on the digamma distribution. For the remainder of the article, define $\lambda_{r,\theta} := \psi(\theta + r) - \psi(\theta)$ for some $r, \theta > 0$. Following a representation by Sibuya [21], we may relate the digamma and beta negative binomial distributions as follows: Let $Z \sim \text{digamma}(r, \theta)$ and define $W := Z - 1$, the latter of which has p.m.f.

$$\mathbb{P}\{W = w\} = (\theta\lambda_{r,\theta})^{-1} \frac{w + r}{w + 1} \text{beta-NB}(w; r, 1, \theta), \qquad w \in \mathbb{Z}_+. \tag{3.1}$$

Deriving the Laplace transform of the law of W is straightforward, and because $\mathbb{E}[e^{-tW}] = e^t \mathbb{E}[e^{-tZ}]$, one may verify that the Laplace transform of the digamma distribution is given by

$$\Psi_{r,\theta}(t) := \mathbb{E}\big[e^{-tZ}\big] = 1 - \lambda_{r,\theta}^{-1} \int \left[1 - \left(\frac{1-p}{1-pe^{-t}}\right)^r\right] p^{-1}(1-p)^{\theta-1}\,\mathrm{d}p. \tag{3.2}$$

The form of equation (3.1) suggests the following rejection sampler, which was first proposed by Devroye [6], Proposition 2, Remark 1: Let $r > 0$ and let $(U_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of uniformly distributed random numbers. Let

$$(Y_n)_{n\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \text{beta-NB}(r, 1, \theta),$$

and define $\eta := \inf\{n \in \mathbb{N} : \max\{r, 1\} \cdot U_n < \frac{Y_i + r}{Y_i + 1}\}$. Then

$$Y_\eta + 1 \sim \text{digamma}(r, \theta),$$

and

$$\mathbb{E}\eta = \frac{\max\{r, 1\}}{\theta[\psi(r + \theta) - \psi(\theta)]}; \qquad \mathbb{E}\eta < \max\{r, r^{-1}\}.$$

With digamma random variables, we provide a finitary construction for the beta negative binomial process. The following result generalizes the statement given by Theorem 1.1 (in the Introduction) to a non-homogeneous process, which also has a fixed component.

Theorem 3.1. *Let $r > 0$, and let $\vartheta := (\vartheta_s)_{s\in\mathcal{A}}$ be a collection of independent random variables with*

$$\vartheta_s \sim \text{beta-NB}\big(r, c(s)\bar{b}_s, c(s)(1 - \bar{b}_s)\big), \qquad s \in \mathcal{A}. \tag{3.3}$$

Let Y be a Poisson process on $(\Omega, \mathcal{A})$, independent from ϑ, with (finite) intensity

$$\mathrm{d}s \mapsto c(s)\big[\psi\big(c(s) + r\big) - \psi\big(c(s)\big)\big]\widetilde{B}_0(\mathrm{d}s). \tag{3.4}$$

Write $Y = \sum_{k=1}^{\kappa} \delta_{\gamma_k}$ *for some random element* κ *in* $\mathbb{Z}_+$ *and a.s. unique random elements* $\gamma_1, \gamma_2, \ldots$ *in* Ω, *and put* $\mathcal{F} := \sigma(\kappa, \gamma_1, \gamma_2, \ldots)$. *Let* $(\zeta_j)_{j\in\mathbb{N}}$ *be a collection of random variables that are independent from* ϑ *and are conditionally independent given* $\mathcal{F}$, *and let*

$$\zeta_j | \mathcal{F} \sim \text{digamma}\big(r, c(\gamma_j)\big), \qquad j \in \mathbb{N}. \tag{3.5}$$

Then

$$X = \sum_{s\in\mathcal{A}} \vartheta_s \delta_s + \sum_{j=1}^{\kappa} \zeta_j \delta_{\gamma_j} \sim \text{BNBP}(r, c, B_0). \tag{3.6}$$

Proof. We have

$$\mathbb{E}^{\mathcal{F}}\big[e^{-X(f)}\big] = \prod_{s\in\mathcal{A}} \mathbb{E}\big[e^{-\vartheta_s f(s)}\big] \times \prod_{j=1}^{\kappa} \mathbb{E}^{\mathcal{F}}\big[e^{-\zeta_j f(\gamma_j)}\big], \tag{3.7}$$

for every $f\colon \Omega \to \mathbb{R}_+$ measurable. For $s \in \Omega$, write $g(s) = \Psi_{r,c(s)}(f(s))$ for the Laplace transform of the digamma distribution evaluated at $f(s)$, where $\Psi_{r,\theta}(t)$ is given by equation (3.2). We may then write

$$\prod_{j=1}^{\kappa} \mathbb{E}^{\mathcal{F}}\big[e^{-\zeta_j f(\gamma_j)}\big] = \prod_{j=1}^{\kappa} g(\gamma_j). \tag{3.8}$$

Then by the chain rule of conditional expectation, complete randomness, and Campbell's theorem,

$$\mathbb{E}\big[e^{-X(f)}\big] = \prod_{s\in\mathcal{A}} \mathbb{E}\big[e^{-\vartheta_s f(s)}\big] \times \exp\bigg[-\int_{\Omega} \big(1 - g(s)\big) c(s) \lambda_{r,c(s)} \widetilde{B}_0(\mathrm{d}s)\bigg] \tag{3.9}$$

$$= \prod_{s\in\mathcal{A}} \bigg[\int \bigg(\frac{1-p}{1-pe^{-f(s)}}\bigg)^r \text{beta}\big(p; c(s)\bar{b}_s, c(s)(1-\bar{b}_s)\big)\,\mathrm{d}p\bigg] \tag{3.10}$$

$$\times \exp\bigg[-\int_{(0,1]\times\Omega} \bigg[1 - \bigg(\frac{1-p}{1-pe^{-f(s)}}\bigg)^r\bigg] c(s) p^{-1} (1-p)^{c(s)-1}\,\mathrm{d}p\,\widetilde{B}_0(\mathrm{d}s)\bigg],$$

which is the desired form of the Laplace functional. □

A finitary construction for conditionally-i.i.d. sequences of negative binomial processes with a common beta process base measure now follows from known conjugacy results. In particular, for every $n \in \mathbb{N}$, let $X_{[n]} := (X_1, \ldots, X_n)$. The following theorem characterizes the conjugacy between the (classes of) beta and negative binomial processes and follows from repeated application of the results by Kim [13], Theorem 3.3 or Hjort [11], Corollary 4.1. This result, which is tailored to our needs, is similar to those already given by Broderick *et el.* [3] and Zhou *et al.* [28], and generalizes the result given by Thibaux [25] for the case $r = 1$.

