The Annals of Statistics
2017, Vol. 45, No. 5, 2016–2045
DOI: 10.1214/16-AOS1517
© Institute of Mathematical Statistics, 2017

BAYESIAN POISSON CALCULUS FOR LATENT FEATURE MODELING VIA GENERALIZED INDIAN BUFFET PROCESS PRIORS

BY LANCELOT F. JAMES[1]

Hong Kong University of Science and Technology

Statistical latent feature models, such as latent factor models, are models where each observation is associated with a vector of latent features. A general problem is how to select the number/types of features, and related quantities. In Bayesian statistical machine learning, one seeks (nonparametric) models where one can learn such quantities in the presence of observed data. The Indian Buffet Process (IBP), devised by Griffiths and Ghahramani (2005), generates a (sparse) latent binary matrix with columns representing a potentially unbounded number of features and where each row corresponds to an individual or object. Its generative scheme is cast in terms of customers entering sequentially an Indian Buffet restaurant and selecting previously sampled dishes as well as new dishes. Dishes correspond to latent features shared by individuals. The IBP has been applied to a wide range of statistical problems. Recent works have demonstrated the utility of generalizations to nonbinary matrices. The purpose of this work is to describe a unified mechanism for construction, Bayesian analysis, and practical sampling of broad generalizations of the IBP that generate (sparse) matrices with general entries. An adaptation of the Poisson partition calculus is employed to handle the complexities, including combinatorial aspects, of these models. Our work reveals a spike and slab characterization, and also presents a general framework for multivariate extensions. We close by highlighting a multivariate IBP with condiments, and the role of a stable-Beta Dirichlet multivariate prior.

1. Introduction. Statistical models involving latent random variables play a fundamental role in numerous applications across a wide variety of disciplines. Parametric latent models often impose various constraints that can be violated when faced with new observations. While model selection is a standard problem, the need for more flexible models is particularly true for modern data sets such as those arising in genetics or internet based data for social networks, etc. For example, in a mixture model one may wish to specify the number, say D, of components, corresponding to classes or groups. This may be infeasible in dynamic environments such as the internet, where new groups (websites) (for instance, social

Received April 2015; revised November 2015.

[1]Supported in part by the grant RGC-HKUST 601712 of the HKSAR.

MSC2010 subject classifications. Primary 60C05, 60G09; secondary 60G57, 60E99.

Key words and phrases. Bayesian statistical machine learning, Indian buffet process, nonparametric latent feature models, Poisson process calculus, spike and slab priors.

networks) are frequently being formed. As another example, in an online recommendation system one may want to specify the dimension, again call this D, of the feature vector corresponding to relevant attributes of users. It has been noted that it is particularly hard to realistically pre-choose features for such data sets which focus on attributes of individuals rather than objects, as naturally the potential features of human beings are unlimited.

Bayesian Machine Learning (ML), takes the position that rather than imposing rigid parametric constraints one can *learn* the structural features of these models through observations. The main task is to first specify a suitable probabilistic model (a nonparametric prior distribution) which explains, perhaps naively, how the data arises. That is to say a description of the generative process. Learning occurs in the presence of observed data and is essentially equivalent to the notion of updating the prior distribution to obtain the posterior distribution, and hence to obtain better predictive models using observed data. In the examples mentioned above, one seeks Bayesian nonparametric models that allow one to learn respectively the number of classes or the number of features by replacing models with fixed D with one that has potentially an unbounded number of classes or features. Naturally, other information such as class membership, size, and values associated with such classes are desired. Similarly, one wants to know statistics such as feature labels and values and the most common features shared by observations. In the parlance of [12, 15, 39], mixture models are examples of latent class models. *Latent class models* are models where each observation is associated with one class or group. In the presence of n observations, a natural way of assigning groups is to think of priors on random partitions of the integers $\{1, \ldots, n\}$. *Latent feature models*, with examples described above and in [12, 15, 39], are statistical models where each observation is associated with a possibly infinite, albeit sparse, vector of features. Broadly speaking Bayesian nonparametric latent feature models employ prior distributions over (sparse) matrices, with an infinite number of columns representing feature labels that are learned, and corresponding latent entries. Each row say $i = 1, \ldots, M$, represents the features and their values possessed by an individual or object i. Note that since there are a number of ways that one can use these processes to record entries in a matrix, it is perhaps more formal to say that these schemes induce priors over equivalence classes of matrices. Structurally, combinatorially, latent class models and latent feature models require different constructions for tractable priors/generative processes. A thorough discussion about these differences in relation to Chinese restaurant processes and basic Indian Buffet models that generate random matrices with binary entries is given in [5].

1.1. *Objectives and outline.* Our interest in this paper is to present a unified approach for construction, implementation, and (posterior) analysis of Bayesian nonparametric latent feature models which can be seen as broad generalizations of the Indian Buffet Process (IBP). Specifically, Sections 2 to 4 will focus on models that generate nonbinary matrix entries. This covers, and substantially extends results, for the known cases involving Bernoulli, Poisson and negative-binomial distributed entries. We, for the first time show how to handle entries based on continuous distributions. Section 5 will show how to extend IBP models to a multivariate setting. An adaptation of the Poisson Partition Calculus (PPC), developed by the author in [23, 25] suitable for nonparametric *Latent Class* models, is employed to handle the otherwise challenging and not well understood infinite-dimensional and combinatorial aspects of the *Latent Feature* models we shall discuss. Distinct from the few existing statistical analyses in the literature, we (1) treat the case where entries are general random variables. (2) Starting in Section 2.1, we highlight a dependent spike and slab characterization which reveals feature specific Bayesian parametric (spike/slab) models based on improper priors and demonstrates how the matrix values affect feature selection. (3) In Section 3.1, relevant to the description of marginal processes and more tractable descriptions of generative processes, we introduce infinite sequences of random pairs $((H_i, X_i))$ which facilitate wider practical implementation. (4) We close by constructing and analyzing general classes of multivariate processes. In Sections 5.2 and 5.3, we describe an Indian buffet process with condiments, and the role of a stable-Beta Dirichlet multivariate prior. Our work allows the user to focus on the practical implementation and development of flexible models for a wide range of statistical applications, rather than being limited by complexities generally arising from the use of discrete random measures, which naturally involve nontrivial combinatorial components. We first describe details for the original IBP process.

1.2. *The Indian Buffet Process* (*IBP*). The basic Indian buffet process, say IBP(θ) process, for $\theta > 0$, was ingeniously formulated by Griffiths and Ghahramani [12, 15, 16] whereby a (sparse) random binary matrix is formed with $(\tilde{\omega}_k, k = 1, \ldots)$ distinct features or attributes labeling an unbounded number of columns, and M rows, where each row i represents the attributes/features/preferences possessed by a single individual (or object) by entering 1 in the (i, k) entry if the feature $\tilde{\omega}_k$ is possessed and 0, otherwise. The matrix, with random binary entries $(b_{i,k})$, naturally indicates what features are shared among individuals. The generative process to describe this is cast in terms of individual customers sequentially entering an Indian buffet restaurant whereby the first customer selects a Poisson(θ) number of dishes. Subsequently, after M customers have chosen K distinct dishes, customer $M + 1$ chooses one of the K already sampled dishes (indicating that customer shares features already exhibited by the M previous customers) according to the most popular dishes already sampled, specifically with probability $m_\ell/(M + 1)$, where $m_\ell := m_{\ell,M}$ denotes the number of the M customers who have already sampled dish ℓ. (Note here that the probabilities do not necessarily add to 1.) Additionally, the customer chooses new dishes according to a Poisson($\theta/(M + 1)$) distribution. The precise label for a feature is drawn from a common i.i.d. distribution B_0 over some abstract Polish space Ω. In the generative process, a new feature label is drawn when a new dish is selected, leading to $(\omega_1, \ldots, \omega_K)$ distinct feature labels for the first M customers. It is important to note that Griffiths and Ghahramani derived this scheme based on a limiting argument where one first starts with an M by D binary matrix with conditionally independent entries $(b_{i,k})$ where, for each fixed (i,k), $b_{i,k}|p_k$ are i.i.d. Bernoulli(p_k) random variables and the p_k, for $k = 1, \ldots, D$, are modeled as i.i.d. Beta$(\theta/D, 1)$ variables. The IBP(θ) scheme arises by taking the limit as $D \to \infty$. Note a two-parameter extension of the IBP is described in [12].