Theorem 3.2 (Hjort [11], Zhou *et al.* [28]). *Let $B \sim \mathrm{BP}_{\mathcal{L}}(c, B_0)$ and, conditioned on B, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. negative binomial processes with parameter $r > 0$ and base measure B. Then for every $n \in \mathbb{N}$,*

$$B|X_{[n]} \sim \mathrm{BP}_{\mathcal{L}}\left(c_n, \frac{c}{c_n}B_0 + \frac{1}{c_n}S_n\right), \tag{3.11}$$

where $S_n := \sum_{i=1}^{n} X_i$ and $c_n(s) := c(s) + S_n\{s\} + nr$, for $s \in \Omega$.

Remark 3.1. It follows immediately that, for every $n \in \mathbb{N}$, the law of X_{n+1} conditioned on $X_1, \ldots, X_n$ is given by

$$X_{n+1}|X_{[n]} \sim \mathrm{BNBP}\left(r, c_n, \frac{c}{c_n}B_0 + \frac{1}{c_n}S_n\right). \tag{3.12}$$

We may therefore construct this exchangeable sequence of beta negative binomial processes with Theorem 3.1.

4. Combinatorial structure

We now characterize the combinatorial structure of the exchangeable sequence $X_{[n]}$ in the case when $c > 0$ is constant and $B_0(= \widetilde{B}_0)$ is non-atomic. In order to make this precise, we introduce a quotient of the space of sequences of integer-valued measures. Let $n \in \mathbb{N}$ and for any pair $U := (U_1, \ldots, U_n)$ and $V := (V_1, \ldots, V_n)$ of (finite) sequences of integer-valued measures, write $U \sim V$ when there exists a Borel automorphism ϕ on $(\Omega, \mathcal{A})$ satisfying $U_j = V_j \circ \phi^{-1}$ for every $j \le n$. It is easy to verify that $\sim$ is an equivalence relation. Let $[[U]]$ denote the equivalence class containing U. The quotient space induced by $\sim$ is itself a Borel space, and can be related to the Borel space of sequences of $\mathbb{Z}_+$-valued measures by coarsening the σ-algebra to that generated by the functionals

$$\mathcal{M}_h(U_1, \ldots, U_n) := \#\big\{s \in \Omega : \forall j \le n, U_j\{s\} = h(j)\big\}, \qquad h \in \mathcal{H}_n, j \le n, \tag{4.1}$$

where $\#A$ denotes the cardinality of A, and $\mathcal{H}_n := \mathbb{Z}_+^n \setminus \{0^n\}$ is the space of histories defined in the Introduction. The collection $(M_h)_{h\in\mathcal{H}_n}$ of multiplicities (of histories) corresponding to $X_{[n]}$, also defined in the Introduction, then satisfies $M_h = \mathcal{M}_h(X_{[n]})$ for every $h \in \mathcal{H}_n$. The collection $(M_h)_{h\in\mathcal{H}_n}$ thus identifies a point in the quotient space induced by $\sim$. Our aim is to characterize the distribution of $(M_h)_{h\in\mathcal{H}_n}$, for every $n \in \mathbb{N}$.

Let $\hbar \in \mathcal{H}_n$, and define $\mathcal{H}_{n+1}^{(\hbar)} := \{h \in \mathcal{H}_{n+1} : \forall j \le n, h(j) = \hbar(j)\}$ to be the collection of histories in $\mathcal{H}_{n+1}$ that agree with $\hbar$ on the first n entries. Then note that

$$M_\hbar = \sum_{h\in\mathcal{H}_{n+1}^{(\hbar)}} M_h, \qquad \hbar \in \mathcal{H}_n, \tag{4.2}$$

that is, the multiplicities $(M_h)_{h\in\mathcal{H}_{n+1}}$ at stage $n+1$ completely determine the multiplicities $(M_\hbar)_{\hbar\in\mathcal{H}_n}$ at all earlier stages. It follows that

$$\begin{aligned}\mathbb{P}\{M_h = m_h\colon h\in\mathcal{H}_{n+1}\} &= \mathbb{P}\{M_\hbar = m_\hbar\colon \hbar\in\mathcal{H}_n\}\\ &\quad\times \mathbb{P}\{M_h = m_h\colon h\in\mathcal{H}_{n+1} | M_\hbar = m_\hbar\colon \hbar\in\mathcal{H}_n\},\end{aligned} \tag{4.3}$$

where $m_\hbar = \sum_{h\in\mathcal{H}_{n+1}^{(\hbar)}} m_h$ for $\hbar\in\mathcal{H}_n$. The structure of equation (4.3) suggests an inductive proof for Theorem 1.2.

4.1. The law of M_h for $h\in\mathcal{H}_1$

Note that $\mathcal{H}_1$ is isomorphic to $\mathbb{N}$ and that the collection $(M_h)_{h\in\mathcal{H}_1}$ counts the number of atoms of each positive integer mass. It follows from Theorem 1.1 and a transfer argument [12], Propositions 6.10, 6.11 and 6.13, that there exists:

1. a Poisson random variable κ with mean $cT\lambda_{r,c}$, where $T := \widetilde{B}_0(\Omega) < \infty$;
2. an i.i.d. collection of a.s. unique random elements $\gamma_1, \gamma_2, \ldots$ in Ω;
3. an i.i.d. collection $(\zeta_j)_{j\in\mathbb{N}}$ of digamma(r,c) random variables;

all mutually independent, such that

$$X_1 = \sum_{j=1}^{\kappa} \zeta_j \delta_{\gamma_j} \qquad \text{a.s.}$$

It follows that

$$M_h = \#\big\{j\le\kappa\colon \zeta_j = h(1)\big\} \qquad \text{a.s., for } h\in\mathcal{H}_1, \tag{4.4}$$

and $\kappa = \sum_{h\in\mathcal{H}_1} M_h$ a.s. Therefore,

$$\begin{aligned}&\mathbb{P}\{M_h = m_h\colon h\in\mathcal{H}_1\}\\ &\quad= \mathbb{P}\Big\{\kappa = \sum_{h\in\mathcal{H}_1} m_h\Big\}\mathbb{P}\Big\{M_h = m_h\colon h\in\mathcal{H}_1\Big|\kappa = \sum_{h\in\mathcal{H}_1} m_h\Big\}.\end{aligned} \tag{4.5}$$

Because $\zeta_1, \zeta_2, \ldots$ are i.i.d., the collection $(M_h)_{h\in\mathcal{H}_1}$ has a multinomial distribution conditioned on its sum κ. Namely, M_h counts the number of times, in κ independent trials, that the multiplicity $h(1)$ arises from a digamma(r,c) distribution. In particular,

$$\begin{aligned}&\mathbb{P}\Big\{M_h = m_h\colon h\in\mathcal{H}_1\Big|\kappa = \sum_{h\in\mathcal{H}_1} m_h\Big\}\\ &\quad= \frac{(\sum_{h\in\mathcal{H}_1} m_h)!}{\prod_{h\in\mathcal{H}_1}(m_h!)} \prod_{h\in\mathcal{H}_1}\big[\text{digamma}\big(h(1); r, c\big)^{m_h}\big].\end{aligned} \tag{4.6}$$

It follows that

$$\begin{aligned}&\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_1\}\\&= \frac{(cT\lambda_{r,c})^{\sum_{h\in\mathcal{H}_1} m_h}}{\prod_{h\in\mathcal{H}_1}(m_h!)} \exp(-cT\lambda_{r,c}) \prod_{h\in\mathcal{H}_1} \big[\text{digamma}\big(h(1); r, c\big)^{m_h}\big].\end{aligned} \tag{4.7}$$

4.2. The conditional law of M_h for $h \in \mathcal{H}_{n+1}$

Let $S_n := \sum_{j=1}^n X_j$. Recall that $s(\hbar) := \sum_{j\le n} \hbar(j)$ for $\hbar \in \mathcal{H}_n$. We may write

$$S_n = \sum_{\hbar\in\mathcal{H}_n} \sum_{j=1}^{M_\hbar} s(\hbar)\delta_{\omega_{\hbar,j}}, \tag{4.8}$$

for some collection $\omega := (\omega_{\hbar,j})_{\hbar\in\mathcal{H}_n, j\in\mathbb{N}}$ of a.s. distinct random elements in Ω. It follows from Remark 3.1, Theorem 1.1, and a transfer argument that there exists:

1. a Poisson random variable κ with mean $cT\lambda_{r,c+nr}$;
2. an i.i.d. collection of a.s. unique random elements $\gamma_1, \gamma_2, \ldots$ in Ω, a.s. distinct also from ω;
3. an i.i.d. collection $(\zeta_j)_{j\in\mathbb{N}}$ of digamma$(r, c + nr)$ random variables;
4. for each $\hbar \in \mathcal{H}_n$, an i.i.d. collection $(\vartheta_{\hbar,j})_{j\in\mathbb{N}}$ of random variables satisfying

$$\vartheta_{\hbar,j} \sim \text{beta-NB}\big(r, s(\hbar), c + nr\big) \qquad \text{for } j \in \mathbb{N};$$

all mutually independent and independent of $X_{[n]}$, such that

$$X_{n+1} = \sum_{\hbar\in\mathcal{H}_n} \sum_{j=1}^{M_\hbar} \vartheta_{\hbar,j}\delta_{\omega_{\hbar,j}} + \sum_{j=1}^{\kappa} \zeta_j\delta_{\gamma_j} \qquad \text{a.s.} \tag{4.9}$$

Conditioned on $X_{[n]}$, the first and second terms on the right-hand side correspond to the fixed and ordinary components of X_{n+1}, respectively. Let

$$\mathcal{H}_{n+1}^{(0)} := \big\{h \in \mathcal{H}_{n+1} : h(j) = 0,\ j \le n\big\} \tag{4.10}$$

be the set of histories h for which $h(n + 1)$ is the first non-zero element. Then, with probability one,

$$M_h = \#\big\{j \le \kappa : \zeta_j = h(n + 1)\big\} \qquad \text{for } h \in \mathcal{H}_{n+1}^{(0)}, \tag{4.11}$$

and

$$M_h = \#\big\{j \le M_\hbar : \vartheta_{\hbar,j} = h(n + 1)\big\} \qquad \text{for } \hbar \in \mathcal{H}_n \text{ and } h \in \mathcal{H}_{n+1}^{(\hbar)}. \tag{4.12}$$

By the stated independence of the variables above, we have

$$\begin{aligned}&\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_{n+1} | M_\hbar = m_\hbar : \hbar \in \mathcal{H}_n\} \\ &\quad = \mathbb{P}\big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(0)}\big\} \prod_{\hbar \in \mathcal{H}_n} \mathbb{P}\big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(\hbar)} | M_\hbar = m_\hbar\big\}.\end{aligned} \tag{4.13}$$

Let $\mathcal{H}_{n+1}^{+} := \bigcup_{\hbar \in \mathcal{H}_n} \mathcal{H}_{n+1}^{(\hbar)}$. For every $\hbar \in \mathcal{H}_n$, the random variables $\vartheta_{\hbar,1}, \vartheta_{\hbar,2}, \ldots$ are i.i.d., and therefore, conditioned on $M_\hbar$, the collection $(M_h)_{h \in \mathcal{H}_{n+1}^{(\hbar)}}$ has a multinomial distribution. In particular, the product term in equation (4.13) is given by

$$\begin{aligned}&\prod_{\hbar \in \mathcal{H}_n} \mathbb{P}\big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(\hbar)} | M_\hbar = m_\hbar\big\} \\ &\quad = \frac{\prod_{\hbar \in \mathcal{H}_n} (m_\hbar !)}{\prod_{h \in \mathcal{H}_{n+1}^{+}} (m_h !)} \prod_{h \in \mathcal{H}_{n+1}^{+}} \big[\text{beta-NB}\big(h(n+1); r, S(h) - h(n+1), c + nr\big)^{m_h}\big].\end{aligned}$$

The p.m.f. of the beta negative binomial distribution is given by

$$\text{beta-NB}(z; r, \alpha, \beta) = \frac{(r)_z}{z} \frac{\mathcal{B}(z+\alpha, r+\beta)}{\mathcal{B}(\alpha, \beta)}, \qquad z \in \mathbb{Z}_+, \tag{4.14}$$

for positive parameters r, α, and β, where $\mathcal{B}(\alpha, \beta) := \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ denotes the beta function. We have that $\kappa = \sum_{h \in \mathcal{H}_{n+1}^{(0)}} M_h$ a.s., and therefore

$$\begin{aligned}&\mathbb{P}\big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(0)}\big\} \\ &\quad = \mathbb{P}\Big\{\kappa = \sum_{h \in \mathcal{H}_{n+1}^{(0)}} m_h\Big\} \\ &\qquad \times \mathbb{P}\Big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(0)} \Big| \kappa = \sum_{h \in \mathcal{H}_{n+1}^{(0)}} m_h\Big\}.\end{aligned} \tag{4.15}$$