A key insight was made by Thibaux and Jordan [50], which connected this generative process with a formal Bayesian nonparametric framework where $Z_1, \ldots, Z_M|\mu$ are modeled as i.i.d. (latent) Bernoulli processes with base measure μ that is selected to have a modified version of Hjort's [18] Beta process prior distribution where for the IBP the atoms $(\tilde{\omega}_k)$ of μ are i.i.d. B_0. The generative process is given by the distribution of $Z_{M+1}|Z_1, \ldots, Z_M$. In other words, the basic IBP(θ) generates a random matrix with entries $(b_{i,k})$, where conditioned on the collection (p_k), $b_{i,k}$ are independent Bernoulli (p_k) variables. The (p_k) are the points of a Poisson random measure with mean intensity $\rho(s) = \theta s^{-1}\mathbb{I}_{\{0<s<1\}}$. It follows that, conditional on μ, one can represent each $Z_i = \sum_{k=1}^{\infty} b_{i,k}\delta_{\tilde{\omega}_k}$ and the Beta process $\mu = \sum_{k=1}^{\infty} p_k\delta_{\tilde{\omega}_k}$. Thibaux and Jordan use the conjugacy result of Hjort [18], as described in Kim [31], to obtain the posterior distribution of $\mu|Z_1, \ldots, Z_M$ in the case where μ is a two-parameter Beta process with $\rho(s) = \theta s^{-1}(1-s)^{\beta-1}\mathbb{I}_{\{0<s<1\}}$, for $\beta > 0$. This then leads to a description of $Z_{M+1}|Z_1, \ldots, Z_M$. Teh and Görür [48] describe an important three parameter extension of the Beta process, which they call a stable-Beta process, specified by

$$\rho_{\alpha,\beta}(s) = \theta s^{-\alpha-1}(1-s)^{\beta+\alpha-1}\mathbb{I}_{\{0<s<1\}}, \tag{1.1}$$

for $\beta > -\alpha$ and $0 \le \alpha < 1$. Here, we will say μ is a stable-Beta process with parameters $(\alpha, \beta; \theta)$. Unlike the Beta process, this extension allows for power law behavior in terms of the total number of dishes (features) tried by M customers and the number of customers trying each dish. Building on the work of [48, 50] provide analogous posterior results for this case, using the results of [31].

REMARK 1.1. The Beta processes, with $B_0((0,\infty)) = \infty$, was developed in Hjort [18] as a prior for cumulative hazard rates arising in nonparametric Bayesian survival analysis. See also Kim [31] and the related work of Doksum [10] on neutral to the right processes.

REMARK 1.2. As was noted, the original IBP(θ) process is defined by (p_k) which are the points of a Poisson random measure with mean intensity $\rho(s) = \theta s^{-1}\mathbb{I}_{\{0<s<1\}}$. While the corresponding process μ can be viewed as a special case of a Beta process, it has a longer history in regards to its connection to the scale invariant process and Dickman distribution as can be found in [2] and references therein. Some aspects of this, as it relates to the IBP, are discussed in [28].

REMARK 1.3. The use of the stable-Beta process in [48], similar to the role of the Pitman–Yor process [20, 43, 46], illustrates the need for more general choices of ρ which do not lead to conjugate models. A more general version of the stable-Beta process, known as an extended beta process, was first discussed by Kim and Lee [33] in a survival analysis context. See [13, 26] for more choices for ρ.

1.3. *Statistical utility.* The IBP, and its generalizations, represent an exciting class of models well suited to handle high dimensional problems now common in this information age. One main function is to allow for models with potentially an unbounded number of features thus alleviating issues related to model selection. As mentioned in [12, 15, 16], the IBP can be used in a variety of ways when viewed as a common generator for a potentially vast number of latent feature model whereby the observed variables, say $\mathbf{Y}$, have conditional likelihoods of the general form $g(\mathbf{Y}|\mathbf{Z})$, where $\mathbf{Z} := (Z_1, \dots, Z_M)$. MCMC can then be applied utilizing the explicit form of the distribution of $Z_{M+1}|Z_1, \dots, Z_M$, say $p(Z_{M+1}|Z_1, \dots, Z_M)$. For instance, one may calculate the distribution of $Z_{M+1}|\mathbf{Y}, Z_1, \dots, Z_M$ which is proportional to $g(\mathbf{Y}|\mathbf{Z})p(Z_{M+1}|Z_1, \dots, Z_M)$. A model incorporating μ, requires the calculation of $\mu|\mathbf{Y}$ using the distribution of $\mu|\mathbf{Z}$ and $\mathbf{Z}|\mathbf{Y}$. Ghahramani et al. [12] and [15, 16] note applications for choice models, modeling protein interactions, independent components analysis and sparse factor analysis, among others.

A particular class of latent feature models are latent factor models whereby each observation, say $\mathbf{Y}_i := (Y_{i,1}, \dots, Y_{i,q})$, for $i = 1, \dots, M$ is postulated to arise from the following model:

$$Y_{i,j} = \sum_{k=1}^{D} A_{i,k}\eta_{j,k} + \varepsilon_{i,j},$$

where, for each i, $(A_{i,1}, \dots, A_{i,D})$ is a vector of the values of D latent features associated with $\mathbf{Y}_i$, $\{\eta_{j,k}\}$ is a matrix of unknown factor loadings and $\varepsilon_{i,j}$ are modeled typically as zero mean normal random variables. One of the issues here is the choice of D. In order to alleviate this, Knowles and Ghahramani [34] (see also [41, 47]) propose a nonparametric Bayesian extension of factor analysis by using the IBP to induce a sparse matrix with entries

$$A_{i,k} := b_{i,k}A'_{i,k} \tag{1.2}$$

for $i = 1, \dots, M$ and $k = 1, \dots, \infty$. The $A'_{i,k}$ are taken to be independent normal variables with mean 0 and having common variance for each fixed k. As another example, Miller, Griffiths and Jordan [39] describe a *nonarametric latent feature relational model* which in its simplest form can be described as follows: for individuals i and j, $Y_{i,j}$ is a binary link which is 1 if there is a link (relation) between i and j and 0 otherwise. Using the IBP, one can practically describe the distribution of $Y_{i,j}$ given $\mathbf{Z}$ and a latent weight matrix $(A'_{k,k'})$ as

$$\mathbb{P}(Y_{i,j} = 1|\mathbf{Z}, (A'_{k,k'})) = \frac{1}{1 + \exp(-\sum_{k,k'} b_{i,k}b_{j,k'}A'_{k.k'})}.$$

Note one may replace the sigmoidal function $1/(1+\mathrm{e}^{-x})$ with a more general link function. The expression above only depends on the nonzero entries corresponding to the cases where $b_{i,k}=1$ and $b_{j,k'}=1$, indicating that individuals i and j possess features k and k', respectively. $\mathbf{Z}$ is determined by the IBP which again allows one to learn the number of features rather than facing the impractical problem of preselecting the number of features. The relational models described in Hoff, Raftery and Handcock [19] are a precursor to the work of [39]. Other applications and aspects of the IBP are articulated in, for instance, [5, 7, 11, 12, 14–16, 30, 42, 48, 50], and for placement in a wider context, see [40]. See [4] for a generalization of the IBP with asymptotic results. Hybrid versions of the IBP for topic modeling have also been proposed in [1, 52].

2. Nonbinary generalizations-spike and slab IBP priors. The IBP generates binary random matrices $(b_{i,k})$ based on the usage of conditionally independent Bernoulli random variables. Generally latent feature models may require the generation of (sparse) random matrices with entries indicating feature values or as a special case repeated usage of features. This suggests the need for generalizations of the IBP to nonparametric processes which generate (sparse) matrices with nonbinary entries. Of note in the current literature are models employing Poisson and negative-binomial random variables as described in, for instance, [6, 17, 49–51, 53, 54]. These are generally coupled with gamma and beta process priors, respectively. These works demonstrate the desirability of IBP-type models which allow for more flexible latent feature values and discuss applications to, for instance, Poisson matrix factorization and admixture modeling. Using the theory of marked Poisson point processes, we describe the generalization of the IBP that includes the models already mentioned above. Simply replace the $((b_{i,k}, p_k))$ with more general variables $((A_{i,k}, \tau_k))$ where conditional on (τ_k), the collection $(A_{i,k})$ are independent random variables with, for each fixed k, distribution denoted as $G_A(da|\tau_k)$. The collection (τ_k) are the points of a Poisson random measure with more general Lévy density $\rho(s|\omega)$, not restricted to [0, 1] and possibly depending on ω, which reflects dependence on the distribution of the $(\tilde{\omega}_k)$, B_0, on a space Ω. Formally, we choose ρ to satisfy $\int_0^\infty \min(s,1)\rho(s|\omega)\,ds<\infty$. Since B_0 is a finite measure, this condition ensures that the total mass $\mu(\Omega)$ is an infinitely divisible random variable that is finite almost surely (see page 5 of [45]). However, we allow both infinite activity models corresponding to $\int_0^\infty \rho(s|\omega)\,ds=\infty$ and finite activity/compound models where $\int_0^\infty \rho(s|\omega)\,ds=c(\omega)<\infty$. In the latter case, $\rho(s|\omega)/c(\omega)$ is a proper density. For each i, $((A_{i,k}), \tau_k, \tilde{\omega}_k)$ are the points of a Poisson random measure with mean intensity, $G_A(da|s)\rho(s|\omega)\,ds\,B_0(d\omega)$. Furthermore, we require that $A_{i,k}$ can take on the value zero with positive probability. So relative to $G_A(\cdot|s)$, write $\pi_A(s)=\mathbb{P}(A\neq 0|s)$, and hence $1-\pi_A(s)=\mathbb{P}(A=0|s)$. We require the condition that $\pi_A(s)$ satisfies

$$\varphi := \int_\Omega \int_0^\infty \pi_A(s)\rho(s|\omega)\,ds\,B_0(d\omega) < \infty. \tag{2.1}$$