Because $\zeta_1, \zeta_2, \ldots$ are i.i.d., conditioned on the sum κ, the collection $(M_h)_{h \in \mathcal{H}_{n+1}^{(0)}}$ has a multinomial distribution, and so

$$\begin{aligned}&\mathbb{P}\Big\{M_h = m_h : h \in \mathcal{H}_{n+1}^{(0)} \Big| \kappa = \sum_{h \in \mathcal{H}_{n+1}^{(0)}} m_h\Big\} \\ &\quad = \frac{(\sum_{h \in \mathcal{H}_{n+1}^{(0)}} m_h)!}{\prod_{h \in \mathcal{H}_{n+1}^{(0)}} (m_h !)} \prod_{h \in \mathcal{H}_{n+1}^{(0)}} \big[\text{digamma}\big(h(n+1); r, c + nr\big)^{m_h}\big].\end{aligned} \tag{4.16}$$

It follows that

$$\begin{aligned}
&\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_{n+1} | M_{\hbar} = m_{\hbar} : \hbar \in \mathcal{H}_n\} \\
&= \frac{(cT\lambda_{r,c+nr})^{\sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h}}{(\sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h)!} \exp(-cT\lambda_{r,c+nr}) \\
&\quad \times \frac{\prod_{\hbar\in\mathcal{H}_n}(m_{\hbar}!)}{\prod_{h\in\mathcal{H}_{n+1}^{+}}(m_h!)} \prod_{h\in\mathcal{H}_{n+1}^{+}} \big[\text{beta-NB}\big(h(n+1); r, S(h) - h(n+1), c+nr\big)^{m_h}\big] \\
&\quad \times \frac{(\sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h)!}{\prod_{h\in\mathcal{H}_{n+1}^{(0)}}(m_h!)} \prod_{h\in\mathcal{H}_{n+1}^{(0)}} \big[\text{digamma}\big(h(n+1); r, c+nr\big)^{m_h}\big].
\end{aligned} \tag{4.17}$$

Proof of Theorem 1.2. The proof is by induction. The p.m.f. $\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_1\}$ is given by equation (4.7), which agrees with equation (1.10) for the case $n = 1$. The conditional p.m.f. $\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_{n+1} | M_{\hbar} = m_{\hbar} : \hbar \in \mathcal{H}_n\}$ is given by equation (4.17). By the inductive hypothesis, the p.m.f. $\mathbb{P}\{M_{\hbar} = m_{\hbar} : \hbar \in \mathcal{H}_n\}$ is given by equation (1.10). Then by equation (4.3), we have

$$\begin{aligned}
&\mathbb{P}\{M_h = m_h : h \in \mathcal{H}_{n+1}\} \\
&= \frac{(cT)^{(\sum_{\hbar\in\mathcal{H}_n} m_{\hbar})}(cT\lambda_{r,c+nr})^{(\sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h)}}{\prod_{h\in\mathcal{H}_{n+1}^{+}}(m_h!)\prod_{h\in\mathcal{H}_{n+1}^{(0)}}(m_h!)} \exp\left(-cT\sum_{j=1}^{n+1}\lambda_{r,c+(j-1)r}\right) \\
&\quad \times \prod_{h\in\mathcal{H}_{n+1}^{+}} \left[\mathcal{B}\big(S(h) - h(n+1), c+nr\big)\prod_{j=1}^{n}\frac{(r)_{h(j)}}{h(j)!}\right. \\
&\quad \left. \times \text{beta-NB}\big(h(n+1); r, S(h) - h(n+1), c+nr\big)\right]^{m_h} \\
&\quad \times \prod_{h\in\mathcal{H}_{n+1}^{(0)}} \big[\text{digamma}\big(h(n+1); r, c+nr\big)\big]^{m_h}.
\end{aligned} \tag{4.18}$$

In the first product term on the right-hand side of equation (4.18), note that, for every $h \in \mathcal{H}_{n+1}^{+}$,

$$\begin{aligned}
&\mathcal{B}\big(S(h) - h(n+1), c+nr\big)\prod_{j=1}^{n}\frac{(r)_{h(j)}}{h(j)!}\text{beta-NB}\big(h(n+1); r, S(h) - h(n+1), c+nr\big) \\
&= \mathcal{B}\big(S(h), c+(n+1)r\big)\prod_{j=1}^{n+1}\frac{(r)_{h(j)}}{h(j)!}.
\end{aligned}$$

In the second product term, note that

$$\begin{aligned}
&\prod_{h\in\mathcal{H}_{n+1}^{(0)}} \big[\text{digamma}\big(h(n+1);r,c+nr\big)\big]^{m_h} \\
&\quad= \prod_{h\in\mathcal{H}_{n+1}^{(0)}} \left[\lambda_{r,c+nr}^{-1}\frac{(r)_{h(n+1)}}{h(n+1)!}\mathcal{B}\big(h(n+1),c+(n+1)r\big)\right]^{m_h} \\
&\quad= \lambda_{r,c+nr}^{-(\sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h)} \prod_{h\in\mathcal{H}_{n+1}^{(0)}} \left[\mathcal{B}\big(h(n+1),c+(n+1)r\big)\prod_{j=1}^{n+1}\frac{(r)_{h(j)}}{h(j)!}\right]^{m_h},
\end{aligned}$$

where for the last equality, we have used the fact that $h(j)=0$ for every $j\le n$ and $h\in\mathcal{H}_{n+1}^{(0)}$. Note that $\sum_{\hbar\in\mathcal{H}_n} m_\hbar + \sum_{h\in\mathcal{H}_{n+1}^{(0)}} m_h = \sum_{h\in\mathcal{H}_{n+1}} m_h$. Then equation (4.18) is equal to

$$\begin{aligned}
&\frac{(cT)^{\sum_{h\in\mathcal{H}_{n+1}} m_h}}{\prod_{h\in\mathcal{H}_{n+1}}(m_h!)}\exp\left(-cT\sum_{j=1}^{n+1}\big[\psi(c+jr)-\psi\big(c+(j-1)r\big)\big]\right) \\
&\qquad\times\prod_{h\in\mathcal{H}_{n+1}}\left[\mathcal{B}\big(S(h),c+(n+1)r\big)\prod_{j=1}^{n+1}\frac{(r)_{h(j)}}{h(j)!}\right]^{m_h}.
\end{aligned} \tag{4.19}$$

Noting that $\sum_{j=1}^{n+1}[\psi(c+jr)-\psi(c+(j-1)r)] = \psi(c+(n+1)r)-\psi(c)$, we obtain the expression in equation (1.10) for $n+1$, as desired. □

By construction, equation (1.10) defines the finite-dimensional marginal distributions of the stochastic process $(M_h)_{h\in\mathcal{H}_\infty}$ with index set $\mathcal{H}_\infty := \bigcup_{n\in\mathbb{N}}\mathcal{H}_n$. The exchangeability result given by Theorem 1.3 then follows from the exchangeability of the sequence $X_{[n]}$.

5. Applications in Bayesian nonparametrics

In Bayesian latent feature models, we assume that there exists a latent set of *features* and that each data point possesses some (finite) subset of the features. The features then determine the distribution of the observed data. In a nonparametric setting, exchangeable sequences of simple point processes can serve as models for the latent sets of features. Similarly, exchangeable sequences of point processes, like those that can be constructed from beta negative binomial processes, can serve as models of latent *multisets* of features. In particular, atoms are features and their (integer-valued) masses indicate multiplicity. In this section, we develop posterior inference procedures for exchangeable sequences of beta negative binomial processes.

5.1. Representations as random arrays/matrices

A convenient way to represent the combinatorial structure of an exchangeable sequence of point processes is via an array/matrix W of non-negative integers, where the rows correspond to point processes and columns correspond to atoms appearing among the point processes. Informally, given an enumeration of the set of all atoms appearing in $X_{[n]}$, the entry $W_{i,j}$ associated with the ith row and jth column is the multiplicity/mass of the atom labeled j in the ith point process X_i.

More carefully, fix $n \in \mathbb{N}$ and let $(M_h)_{h\in\mathcal{H}_n}$ be the combinatorial structure of a sequence $X_1, \ldots, X_n$ of conditionally i.i.d. negative binomial processes, given a shared beta process base measure with concentration parameter $c > 0$ and non-atomic base measure $\widetilde{B}_0$ of finite mass T. Let $\kappa := \sum_{h\in\mathcal{H}_n} M_h$ be the number of unique atoms among $X_{[n]}$. Then W is an $n \times \kappa$ array of non-negative integers such that, for every $h \in \mathcal{H}_n$, there are exactly M_h columns of W equal to h, where h is thought of as a length-n column vector. Note that W will have no columns when $\kappa = 0$.

All that remains is to order the columns of W. Every total order on $\mathcal{H}_n$ induces a unique ordering of the columns of W. Titsias [26] defined a unique ordering in this way, analogous to the left-ordered form defined by Griffiths and Ghahramani [10] for the IBP. In particular, for $h, h' \in \mathcal{H}_n$, let $\preceq$ denote the lexicographic order given by: $h \preceq h'$ if and only if $h = h'$ or $h(\eta) < h'(\eta)$, where η is the first coordinate where h and h' differ. We say W is *left-ordered* when its columns are ordered according to $\preceq$. Because there is a bijection between combinatorial structures $(M_h)_{h\in\mathcal{H}_n}$ and their unique representations by left-ordered arrays, the probability mass function of W is given by equation (1.10).

Other orderings have been introduced in the literature: If we permute the columns of W uniformly at random, then W is the analogue of the *uniform random labeling* scheme described by Broderick, Pitman and Jordan [4] for the IBP. Note that the number of distinct ways of ordering the κ columns is given by the multinomial coefficient

$$\frac{\kappa!}{\prod_{h\in\mathcal{H}_n} M_h!}, \tag{5.1}$$

where the denominator arises from the fact that there are M_h indistinguishable columns for every history $h \in \mathcal{H}_n$. The following result is then immediate:

Theorem 5.1. *Let W be a uniform random labeling of $(M_h)_{h\in\mathcal{H}_n}$ described above, let $w \in \mathbb{Z}_+^{n\times k}$ be an array of non-negative integers with n rows and $k \geq 0$ non-zero columns, and for every $j \leq k$, let $s_j := \sum_{i=1}^n w_{i,j}$ be the sum of column j. Then*

$$\mathbb{P}\{W = w\} = \frac{(cT)^k}{k!} \exp\bigl(-cT\bigl[\psi(c+nr) - \psi(c)\bigr]\bigr) \prod_{j=1}^{k} \left[\frac{\Gamma(s_j)\Gamma(c+nr)}{\Gamma(s_j+c+nr)} \prod_{i=1}^{n} \frac{(r)_{w_{i,j}}}{w_{i,j}!} \right]. \tag{5.2}$$

An array representation makes it easy to visualize some properties of the model. For example, in Figure 1 we display several simulations from the NB-IBP with varying values of the parameters T, c, and r. The columns are displayed in the order of first appearance, and are otherwise ordered uniformly at random. (A similar ordering was used by Griffiths and Ghahramani [10] to