REMARK 2.1. The condition (2.1) is not true for arbitrary $\pi_A(s)$. If for example one chooses $\pi_A(s) = (1 - \mathrm{e}^{-s})$, then $\varphi < \infty$. However, if $\pi_A(s) = \mathrm{e}^{-s}$, then generally $\varphi = \infty$, unless ρ is a finite measure. Condition (2.1) is satisfied if $\pi_A(s)$ behaves like a cumulative distribution function.

The general construction is now

$$Z_i = \sum_{k=1}^{\infty} A_{i,k}\delta_{\tilde{\omega}_k} \quad \text{and} \quad \mu = \sum_{k=1}^{\infty} \tau_k \delta_{\tilde{\omega}_k}. \tag{2.2}$$

Here, key to our exposition, it is important to note that μ can always be represented as $\mu(d\omega) = \int_0^\infty s N(ds, d\omega)$, where $N = \sum_{k=1}^\infty \delta_{\tau_k, \tilde{\omega}_k}$ is a Poisson random measure with mean intensity $\mathbb{E}[N(ds, d\omega)] = \rho(s|\omega)\, ds\, B_0(d\omega) := \nu(ds, d\omega)$.

DEFINITION 2.1. We say that N is PRM(ρB_0), or more generally PRM(ν) and also write $\mathcal{P}(dN|\nu)$, when discussing calculations. μ is by construction a completely random measure and we shall specify its law by saying μ is CRM(ρB_0), or CRM(ν). Using this, we shall say that $Z_1, \ldots, Z_M|\mu$ are i.i.d. IBP$(G_A|\mu)$, if they have the specifications in (2.2) and call the marginal distribution of Z_i IBP$(A, \rho B_0)$ or equivalently IBP(A, ν).

2.1. *Spike and slab viewpoint for feature selection.* Notice that one can always represent $A_{i,k} = b_{i,k} A'_{i,k}$, where $b_{i,k}$ are Bernoulli variables and $A'_{i,k}$ is a general random variable that is not necessarily independent of $b_{i,k}$. This evokes the notion of generalized spike and slab type prior distributions (see for instance Ishwaran and Rao [22]) within the different context of variable selection, where here we mean a two-point mixture model made of a general distribution $\tilde{G}_{A'}$ (a generalized slab), not taking mass at zero, and a degenerate distribution at 0 (a spike). That is,

$$G_A(dy|s) = \pi_A(s)\tilde{G}_{A'}(dy|s) + [1 - \pi_A(s)]\delta_0(dy).$$

The specification of $\rho(s|\omega)$ induces a collection of improper priors, on s, in a manner that shall be clarified by our analysis. Naturally, one has in the Bernoulli case that $A' = 1$. In the Poisson case of [51] where $A_{i,k}$ is Poisson$(r\lambda_k)$, it follows that $b_{i,k}|\lambda_k$ is Bernoulli$(1 - \mathrm{e}^{-r\lambda_k})$ and $A'_{i,k}|\lambda_k$ has (conditional) distribution $\tilde{G}_{A'}(\cdot|\lambda_k)$ specified by

$$\mathbb{P}(A'_{i,k} = a|\lambda_k) = \mathbb{P}(A_{i,k} = a|b_{i,k} = 1, \lambda_k) = \frac{r^a \lambda_k^a \mathrm{e}^{-r\lambda_k}}{a!(1 - \mathrm{e}^{-r\lambda_k})}, \tag{2.3}$$

for $a = 1, 2, \ldots$, and further depending on λ_k, where the Lévy density $\rho(\lambda_k|\omega_k)$ is a mixing measure. In general, and in contrast to (1.2), since the distributions of $(b_{i,k}, A'_{i,k})$ depend on common parameters where ρ acts as a mixing measure, the slab values, $(A'_{i,k})$, will play a direct role, within an IBP-type selection scheme, in terms of how future customers will select existing features/dishes. It is known for the Poisson and negative-binomial models appearing in the literature that larger entries, corresponding to popularity, increase the chance of selection. However, as we shall show, in general the effects are distribution specific and do not necessarily follow this simple rule. A more challenging example that we shall discuss in detail later and has not appeared in the literature, is where the $A'|\lambda$ is specified to be a Normal$(\eta, 1/\lambda)$ random variable and the spike is determined by $\pi_A(\lambda) = (1 - \mathrm{e}^{-\lambda})$, which is the same as in the Poisson-IBP model.

2.2. *Preliminaries on posterior distributions and sequential generative schemes.* Now that we have provided constructions for general (A, ρ), we now turn to the task of obtaining posterior quantities that hold for any such choices. Although our models are more general, they share many of the pertinent structures that arise in the Bernoulli case. As such, we follow quite closely the lucid discussion given in Teh and Görür [48], Sections 3.1 and 3.2. Specifically, as noted by those authors, practical implementation and understanding of these processes requires a tractable description of the posterior distribution of $\mu|Z_1, \dots, Z_M$, or $N|Z_1, \dots, Z_M$, and the predictive distribution of $Z_{M+1}|Z_1, \dots, Z_M$. The latter is crucial for describing the IBP-type generative schemes for selecting features. Since these are discrete processes, the description involves a significant combinatorial component related to shared features. Following [48], Section 3.1, let $(\omega_1, \dots, \omega_K)$ be the K unique atoms among the $(Z_1, \dots, Z_M)$, where ω_ℓ has been selected m_ℓ times. Notice that all relevant quantities depend on M, so, for instance, we could write $K = K_M$ and $m_\ell := m_{\ell,M}$, however, unless otherwise indicated we will generally suppress this additional notation. For the Bernoulli cases, the counts $(m_1, \dots, m_K)$ suffice, however, we will need to appropriately index the relevant $(A'_{i,\ell})$ variables. At the most basic level, this involves (random) sets indicating which individuals possess features $(\omega_1, \dots, \omega_K)$, that is $\tilde{\mathcal{B}} := \{\mathcal{B}_1, \dots, \mathcal{B}_K\}$ where, $\mathcal{B}_\ell := \{i : Z_i(\omega_\ell) \neq 0\}$, with respective sizes $m_\ell := |\mathcal{B}_\ell|$, for $\ell = 1, \dots, K$. As we have alluded to, the values of the nonzero $(A_{i,k})$ play a significant role in terms of how new customers will choose existing dishes/features, as such we introduce the *value sets* $\mathcal{A}_\ell = \{A'_{i,\ell} : i \in \mathcal{B}_\ell\}$. The methods in [23, 25], will be adjusted to accommodate these structures to provide a unified framework in the form of a Bayesian Poisson calculus for latent feature models. Furthermore, it is important to note that additional work is required to describe the general distribution of $Z_{M+1}|Z_1, \dots, Z_M$, in this general setting. This is done in Section 3.1.

2.2.1. *Some existing results.* As we mentioned previously, subject to some variations, the cases treated explicitly involve essentially three choices for G_A, that is Bernoulli, negative binomial and Poisson distributions. Explicit results for the Bernoulli and negative-binomial cases are limited to the choice of $\rho = \rho_{\alpha,\beta}$ given in (1.1), that is, for μ a stable-Beta process. We do note that for the Bernoulli case, one can perhaps deduce the form of results for general ρ by reading carefully the exposition in [48] and otherwise applying [31]. However, this is not true for the negative-binomial case, where current results are based on more involved specialized arguments. For the Poisson case, some results are given for only the choice of $\rho(s) = \theta s^{-1} \mathrm{e}^{-s}$, $s > 0$, which corresponds to the case where μ is a Gamma process. Hence, for instance, results are not known for the case of μ being a generalized gamma process which is not a conjugate prior. This choice of μ exhibits power law properties for the Poisson case similar to [48]. More details for the cases above can be found in [6, 17, 49–51, 53, 54]. We will discuss these cases for general ρ in Section 4. In addition, relevant to the Bernoulli case, there is a related class of models considered by [9], where explicit details for a generalized gamma prior has been given. See [25, 26] for the analysis of various models, in different contexts, employing Beta, generalized gamma and other processes.