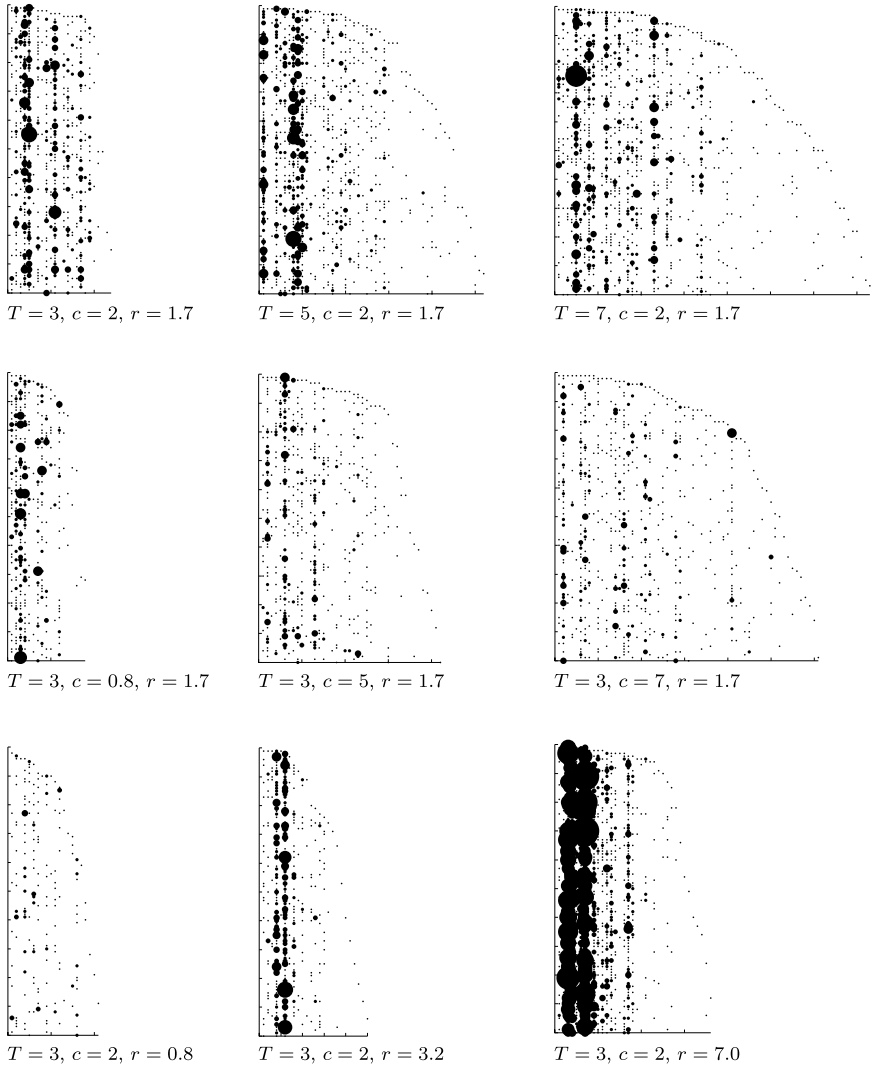

Figure 1. Simulated $\mathbb{Z}_+$-valued arrays from the NB-IBP. Dots are positive entries, the magnitudes of which determine the size of the dot. The total mass parameter T is varied along the top row; the concentration parameter c is varied along the middle row; the negative binomial parameter r is varied along the bottom row. See the text for a summary of how these parameters affect the expected number of features in total, features per row, and feature multiplicities.

introduce the IBP.) The relationship of the model to the values of T and c are similar to the characteristics described by Ghahramani, Griffiths and Sollich [7] for the IBP, with the parameter r providing flexibility with respect to the counts in the array. In particular, the total number of fea-

tures, κ, is Poisson distributed with mean $cT[\psi(c+nr)-\psi(c)]$, which increases with T, c, and r. From the NB-IBP, we know that the expected number of features for the first (and therefore, by exchangeability, every) row is T. Because of the ordering we have chosen here, the rows are not exchangeable, despite the sequence $X_{[n]}$ being exchangeable. (In contrast, a uniform random labeling W is row exchangeable and, conditioned on κ, column exchangeable.) Finally, note that the mean of the digamma(r, c) distribution exists for $c > 1$ and is given by

$$\frac{r}{(c-1)(\psi(r+c)-\psi(c))}, \tag{5.3}$$

which increases with r and decreases with c. This is the expected multiplicity of each feature for the first row, which again, by exchangeability, must hold for every row. We may therefore summarize the effects of changing each of these parameters (as we hold the others constant) as follows:

- Increasing the mass parameter T increases both the expected total number of features and the expected number of features per row, while leaving the expected multiplicities of the features unchanged.
- Increasing the concentration parameter c increases the expected total number of features and decreases the expected multiplicites of the features, while leaving the expected number of features per row unchanged.
- Increasing the parameter r increases both the expected total number of features and the expected multiplicities of the features, while leaving the expected number of features per row unchanged.

These effects can be seen in the first, second, and third rows of Figure 1, respectively. We note that r has a weak effect on the expected total number of features (seen in the third row of Figure 1), and c has a weak effect on the expected multiplicities of the features (seen in the second row of Figure 1). The model may therefore be effectively tuned with T and c determining the size and density of the array, and r determining the multiplicities. The most appropriate model depends on the application at hand, and in Section 5.3 we discuss how these parameters may be inferred from data.

5.2. Examples

Latent feature models with associated multiplicities and unbounded numbers of features have found several applications in Bayesian nonparametric statistics, and we now provide some examples. In these applications, the features represent latent objects or factors underlying a dataset comprised of n groups of measurements $y_1, \ldots, y_n$, where each group y_i is comprised of D_i measurements $y_i = (y_{i,1}, \ldots, y_{i,D_i})$. In particular, $W_{i,j}$ denotes the number of instances of object/factor j in group i.

These nonparametric latent feature representations lend themselves naturally to mixture models with an unbounded number of components. For example, consider a variant of the models by Sudderth *et al.* [22] and Titsias [26] for a dataset of n street camera images where the latent features are interpreted as object classes that may appear in the images, such as "building", "car", "road", etc. The count $W_{i,j}$ models the relative number of times object class j

appears in image i. For every $i \leq n$, image y_i consists of D_i local patches $y_{j,1}, \ldots, y_{j,D_i}$ detected in the image, which are (collections of) continuous variables representing, for example, color, hue, location in the image, etc. Let κ be the number of columns of W, that is, the number of features. The local patches in image i are modeled as conditionally i.i.d. draws from a mixture of $S_i = \sum_{j=1}^{\kappa} W_{i,j}$ Gaussian distributions, where $W_{i,j}$ of these components are associated with feature j. For $k = 1, 2, \ldots$, let $\Theta_i^{(j,k)} := (m_i^{(j,k)}, \Sigma_i^{(j,k)})$ denote the mean and covariance of the Gaussian components associated with feature j for image i. Let $z_{i,d} = (j,k)$ when $y_{i,d}$ is assigned to component $k \leq W_{i,j}$ associated with feature $j \leq \kappa$. Conditioned on $\Theta := (\Theta_i^{(j,k)})_{i \leq n, j \leq \kappa, k \leq W_{i,j}}$ and the assignments $z := (z_{i,d})_{i \leq n, d \leq D_i}$, the distribution of the measurements admits a conditional density

$$p(y|W, \Theta, z) = \prod_{i=1}^{n} \prod_{d=1}^{D_i} \mathcal{N}\big(y_{i,d}; m_i^{z_{i,d}}, \Sigma_i^{z_{i,d}}\big). \tag{5.4}$$

To share statistical strength across images, the parameters $\Theta_i^{(j,k)}$ are given a hierarchical Bayesian prior:

$$\Theta_i^{(j,k)} | \Theta^{(j)} \overset{\text{i.i.d.}}{\sim} \nu\big(\Theta^{(j)}\big) \qquad \text{for every } i \text{ and } k, \tag{5.5}$$

$$\Theta^{(j)} \overset{\text{i.i.d.}}{\sim} \nu_0 \qquad \text{for every } j. \tag{5.6}$$

A typical choice for $\nu(\cdot)$ is the family of Gaussian–inverse-Wishart distributions with feature-specific parameters $\Theta^{(j)}$ drawn i.i.d. from a distribution ν_0. Finally, for every image $i \leq n$, conditioned on W, the assignment variables $z_{i,1}, \ldots, z_{i,D_i}$ for the local patches in image n are assumed to form a multivariate Pólya urn scheme, arising from repeated draws from a Dirichlet-distributed probability vector over $\{(j,k) : j \leq \kappa, k \leq W_{i,j}\}$. The parameters for the Dirichlet distributions are tied in a similar fashion to Θ. The interpretation here is that local patch d in image i is assigned to one of the S_i instances of the latent objects appearing in the image. The number of object instances to which a patch may be assigned is specific to the image, but components across all images that correspond to the same feature will be similar.