3. General posterior and marginal distributions.

We now describe the formal pertinent posterior quantities for understanding and implementing the various generalized Indian buffet schemes. The results are obtained by an adapted version of the original partition based PPC to a *Poisson Latent Feature Calculus* that can now handle the appropriate combinatorial structure for latent feature models. Specifically, for our purposes here we will need two ingredients of the PPC that can be found as Propositions 2.1 and 2.2 in James [25], pages 6–8 (see also [23], pages 7–8). These represent direct extensions of updating mechanisms for the Dirichlet and gamma processes employed in [36–38] and also [21, 24]. Similar to the PPC, our adapted version allows one to obtain intricate combinatorial expressions without combinatorial arguments and does not rely on any particular specification of ρ. Formal details of these derivations are provided in the Appendix.

Now, define $\rho_j(s|\omega) = [1 - \pi_A(s)]^j \rho(s|\omega)$ for $j = 0, 1, 2, \ldots$. Hence, $\rho_0(s|\omega) := \rho(s|\omega)$. Define for $k = 1, 2, \ldots,$

$$\varphi_k := \int_\Omega \int_0^\infty \pi_A(s) \rho_{k-1}(s|\omega)\, ds\, B_0(d\omega). \tag{3.1}$$

Throughout, for each M, we use, $\nu_M(ds, d\omega) := \rho_M(s|\omega) B_0(d\omega)\, ds$, where $\rho_M(s|\omega) = [1 - \pi_A(s)]^M \rho(s|\omega)$. Note that ρ_M depends on π_A and so will differ over various models. Since it should be clear from context, we shall suppress this dependence within the notation.

PROPOSITION 3.1. *Let (J_ℓ, ω_ℓ) for $\ell = 1, \ldots, K$ denote the $K := K_M$ unique points among $((\tau_k, \tilde{\omega}_k))$ picked from the Poisson random measure N, based on the conditionally i.i.d. processes $(Z_1, \ldots, Z_M)$. Then one can write the joint distribution of $((Z_1, \ldots, Z_M), (J_1, \ldots, J_K))$, where $(J_1, \ldots, J_K)$ are the unique jumps picked from μ, with arguments $(s_1, \ldots, s_K)$,*

$$\left[\prod_{k=1}^{M} \mathrm{e}^{-\varphi_k}\right] \prod_{\ell=1}^{K} \mathrm{S}_{\mathrm{spike} \times \mathrm{slab}}(s_\ell | \mathcal{A}_\ell) \rho(s_\ell | \omega_\ell) B_0(d\omega_\ell), \tag{3.2}$$

where, for $\mathcal{A}_\ell = \{A'_{i,\ell} : i \in \mathcal{B}_\ell\}$,

$$\mathrm{S}_{\text{spike}\times\text{slab}}(s|\mathcal{A}_\ell) = \left[1 - \pi_A(s)\right]^{M-m_\ell} \pi_A^{m_\ell}(s) \prod_{i\in\mathcal{B}_\ell} \tilde{G}_{A'}(da_{i,\ell}|s). \tag{3.3}$$

Hence, applying Bayes rule, the $(J_\ell)|(Z_i)$ *are conditionally independent with density,*

$$\mathbb{P}_{\ell,M}(J_\ell \in ds) = \frac{\mathrm{S}_{\text{spike}\times\text{slab}}(s|\mathcal{A}_\ell)\rho(s|\omega_\ell)\,ds}{\int_0^\infty \mathrm{S}_{\text{spike}\times\text{slab}}(t|\mathcal{A}_\ell)\rho(t|\omega_\ell)\,dt} \tag{3.4}$$

and the joint marginal distribution of $(Z_1, \dots, Z_M)$, *denoted as* $\mathbb{P}(Z_1, \dots, Z_M)$, *is*

$$\left[\prod_{k=1}^{M} \mathrm{e}^{-\varphi_k}\right] \prod_{\ell=1}^{K} \left[\int_0^\infty \mathrm{S}_{\text{spike}\times\text{slab}}(t|\mathcal{A}_\ell)\rho(t|\omega_\ell)\,dt\right] B_0(d\omega_\ell). \tag{3.5}$$

REMARK 3.1. One may compare our general expression for $\mathbb{P}(Z_1, \dots, Z_M)$, with the special case of Teh and Görür [48], equation (10), in the Bernoulli setting. See also their ensuing remarks in regards to [15], equation (4). More details are given in Section 4.

We now describe the posterior distribution of N and μ given $(Z_1, \dots, Z_M)$.

THEOREM 3.1. *Suppose that* $Z_1, \dots, Z_M|\mu$ *are i.i.d.* $\mathrm{IBP}(G_A|\mu)$, μ *is* $\mathrm{CRM}(\rho B_0)$:

(i) *Then the posterior distribution of* $N|Z_1, \dots, Z_M$ *is equivalent to the distribution of* $N_M + \sum_{\ell=1}^{K} \delta_{J_\ell,\omega_\ell}$ *where* N_M *is* $\mathrm{PRM}(\rho_M B_0)$ *and the distribution of the conditionally independent* (J_ℓ) *is given by* (3.4).

(ii) *The posterior distribution of* $\mu|Z_1, \dots, Z_M$ *is equivalent to the distribution of* $\mu_M + \sum_{\ell=1}^{K} J_\ell \delta_{\omega_\ell}$ *where* μ_M *is* $\mathrm{CRM}(\rho_M B_0)$.

(iii) *Note that* $(\omega_1, \dots, \omega_K)$, *and* K, *appearing in the expressions above, are known fixed values from* $(Z_1, \dots, Z_M)$.

This immediately leads to the next result.

PROPOSITION 3.2. *Suppose that* $Z_1, \dots, Z_M, Z_{M+1}|\mu$ *are i.i.d.* $\mathrm{IBP}(G_A|\mu)$, *where* μ *is* $\mathrm{CRM}(\rho B_0)$. *Then results* (i) *or* (ii) *in Theorem* 3.1, *shows that the distribution of* $Z_{M+1}|Z_1, \dots, Z_M$ *is equivalent to the distribution of*

$$\tilde{Z}_{M+1} + \sum_{\ell=1}^{K} A_{M+1,\ell}\delta_{\omega_\ell},$$

where $\tilde{Z}_{M+1}$ *is* $\mathrm{IBP}(A, \rho_M B_0)$, *and each* $A_{M+1,\ell}|J_\ell = s$ *has distribution* $G_A(da|s)$, *and the marginal distribution of* J_ℓ *is specified by* (3.4). *That is to* *say* $A_{M+1,\ell}$ *has the distribution* $\int_0^\infty G_A(dy|s)\mathbb{P}_{\ell,M}(ds)$, *which can be expressed as*

$$\mathbb{E}[(1-\pi_A(J_\ell))]\delta_0(dy)+\mathbb{E}[\pi_A(J_\ell)]\frac{\int_0^\infty \pi_A(s)\tilde{G}_{A'}(dy|s)\mathbb{P}_{\ell,M}(ds)}{\mathbb{E}[\pi_A(J_\ell)]},$$

where $\pi_A(s)\mathbb{P}_{\ell,M}(ds)/\mathbb{E}[\pi_A(J_\ell)]$ *is the distribution of* $J_\ell|Z_1,\dots,Z_M$, $b_{M+1,\ell}=1$. (ω_ℓ) *are the already chosen features*.

3.1. *Describing the marginal distribution of* Z *via a method of decompositions*. The results above provide a general framework to describe generative processes analogous to the Indian buffet sequential latent feature selection scheme. What remains is to describe the marginal distributions of $Z_1 \sim \mathrm{IBP}(A,\rho B_0)$, for arbitrary (A,ρ). Substituting ρ with ρ_M then leads to the representation of $\tilde{Z}_{M+1}\sim \mathrm{IBP}(A,\rho_M B_0)$, which consists of features/atoms $(\tilde{\omega}_k)/(\omega_1,\dots,\omega_K)$ that have not been selected from $(Z_1,\dots,Z_M)$. That is to say, features not possessed by the first M customers. We now describe the marginal distribution of Z utilizing an infinite sequence of i.i.d. pairs of variables $((H_i,X_i))$.