Latent feature representations are also a natural choice for factor analysis models. Canny [5] and Zhou *et al.* [28] proposed models for text documents in terms of latent features representing *topics*. More carefully, let $y_{i,v}$ be the number of occurrences of word v in document i. Conditioned on W and a collection of non-negative topic-word weights $\Theta := (\theta_{j,v})_{j \leq \kappa, v \leq V}$, the word counts are assumed to be conditionally i.i.d. and

$$y_{i,v} | W, \Theta \sim \text{Poisson}\left(\sum_{j=1}^{\kappa} W_{i,j} \theta_{j,v}\right). \tag{5.7}$$

In other words, the expected number of occurrences of word v in document i is a linear sum of a small number of weighted factors. The features here are interpreted as topics: words v such that $\theta_{j,v}$ is large are likely to appear many times. There are a total of κ topics that are shared across the documents. The topic-word weights Θ are typically chosen to be i.i.d. Gamma random variates,

although there may be reason to prefer priors with dependency enforcing further sparsity. This general setup has been applied to other types of data including, for example, recommendations [8], where $y_{i,v}$ represents the rating a Netflix user i assigns to a film v.

5.3. Conditional distributions

Let W be a uniform random labelling of a NB-IBP as described in Section 5.1. In the applications described above, computing the posterior distribution of W is the first step towards most other inferential goals. Existing inference schemes use stick-breaking representations, that is, they represent (a truncation of) the beta process underlying W. This approach has some advantages, including that the entries of W are then conditionally independent negative binomial random variables. On the other hand, the random variables representing the truncated beta process, as well as the truncation level itself, must be marginalized away using auxiliary variable methods or other techniques [3,19,23,28]. Here, we take advantage of the structure of the NB-IBP and do not represent the beta process. The result is a set of Markov (proposal) kernels analogous to those originally derived for the IBP [7,10].

The models described in Section 5.2 associate every feature with a latent parameter. Therefore, conditioned on the number of columns κ, let $\Theta = (\theta_1, \ldots, \theta_\kappa)$ be an i.i.d. sequence drawn from some non-atomic distribution ν_Θ, and assume that the data y admits a conditional density $p(y|W, \Theta)$. We will associate the jth column of W with Θ_j, and so the pair (W, Θ) can be seen as an alternative representation for an exchangeable sequence $X_{[n]}$ of beta negative binomial processes. By Bayes' rule, the posterior distributions admits a conditional density

$$p(W, \Theta|y) \propto p(y|W, \Theta) \times p(W, \Theta), \tag{5.8}$$

where $p(W, \Theta)$ is a density for the joint distribution of (W, Θ). We describe two Markov kernels that leave this distribution invariant. Combined, these kernels give a Markov chain Monte Carlo (MCMC) inference procedure for the desired posterior.

The first kernel resamples individual elements $W_{i,j}$, conditioned on the remaining elements of the array (collectively denoted by $W_{-(i,j)}$), the data y, and the parameters Θ. By Bayes' rule, and the independence of Θ and W given κ, we have

$$\begin{aligned}&\mathbb{P}\{W_{i,j} = z|y, W_{-(i,j)}, \Theta\}\\ &\quad\propto p\big(y|\{W_{i,j} = z\}, W_{-(i,j)}, \Theta\big) \times \mathbb{P}\{W_{i,j} = z|W_{-(i,j)}\}.\end{aligned} \tag{5.9}$$

Recall that the array W is row-exchangeable, and so, in the language of the NB-IBP, we may associate the ith row with the final customer at the buffet. The count $W_{i,j}$ is the number of servings the customer takes of dish j, which has been served $S_j^{(-i)} := \sum_{i' \neq i} W_{i',j}$ times previously. When $S_j^{(-i)} > 0$, we have

$$W_{i,j}|W_{-(i,j)} \sim \text{beta-NB}\big(r, S_j^{(-i)}, c + (n-1)r\big). \tag{5.10}$$

Therefore, we can simulate from the unnormalized, unbounded discrete distribution in equation (5.9) using equation (5.10) as a Metropolis–Hastings proposal, or we could use inverse

transform sampling where the normalization constant is approximated by an importance sampling estimate.

Following Meeds *et al.* [17], the second kernel resamples the number, positions, and values of those *singleton* columns j' such that $S_{j'}^{(-i)} = 0$. Simultaneously, we propose a corresponding change to the sequence of latent parameters Θ, preserving the relative ordering with the columns of W. This corresponding change to Θ cancels out the effect of the $\kappa!$ term appearing in the p.m.f. of the array W. Let J_i be the number of singleton columns, that is, let

$$J_i = \#\big\{j \leq \kappa \colon W_{i,j} > 0 \text{ and } S_j^{(-i)} = 0\big\}, \tag{5.11}$$

which we note may be equal to zero. Because we are treating the customer associated with row i as the final customer at the buffet, J_i may be interpreted as the number of new dishes sampled by the final customer, in which case, we know that

$$J_i \sim \text{Poisson}\big(cT\big[\psi(c+nr) - \psi\big(c+(n-1)r\big)\big]\big). \tag{5.12}$$

We therefore propose a new array W^* by removing the J_i singleton columns from the array and insert J_i^* new singleton columns at positions drawn uniformly at random, where J_i^* is sampled from the (marginal) distribution of J_i given in equation (5.12). Like those columns that were removed, each new column has exactly one non-zero entry in the ith row: We draw each non-zero entry independently and identically from a digamma$(r, c+(n-1)r)$ distribution, which matches the distribution of the number of servings the last customer takes of each newly sampled dish.