PROPOSITION 3.3. *Let* $Z|\mu$ *be* $\mathrm{IBP}(G_A|\mu)$ *where* μ *is a* $\mathrm{CRM}(\rho B_0)$ *with* $\rho(s)$ *a homogeneous Lévy density such that* $\varphi:=\int_0^\infty \pi_A(s)\rho(s)\,ds<\infty$. *Then* Z *is said to have an* $\mathrm{IBP}(A,\rho B_0)$ *marginal distribution*, *where there exists a sequence of i.i.d. pairs* $((H_i,X_i))$, *which we refer to as an* $\mathrm{IBP}(A,\rho)$ *sequence*, *such that*

$$Z\stackrel{d}{=}\sum_{i=1}^{\xi(\varphi)}X_i\delta_{\tilde{\omega}_i},$$

where $\xi(\varphi)$ *is a Poisson random variable*, *with mean* φ, *independent of* $((H_i,X_i),\tilde{\omega}_i)$. *The random variables* $(\tilde{\omega}_i)$ *are i.i.d.* B_0. *The collection of pairs* $((H_i,X_i))$, *have the following distributional properties*:

(i) $X_i|H_i=s$ *has distribution*, *not depending on* ρ,

$$\tilde{G}_{A'}(da|s)=\frac{I_{\{a\neq 0\}}G_A(da|s)}{\pi_A(s)},$$

which is equivalent to the conditional (*slab*) *distribution of* $A|A\neq 0$ *under* $G_A(\cdot|s)$, *and* H_i *has marginal density*:

$$\mathbb{P}_{H_i}(ds)=\frac{\pi_A(s)\rho(s)\,ds}{\varphi}.$$

(ii) *The marginal distribution of* X_i *is equivalent to the distribution of* $A_i|b_i=1$,

$$\mathbb{P}(X_i\in da)=\int_0^\infty \tilde{G}_{A'}(da|s)\mathbb{P}_{H_i}(ds)=\frac{I_{\{a\neq 0\}}\int_0^\infty G_A(da|s)\rho(s)\,ds}{\varphi}$$

and the distribution of $H_i|X_i = a$ *is given by*

$$\mathbb{P}(H_i \in ds|X_i = a) = \frac{G_{A'}(da|s)\mathbb{P}_{H_i}(ds)}{\int_0^\infty G_{A'}(da|v)\mathbb{P}_{H_i}(dv)}.$$

(iii) *If* $\rho(s)$ *is replaced by* $\rho(s|\omega)$, *then replace* $\rho(s)$ *in the above expressions with* $\rho(s|\tilde{\omega}_i)$ *to obtain results for* $((H_i, X_i))|(\tilde{\omega}_i)$.

PROOF. Since $Z := Z_1 \stackrel{d}{=} \sum_{k=1}^{\infty} A_k \delta_{\tilde{\omega}_k}$ is composed of $(A_k, \tau_k, \tilde{\omega}_k)$ the theory of marked Poisson point processes tells us that the unconditional Laplace functional of $Z(f) = \sum_{k=1}^{\infty} A_k f(\omega_k)$ is given by $\mathbb{E}[\mathrm{e}^{-Z(f)}] = \mathrm{e}^{-\Phi(f)}$, where

$$\Phi(f) = \int_{\Omega}\int_{\mathbb{R}} (1 - \mathrm{e}^{-af(\omega)}) I_{\{a \neq 0\}} Q_A(da|\omega) B_0(d\omega), \tag{3.6}$$

where $Q_A(da|\omega) = [\int_0^\infty G_A(da|s)\rho(s|\omega)\, ds]$, and the distribution of the A_k is determined by the measure $I_{a\neq 0} Q_A(da|\omega)$. This is clear since at $a = 0$, $(1 - \mathrm{e}^{-af(\omega)})$ is zero. It follows from (2.1), that $\varphi(\omega) = \int_0^\infty \pi_A(s)\rho(s|\omega)\, ds < \infty$ which is the total mass of the measure $I_{a\neq 0} Q_A(da|\omega)$. This, along with the representation (3.6) establishes that Z is a compound Poisson process as described where $A_k|\tilde{\omega}_k$ are independent with distribution $\tilde{Q}_A(da|\tilde{\omega}_k) = I_{\{a\neq 0\}} Q_A(da|\tilde{\omega}_k)/\varphi(\tilde{\omega}_k)$. Augmenting this expression reveals a joint distribution proportional to $I_{\{a\neq 0\}} G_A(da|s)\rho(s|\omega)\, ds$, concluding the result. $\square$

3.2. *Generalized Indian Buffet Processes*: *The sequential generative process for* IBP$(A, \rho B_0)$. We can now use Theorem 3.1 and Propositions 3.2 and 3.3 to describe the sequential generative process for IBP$(A, \rho B_0)$, in the homogeneous case. That is to say how to sample from Z_1, and subsequently $Z_{M+1}|Z_1, \dots, Z_M$. Since under G_A every matrix entry value for a feature can take a wider range of values, one needs to give this a proper interpretation. This is mostly left to the reader, but here we shall naively say that a customer gives a scoring to that dish. At any rate, we use the basic IBP metaphor of people sequentially entering an Indian buffet restaurant. Set $\tilde{K}_i = \text{Poisson}(\varphi_i)$ for $i = 1, 2, \dots$, hence $K := K_M = \sum_{i=1}^{M} \tilde{K}_i$:

1. Customer 1 selects dishes according to a Poisson(φ_1) distribution and gives them scores according to $((H_i, X_i))$ following an IBP$(A, \rho B_0)$ sequence. This is done precisely as follows:

   (i) Draw a Poisson$(\varphi_1) = \tilde{K}_1$ number of variables.

   (ii) Draw $((\omega_1, H_1), \dots (\omega_{\tilde{K}_1}, H_{\tilde{K}_1}))$ i.i.d. from B_0 and the distribution for H_i, $\mathbb{P}_{H_i}(ds) = \pi_A(s)\rho(s)\, ds/\varphi_1$.

   (iii) Draw $X_i|H_i$ following $\tilde{G}_{A'}(da|H_i)$, for $i = 1, \dots, \tilde{K}_1$. If it is straightforward to sample directly from the marginal distribution of X_i, then one may bypass sampling H_i. These values may be set to $A'_{1,\ell}$ for $\ell = 1, \dots, \tilde{K}_1$.

2. After M customers have chosen collectively $\ell = 1, \dots, K$ distinct dishes, and assigned their scores $\{A'_{i,\ell} : i \in \mathcal{B}_\ell\}$, customer $M+1$ selects each dish ω_ℓ, $\ell = 1, \dots, K$, and assigns respective scores, according to the distribution of $A_{M+1,\ell}$, where $A_{M+1,\ell}|J_\ell$ has conditional distribution $G_A(da|J_\ell)$, in particular the chance that $A_{M+1,\ell}|J_\ell$ takes the value zero is $1 - \pi_A(J_\ell)$. The distribution of (J_ℓ) is specified in (3.4).
3. Customer $M+1$ also chooses and scores new dishes according to a IBP(A, $\rho_M B_0$) process $\tilde{Z}_{M+1}$. This follows the same scheme as customer 1 with $\rho_M(s) = [1 - \pi_A(s)]^M \rho(s)$ in place of ρ. Specifically, a Poisson(φ_{M+1}) $= \tilde{K}_{M+1}$ number of new dishes are chosen and scores are assigned according to $((H_i, X_i))$ following an IBP(A, ρ_M) sequence. For accounting purposes, one may index the new $\tilde{K}_{M+1}$ selected features and corresponding scores (X_i) as $(\omega_\ell, A'_{M+1,\ell})$ for $\ell = K_M + 1, \dots, K_{M+1}$.
4. Notice that $\mathbb{P}(\tilde{K}_k = 0) = \mathbb{P}(\xi(\varphi_k) = 0) = \mathrm{e}^{-\varphi_k}$, for $k = 1, 2, \dots$.

4. Examples. We now first present details for the cases where A is Bernoulli, Poisson and negative binomial, with respect to general choices for homogeneous ρ. We then present details for a normal model in Section 4.4. The choice of a generalized gamma prior is highlighted in the Poisson and normal cases.

4.1. Bernoulli(p). In the simplest case $A|p$ is Bernoulli(p), where $\pi_A(p) = p$ and $A' = 1$. It follows that $\mathrm{S}_{\mathrm{spike}\times\mathrm{slab}}(p|\mathcal{A}_\ell) = (1-p)^{M-m_\ell} p^{m_\ell}$. For generic ρ, J_ℓ has density proportional to $(1-p)^{M-m_\ell} p^{m_\ell} \rho(p)$ and $\varphi_k = \int_0^1 p(1-p)^{k-1} \rho(p)\, dp$. $\rho_M(p) = (1-p)^M \rho(p)$ which determines N_M, μ_M. For a previously chosen feature, ω_ℓ, the distribution of $A_{M+1,\ell}$ is Bernoulli($\mathbb{E}[J_\ell]$) where

$$\mathbb{E}[J_\ell] = \frac{\int_0^1 (1-p)^{M-m_\ell} p^{m_\ell+1} \rho(p)\, dp}{\int_0^1 (1-p)^{M-m_\ell} p^{m_\ell} \rho(p)\, dp},$$

which is the chance customer $M+1$ has the existing feature ω_ℓ shared by m_ℓ out of the M previous customers. The customer will also choose Poisson(φ_{M+1}) new dishes, where since $A' = 1$ the scores (X_i) are always 1. In order to see how to use our results to recover the known cases in the literature we look at the case of the stable-Beta process as in [48]. That is μ is a stable-Beta process with parameters $(\alpha, \beta; \theta)$, with $\rho(s) = \rho_{\alpha,\beta}(s)$ specified in (1.1). It follows that μ_M is also a stable-Beta process now with parameters $(\alpha, M+\beta; \theta)$, and the corresponding jumps are such that J_ℓ is a Beta($m_\ell - \alpha, M - m_\ell + \beta + \alpha$) random variable. Hence, $\mathbb{E}[J_\ell] = (m_\ell - \alpha)/(M + \beta)$.

4.2. Poisson($r\lambda$). We now look at the case where $A|\lambda$ is a Poisson($r\lambda$) random variable. When μ is specified to be a gamma process by taking $\rho(\lambda) = \theta\lambda^{-1}\mathrm{e}^{-r\lambda}$ this coincides with the Poisson-Gamma model considered by [49, 51]. This is related to a Poisson-Gamma model of Lo [36] where posterior conjugacy was established. Here, $\pi_A(\lambda) = (1 - \mathrm{e}^{-r\lambda})$ and the slab distribution of $A'|\lambda$, and hence $X_i|H_i = \lambda$, is specified by (2.3). It follows that

$$\mathrm{S}_{\text{spike}\times\text{slab}}(\lambda|\mathcal{A}_\ell) = \frac{r^{c_{\ell,M}}}{\prod_{i\in\mathcal{B}_\ell} A'_{i,\ell}!}\lambda^{c_{\ell,M}}\mathrm{e}^{-\lambda M r},$$

where $c_{\ell,M} = \sum_{i\in\mathcal{B}_\ell} A'_{i,\ell}$. For a general ρ, J_ℓ has density proportional to $\lambda^{c_{\ell,M}}\mathrm{e}^{-\lambda M r}\rho(\lambda)$. For sampling new dishes, it follows that

$$\varphi_k := \phi_{r,k} = \int_0^\infty (1 - \mathrm{e}^{-ry})\mathrm{e}^{-(k-1)ry}\rho(y)\,dy. \tag{4.1}$$

Additionally, $\rho_j(\lambda) = \mathrm{e}^{-jr\lambda}\rho(\lambda)$, $j = 0, 1, 2, \dots$. For a previously chosen feature, ω_ℓ, the distribution of $A_{M+1,\ell}|J_\ell$ is Poisson(rJ_ℓ), hence the conditional probability of $b_{M+1,\ell} = 1|J_\ell$ is $(1 - \mathrm{e}^{-rJ_\ell})$. It follows that

$$\mathbb{E}\big[(1 - \mathrm{e}^{-rJ_\ell})\big] = \frac{\int_0^\infty (1 - \mathrm{e}^{-r\lambda})\mathrm{e}^{-\lambda M r}\lambda^{c_{\ell,M}}\rho(\lambda)\,d\lambda}{\int_0^\infty \mathrm{e}^{-\lambda M r}\lambda^{c_{\ell,M}}\rho(\lambda)\,d\lambda}$$

is the probability of choosing an existing dish ω_ℓ. Now for $((H_i, X_i))$ following a IBP(Poisson$(r\lambda)$, ρ_M) sequence, it follows for customer $M + 1$:

$$\mathbb{P}(H_i \in d\lambda) = \frac{(1 - \mathrm{e}^{-r\lambda})\mathrm{e}^{-\lambda M r}\rho(\lambda)\,d\lambda}{\phi_{r,M+1}} \tag{4.2}$$

and the marginal of X_i is

$$\mathbb{P}(X_i = j) = \frac{\int_0^\infty r^j\lambda^j\mathrm{e}^{-r\lambda(M+1)}\rho(\lambda)d\lambda}{j!\phi_{r,M+1}}, \qquad j = 1, 2, \dots.$$

For customer 1, setting $M = 0$, it is easy to check that $\phi_{r,1}\mathbb{E}[X_i] = r\int_0^\infty \lambda\rho(\lambda)d\lambda$ which may not be finite.

REMARK 4.1. Note that the general $((H_i, X_i))$ IBP(Poisson$(r\lambda)$, ρ) sequence appears in Pitman [44], Section 3, in a very different context, where also the stable case is highlighted.

4.2.1. *Calculations for the IBP-Poisson-generalized gamma model.* We say that μ is a generalized gamma process with law denoted as GG$(\alpha, r\zeta; \theta)$ if

$$\rho(\lambda) = \frac{\theta}{\Gamma(1-\alpha)}\lambda^{-\alpha-1}\mathrm{e}^{-r\zeta\lambda} \qquad \text{for } 0 < \lambda < \infty,$$

for $\theta > 0$, and the ranges $0 < \alpha < 1$, $\zeta \ge 0$, or $\alpha \le 0$ and $\zeta > 0$. When $\alpha = 0$, this is the case of the gamma process. When $\alpha = -\kappa < 0$, this results in a class of gamma compound Poisson processes. The posterior distribution of $\mu|Z_1, \dots, Z_M$ is equivalent in distribution to $\mu_M + \sum_{\ell=1}^K J_\ell\delta_{\omega_\ell}$ where μ_M is a GG$(\alpha, r(M + \zeta); \theta)$ and the corresponding jumps satisfy $G_\ell \overset{d}{=} r(\zeta+M)J_\ell$ are conditionally independent Gamma$(c_{\ell,M}-\alpha,1)$ variables. $A_{M+1,\ell}$ is Poisson$(G_\ell/(M+\zeta))$, hence it has a negative-binomial distribution with parameters $(c_{\ell,M}-\alpha, 1/(M+1+\zeta))$. This implies that customer $M+1$ will sample a dish ω_ℓ that has been sampled by m_ℓ out of M previous customers who have assigned a total score of $c_{\ell,M}$, with probability

$$1-\left[\frac{M+\zeta}{M+\zeta+1}\right]^{c_{\ell,M}-\alpha}.$$

This shows for the Poisson model that dishes/features having larger scores $c_{\ell,M}$ have a larger chance of being selected by future customers. Customer $M+1$ chooses a Poisson$(\phi_{r,M+1})$ number of new dishes where in this case, $\phi_{r,M+1} = \theta\tilde{\psi}_{\alpha,r}(M+\zeta)$ defined as

$$\tilde{\psi}_{\alpha,r}(M+\zeta) = \begin{cases} \dfrac{r^\alpha}{\alpha}\big[(M+1+\zeta)^\alpha-(M+\zeta)^\alpha\big], & \text{if } 0<\alpha<1, \zeta\ge 0,\\ \log\big(1+1/(M+\zeta)\big), & \text{if } \alpha=0, \zeta>0,\\ \dfrac{r^\kappa}{\kappa}\big[(M+\zeta)^{-\kappa}-(M+1+\zeta)^{-\kappa}\big], & \text{if } \alpha=-\kappa<0, \zeta>0. \end{cases} \tag{4.3}$$

Now for customer $M+1$, $((H_i,X_i))$ are specified as follows: $rH_i \overset{d}{=} \Sigma_{\alpha,M+\zeta}$, where $\Sigma_{\alpha,M+\zeta}$ is a random variable with density

$$f_{\Sigma_{\alpha,M+\zeta}}(\lambda) = \begin{cases} \dfrac{\alpha(1-e^{-\lambda})\lambda^{-\alpha-1}e^{-\lambda(M+\zeta)}}{\Gamma(1-\alpha)[(M+1+\zeta)^\alpha-(M+\zeta)^\alpha]}, & \text{if } 0<\alpha<1, \zeta\ge 0,\\ \dfrac{(1-e^{-\lambda})\lambda^{-1}e^{-\lambda(M+\zeta)}}{\log(1+1/(M+\zeta))}, & \text{if } \alpha=0, \zeta>0,\\ \dfrac{(1-e^{-\lambda})\lambda^{\kappa-1}e^{-\lambda(M+\zeta)}}{\Gamma(\kappa)[(M+\zeta)^{-\kappa}-(M+1+\zeta)^{-\kappa}]}, & \text{if } \alpha=-\kappa<0, \zeta>0. \end{cases} \tag{4.4}$$

Integrating the slab distribution of $A'|\lambda$, specified by (2.3), with respect to these densities leads to the marginal distribution of X_i, for $j=1,2,\dots,$

$$\mathbb{P}(X_i=j) = \begin{cases} \dfrac{(M+1+\zeta)^{\alpha-j}}{[(M+1+\zeta)^\alpha-(M+\zeta)^\alpha]}\dfrac{\alpha\Gamma(j-\alpha)}{j!\Gamma(1-\alpha)}, & \text{if } 0<\alpha<1, \zeta\ge 0,\\ \dfrac{(M+1+\zeta)^{-j}}{j\log(1+1/(M+\zeta))}, & \text{if } \alpha=0, \zeta>0,\\ \dfrac{(M+1+\zeta)^{-(\kappa+j)}}{[(M+\zeta)^{-\kappa}-(M+1+\zeta)^{-\kappa}]}\dfrac{\Gamma(j+\kappa)}{j!\Gamma(\kappa)}, & \text{if } \alpha=-\kappa<0,\ \zeta>0. \end{cases} \tag{4.5}$$

REMARK 4.2. The random variables with densities (4.4) are easy to sample as they can be represented as simple scale mixtures of gamma random variables. These variables appear in various places. See [29], Section 3, for further details. For $M + \zeta = 0$ and $0 < \alpha < 1$, the distribution in (4.5) is sometimes referred to as Sibuya's distribution, and has an infinite mean.