Finally, we form a new sequence of latent parameters Θ^* by removing those entries from Θ associated with the J_i columns that were removed from W and inserting J_i^* new entries, drawn i.i.d. from ν_Θ, at the same locations corresponding to the J_i^* newly introduced columns. Let $\kappa^* := \kappa - J_i + J_i^*$, and note that there were $\binom{\kappa^*}{J_i^*}$ possible ways to insert the new columns. Therefore, the proposal density is

$$\begin{aligned} q\big(W^*, \Theta^* | W, \Theta\big) = \binom{\kappa^*}{J_i^*}^{-1} &\text{Poisson}\big(J_i^*; cT\big[\psi(c+nr) - \psi\big(c+(n-1)r\big)\big]\big) \\ &\times \prod_{j \leq \kappa^*} \text{digamma}\big(W_{i,j}^*; r, c+(n-1)r\big) \prod_{\theta \in \Theta^* \setminus \Theta} \nu_\Theta(\theta). \end{aligned} \tag{5.13}$$

With manipulations similar to those in the proof of Theorem 1.2, it is straightforward to show that a Metropolis–Hastings kernel accepts a proposal (W^*, Θ^*) with probability $\min\{1, \alpha^*\}$, where

$$\alpha^* = \frac{p(y|W^*, \Theta^*)}{p(y|W, \Theta)}. \tag{5.14}$$

Combined with appropriate Metropolis–Hastings moves that shuffle the columns of W and resample the latent parameters Θ, we obtain a Markov chain whose stationary distribution is the conditional distribution of W and Θ given the data y.

Another benefit of the characterization of the distribution of W in (5.1) is that numerically integrating over the real-valued concentration, mass, and negative binomial parameters c, T, and

r, respectively, are straightforward with techniques such as slice sampling [18]. In the particular case when T is given a gamma prior distribution, say $T \sim \text{gamma}(\alpha, \beta)$ for some positive parameters α and β, the conditional distribution again falls into the class of gamma distributions. In particular, the conditional density is

$$p(T|W,\kappa) \propto T^{\alpha+\kappa-1} \exp\bigl(-cT\bigl[\psi(c+nr)-\psi(c)\bigr]-\beta T\bigr) \tag{5.15}$$

$$\propto \text{gamma}\bigl(T;\alpha+\kappa,\beta+cT\bigl[\psi(c+nr)-\psi(c)\bigr]\bigr). \tag{5.16}$$

Acknowledgements

We thank Mingyuan Zhou for helpful feedback and for pointing out the relation of our work to that of Sibuya [21]. We also thank Yarin Gal and anonymous reviewers for feedback on drafts. This research was carried out while C. Heaukulani was supported by the Stephen Thomas studentship at Queens' College, Cambridge, with funding also from the Cambridge Trusts, and while D.M. Roy was a research fellow of Emmanuel College, Cambridge, with funding also from a Newton International Fellowship through the Royal Society.

References

[1] Barndorff-Nielsen, O. and Yeo, G.F. (1969). Negative binomial processes. *J. Appl. Probab.* **6** 633–647. MR0260001

[2] Broderick, T., Jordan, M.I. and Pitman, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statist. Sci.* **28** 289–312. MR3135534

[3] Broderick, T., Mackey, L., Paisley, J. and Jordan, M.I. (2014). Combinatorial clustering and the beta-negative binomial process. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 290–306. Special issue on Bayesian nonparametrics.

[4] Broderick, T., Pitman, J. and Jordan, M.I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836. MR3150470

[5] Canny, J. (2004). Gap: A factor model for discrete data. In *Proceedings of the* 27*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *Sheffield*, *United Kingdom*.

[6] Devroye, L. (1992). Random variate generation for the digamma and trigamma distributions. *J. Stat. Comput. Simul.* **43** 197–216. MR1389440

[7] Ghahramani, Z., Griffiths, T.L. and Sollich, P. (2007). Bayesian nonparametric latent feature models. In *Bayesian Statistics* 8. *Oxford Sci. Publ.* 201–226. Oxford: Oxford Univ. Press. MR2433194

[8] Gopalan, P., Ruiz, F.J.R., Ranganath, R. and Blei, D.M. (2014). Bayesian nonparametric Poisson factorization for recommendation systems. In *Proceedings of the* 17*th International Conference on Artificial Intelligence and Statistics*, *Reykjavik*, *Iceland*.

[9] Grégoire, G. (1984). Negative binomial distributions for point processes. *Stochastic Process. Appl.* **16** 179–188. MR0724064

[10] Griffiths, T.L. and Ghahramani, Z. (2006). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems* 19, *Vancouver*, *Canada*.

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

[12] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1876169
[13] Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27** 562–588. MR1714717
[14] Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185
[15] Kozubowski, T.J. and Podgórski, K. (2009). Distributional properties of the negative binomial Lévy process. *Probab. Math. Statist.* **29** 43–71. MR2553000
[16] Lo, A.Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. Verw. Gebiete* **59** 55–66. MR0643788
[17] Meeds, E., Ghahramani, Z., Neal, R.M. and Roweis, S.T. (2007). Modeling dyadic data with binary latent factors. In *Advances in Neural Information Processing Systems* 20, *Vancouver, Canada*.
[18] Neal, R.M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. With discussions and a rejoinder by the author. MR1994729
[19] Paisley, J., Zaas, A., Woods, C.W., Ginsburg, G.S. and Carin, L. (2010). A stick-breaking construction of the beta process. In *Proceedings of the* 27*th International Conference on Machine Learning*, *Haifa, Israel*.
[20] Roy, D.M. (2014). The continuum-of-urns scheme, generalized beta and Indian buffet processes, and hierarchies thereof. Preprint. Available at arXiv:1501.00208.
[21] Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Ann. Inst. Statist. Math.* **31** 373–390. MR0574816
[22] Sudderth, E.B., Torralba, A., Freeman, W.T. and Willsky, A.S. (2005). Describing visual scenes using transformed Dirichlet processes. In *Advances in Neural Information Processing Systems* 18, *Vancouver, Canada*.
[23] Teh, Y.W., Görür, D. and Ghahramani, Z. (2007). Stick-breaking construction for the Indian buffet process. In *Proceedings of the* 11*th International Conference on Artificial Intelligence and Statistics*, *San Juan, Puerto Rico*.
[24] Thibaux, R. and Jordan, M.I. (2007). Hierarchical beta processes and the Indian buffet process. In *Proceedings of the* 11*th International Conference on Artificial Intelligence and Statistics*, *San Juan, Puerto Rico*.
[25] Thibaux, R.J. (2008). Nonparametric Bayesian models for machine learning. Ph.D. thesis, EECS Department, Univ. California, Berkeley. MR2713095
[26] Titsias, M. (2007). The infinite gamma-Poisson feature model. In *Advances in Neural Information Processing Systems* 20.
[27] Wolpert, R.L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267. MR1649114
[28] Zhou, M., Hannah, L., Dunson, D. and Carin, L. (2012). Beta-negative binomial process and Poisson factor analysis. In *Proceedings of the* 29*th International Conference on Machine Learning*, *Edinburgh, United Kingdom*.
[29] Zhou, M., Madrid, O. and Scott, J.G. (2014). Priors for random count matrices derived from a family of negative binomial processes. Preprint. Available at arXiv:1404.3331v2.

Received June 2014 and revised March 2015