4.3. *Negative*-Binomial(r, p). In the case where $A|p$ is negative-binomial (r, p), denoted as $\mathrm{NB}(r, p)$,

$$\mathbb{P}(A = a|p) = \binom{a + r - 1}{a} p^a (1 - p)^r, \qquad a = 0, 1, \dots.$$

It follows that $\pi_A(p) = 1 - (1 - p)^r$, and A' has a discrete slab distribution $\tilde{G}_{A'}$, specified for $a = 1, 2, \dots$ by

$$\mathbb{P}\bigl(A' = a|p\bigr) = \mathbb{P}(X_i = a|H_i = p) = \frac{\binom{a+r-1}{a} p^a (1 - p)^r}{1 - (1 - p)^r}. \tag{4.6}$$

It follows that

$$\mathrm{S}_{\text{spike}\times\text{slab}}(p|\mathcal{A}_\ell) = (1 - p)^{Mr} p^{c_{\ell,M}} \prod_{i \in \mathcal{B}_\ell} \binom{A'_{i,\ell} + r - 1}{A'_{i,\ell}},$$

where $c_{\ell,M} = \sum_{i \in \mathcal{B}_\ell} A'_{i,\ell}$. Furthermore, $\rho_M(p|\omega) = (1 - p)^{Mr} \rho(p|\omega)$, which implies that each J_ℓ has density proportional to $(1 - p)^{Mr} p^{c_{\ell,M}} \rho(p|\omega_\ell)$. These facts lead immediately to the specification of the distributions of N_M, μ_M and the posterior distributions of N and μ.

We now give the results for describing $Z_{M+1}|Z_1, \dots, Z_M$, and hence a generative scheme for negative-binomial IBP models based on any choice for ρ. For a previously chosen dish/feature, ω_ℓ, the distribution of $A_{M+1,\ell}|J_\ell$ is $\mathrm{NB}(r, J_\ell)$, hence the conditional probability of $b_{M+1,\ell} = 1|J_\ell$ is $1 - [1 - J_\ell]^r$. It follows that

$$\mathbb{E}\bigl[1 - (1 - J_\ell)^r\bigr] = \frac{\int_0^1 [1 - (1 - p)^r](1 - p)^{Mr} p^{c_{\ell,M}} \rho(p)\, dp}{\int_0^1 (1 - p)^{Mr} p^{c_{\ell,M}} \rho(p)\, dp},$$

is the probability that customer $M + 1$ chooses dish ω_ℓ that has been previously selected by m_ℓ customers who have given a total score of $c_{\ell,M}$.

We now describe a mapping which relates some properties of IBP-Poisson models with the IBP-negative binomial model. Let $\tilde{\rho}$ denote a Lévy density on $(0, \infty)$ defined by ρ on $[0, 1]$ by the transformation $\lambda = -\log(1 - p)$. Specifically $\tilde{\rho}(\lambda) := \mathrm{e}^{-\lambda} \rho(1 - \mathrm{e}^{-\lambda})$. Then $\varphi_k := \phi_{r,k}$, and is equivalent to

$$\int_0^1 \bigl[1 - (1 - p)^r\bigr](1 - p)^{(k-1)r} \rho(dp) = \int_0^\infty \bigl(1 - \mathrm{e}^{-ry}\bigr)\mathrm{e}^{-(k-1)ry} \tilde{\rho}(dy). \tag{4.7}$$

See [13, 26] for this identity. This means that customer $M + 1$ will sample a Poisson$(\phi_{r,M+1})$ number of dishes, which is the same under a IBP-Poisson model specified by $\tilde{\rho}$. Now for $((H_i, X_i))$ following a IBP(NB(r, p), ρ_M) sequence, it follows for customer $M + 1$:

$$\mathbb{P}(H_i \in dp) = \frac{[1-(1-p)^r](1-p)^{Mr}\rho(p)\,dp}{\phi_{r,M+1}} \tag{4.8}$$

and the distribution of $X_i|H_i$ is specified by (4.6). Hence, the marginal of X_i is

$$\mathbb{P}(X_i = a) = \frac{\binom{a+r-1}{a}\int_0^1 p^a(1-p)^{(M+1)r}\rho(p)\,dp}{\phi_{r,M+1}}, \qquad a = 1, 2, \ldots,$$

and also $H_i|X_i = a$ is proportional to $p^a(1-p)^{(M+1)r}\rho(p)$. If $\tilde{H}_i$ is determined by an IBP(Poisson$(r\lambda)$, $\tilde{\rho}_M$) sequence, it follows that $H_i \stackrel{d}{=} (1-\mathrm{e}^{-\tilde{H}_i})$. This equivalence and the identity (4.7) are based on the correspondence between the spike probabilities. There is no equivalence in terms of quantities involving slab variables, such as (X_i) and (J_ℓ). We will demonstrate this again in our next examples involving normal random variables. By setting $M = 0$, one can check that for customer 1,

$$\mathrm{E}[X_i] = \frac{r\int_0^1 p(1-p)^{-1}\rho(p)\,dp}{\phi_{r,1}}, \tag{4.9}$$

which is certainly not always finite. A fact that seems not to have been noticed in the literature is that this happens for the commonly used range of the parameters for the Beta/stable-Beta process. Note that if $Z_1 \sim$ IBP(NB(r, p), ρB_0) then one obtains $\mathbb{E}[Z_1(\Omega)] = \phi_{r,1}\mathbb{E}[X_1]$. Using (4.9), and the stable-Beta process specification, $\rho_{\alpha,\beta}(p) = \theta p^{-\alpha-1}(1-p)^{\beta+\alpha-1}$, gives

$$\mathbb{E}\big[Z_1(\Omega)\big] = \theta r\int_0^1 p^{-\alpha}(1-p)^{\beta+\alpha-2}\,dp.$$

So it follows that if $\beta \le 1-\alpha$ then $\mathbb{E}[Z_1(\Omega)] = \infty$. This still means that customer 1 selects a Poisson $(\phi_{r,1})$, number of dishes, but in these cases gives very high scores to each dish, which is reflected by $\mathbb{E}[X_1] = \infty$.

The posterior distribution of $\mu|Z_1, \ldots, Z_M$ under the case where μ is a stable-Beta process with parameters $(\alpha, \beta; \theta)$, with $\rho_{\alpha,\beta}$, was established through some rather involved arguments in [6], see also [8]. Our results can be used to obtain this as follows, since $\rho_M(s) = \rho_{\alpha,\beta+Mr}(s)$ it follows that μ_M is a stable-Beta process with parameters $(\alpha, Mr+\beta; \theta)$, and the (J_ℓ) are independent Beta$(c_{\ell,M}-\alpha, Mr+\beta+\alpha)$. When $\alpha = 0$, that is to say when μ is a Beta process, Heaukulani and Roy [17] describe the explicit generative scheme in this case. Moreover, they give the representation of Z showing that the marginal distributions of the (X_i) have digamma distributions. Decompositions involving the corresponding H_i are not noted. See also [53, 54]. Outside of the Beta process case the results we have established are new. One may use, for instance, a generalized gamma process, GG$(\alpha, r\zeta; \theta)$ by employing the change of variable $p = 1-\mathrm{e}^{-\lambda}$. Calculations for this are very similar to our next example which we now describe.

4.4. *Example*: *Calculations for spike and slab normal distributions*. In the previous examples where G_A is based on classic discrete distributions, (binomial, Poisson, negative-binomial), it is perhaps natural to think in terms of A rather than in terms of their spike, $1-\pi_A$, and their slab components A'. In general, it makes sense to think of modeling via the decomposition (π_A, A') to construct particular distributions A. The purpose of this section is to demonstrate how to obtain calculations for a particularly challenging dependent model involving a continuous distribution for A'. Specifically, here $b_{i,k}, A'_{i,k}|\lambda_k$ are conditionally independent such that $b_{i,k}$ is Bernoulli$(1-\mathrm{e}^{-\lambda_k})$ and $A'_{i,k}|\lambda_k$ is a Normal$(\eta_k, 1/\lambda_k)$ random variable with density $\phi(y|\eta_k,\lambda_k) = \frac{\sqrt{\lambda_k}}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\lambda_k}{2}(y-\eta_k)^2}$.

REMARK 4.3. While this model bears some similarities to [34], and can in principle be used in similar situations, it has not been treated in the literature. Note also that $\mathbb{P}(A \neq 0|\lambda) := \pi_A(\lambda) = (1-\mathrm{e}^{-\lambda})$ depends on the variance parameter $(1/\lambda)$. This indicates that previously chosen latent features, say ω_ℓ, with highly varied values about their mean η_ℓ [representing uncertainty reflected by $\sum_{i\in B_\ell}(A'_{i,\ell}-\eta_\ell)^2$], are less likely to be chosen by new customers.

REMARK 4.4. We do not concern ourselves with how to deal with the (η_k). These may be fixed, estimated externally to the Poisson random measure by Bayesian parametric priors, or modeled as points of a Poisson random measure which arises as a special case of Section 5.

The challenge for this model in terms of tractability is to find a suitable $\rho(\lambda)$ to deal with terms depending on $(1-\mathrm{e}^{-\lambda})\phi(y|\eta,\lambda)$. Note that since $\pi_A(\lambda) = (1-\mathrm{e}^{-\lambda})$ is the same as the IBP-Poisson case with $r=1$, it follows that customer $M+1$ will select a Poisson$(\phi_{1,M+1})$ new dishes. Furthermore, the distribution of the (H_i), N_M, μ_M are the same as in the Poisson setting. The difference is reflected in the specification of the normal distribution of the slab variable A', which affects $((X_i),(J_\ell),(A_{i,\ell}))$.

Based on a sample of size M,

$$\mathrm{S}_{\text{spike}\times\text{slab}}(\lambda|\mathcal{A}_\ell) = \mathrm{e}^{-(M-m_\ell)\lambda}\left(1-\mathrm{e}^{-\lambda}\right)^{m_\ell}\left(\frac{\lambda}{2\pi}\right)^{\frac{m_\ell}{2}}\mathrm{e}^{-\frac{\lambda}{2}\sum_{i\in\mathcal{B}_\ell}(A'_{i,\ell}-\eta_\ell)^2}.$$

This leads to a general expression for the joint distribution of $(J_\ell), Z_1,\dots,Z_M$, sans the unique values $(\omega_1,\dots,\omega_K)$,

$$\left[\prod_{k=1}^{M}\mathrm{e}^{-\varphi_k}\right]\prod_{\ell=1}^{K}\mathrm{e}^{-(M-m_\ell)\lambda_\ell}\left(1-\mathrm{e}^{-\lambda_\ell}\right)^{m_\ell}\left(\frac{\lambda_\ell}{2\pi}\right)^{\frac{m_\ell}{2}}\mathrm{e}^{-\frac{\lambda_\ell}{2}\sum_{i\in\mathcal{B}_\ell}(A'_{i,\ell}-\eta_\ell)^2}\rho(\lambda_\ell).$$

Here, as in the IBP-Poisson case we choose μ to be a $\mathrm{GG}(\alpha,\zeta;\theta)$ process. It follows that for $\mu|Z_1,\dots,Z_M$, μ_M is $\mathrm{GG}(\alpha,\zeta+M;\theta)$ just as in the IBP-Poisson case. Furthermore, $\varphi_{M+1} = \theta\tilde{\psi}_{\alpha,1}(M+\zeta)$ defined in (4.3). Hence, customers select new dishes as in the IBP-Poisson case. With respect to the values

$((H_i, X_i))$ for customer $M+1$ one should think in terms of $(H_\ell, X_\ell)|\eta_\ell$, for $\ell = K_M+1, \ldots, K_{M+1}$ where the relevant indices correspond to the $\tilde{K}_{M+1}$ newly selected dishes. Here, the $H_\ell \stackrel{d}{=} \Sigma_{\alpha, M+\zeta}$, specified by (4.4), and hence are easily sampled. The corresponding $X_\ell|H_\ell, \eta_\ell$ are Normal$(\eta_\ell, 1/H_\ell)$ variables. The density of $X_\ell|\eta_\ell$ takes the form:

$$\frac{\Gamma(\frac{3}{2}-\alpha)}{\Gamma(1-\alpha)} \frac{\tilde{\psi}_{\alpha-\frac{1}{2},1}(M+\zeta+\frac{1}{2}(y-\eta_\ell)^2)}{\sqrt{2\pi}\,\tilde{\psi}_{\alpha,1}(M+\zeta)} \qquad \text{for } -\infty < y < \infty. \tag{4.10}$$

When $\alpha = 1/2$, (4.10) specializes to

$$\frac{\log(1+\frac{1}{M+\zeta+\frac{1}{2}(y-\eta_\ell)^2})}{2^{3/2}\pi[\sqrt{(M+1+\zeta)}-\sqrt{M+\zeta}]}.$$

In the IBP sequential selection scheme, one sets the $X_\ell = A'_{M+1,\ell}$ for $\ell = K_M+1, \ldots, K_{M+1}$.

Customer $M+1$ chooses among the $K := K_M$ existing dishes using the following results. For $m_\ell = |\mathcal{B}_\ell| > 1$, the density of J_ℓ is now:

$$f_{J_\ell}(\lambda) := \frac{\mathrm{e}^{-(S_{\ell,M}-m_\ell)\lambda}(1-\mathrm{e}^{-\lambda})^{m_\ell}\lambda^{\frac{m_\ell}{2}-\alpha-1}}{\Gamma(m_{\ell/2}-\alpha)\sum_{j=0}^{m_\ell}\binom{m_\ell}{j}[S_{\ell,M}-j]^{\alpha-m_\ell/2}(-1)^{m_\ell-j}}, \tag{4.11}$$

where $S_{\ell,M} = M+\zeta+\frac{1}{2}\sum_{i\in\mathcal{B}_\ell}(A'_{i,\ell}-\eta_\ell)^2$.

REMARK 4.5. The form of the normalizing constant arises from the binomial expansion $(1-\mathrm{e}^{-\lambda})^{m_\ell} = \sum_{j=0}^{m_\ell}\binom{m_\ell}{j}\mathrm{e}^{-\lambda(m_\ell-j)}(-1)^{m_\ell-j}$, and Laplace transforms of gamma distributions.

When $m_\ell > 1$, it is easiest to sample from the distribution of $A_{M+1,\ell}|J_\ell$ which is equivalent to $b_{M+1,\ell}A'_{M+1,\ell}|J_\ell$, where $b_{M+1,\ell}$ is Bernoulli$((1-\mathrm{e}^{-J_l}))$, and $A'_{M+1,\ell}|J_\ell$ is Normal$(\eta_\ell, 1/J_\ell)$, where J_ℓ is sampled from $f_{J_\ell}(\lambda)$ given in (4.11). However, for completeness, it follows that the probability of selecting a previously selected dish ω_ℓ is $\mathbb{P}(b_{M+1,\ell}=1|m_\ell>1) = \mathbb{E}[(1-\mathrm{e}^{-J_l})]$, given by

$$\frac{\sum_{j=0}^{m_\ell+1}\binom{m_\ell+1}{j}[S_{\ell,M}+1-j]^{\alpha-m_\ell/2}(-1)^{m_\ell+1-j}}{\sum_{j=0}^{m_\ell}\binom{m_\ell}{j}[S_{\ell,M}-j]^{\alpha-m_\ell/2}(-1)^{m_\ell-j}}.$$

The density of $A_{M+1,\ell}|b_{M+1,\ell}=1, m_\ell>1$ can be expressed as

$$\frac{\Gamma(\frac{m_\ell+1}{2}-\alpha)\sum_{j=0}^{m_\ell+1}\binom{m_\ell+1}{j}[S_{\ell,M}+1-j+\frac{1}{2}(y-\eta_\ell)^2]^{\alpha-\frac{m_\ell+1}{2}}(-1)^{m_\ell+1-j}}{\Gamma(m_{\ell/2}-\alpha)\sum_{j=0}^{m_\ell+1}\binom{m_\ell+1}{j}[S_{\ell,M}+1-j]^{\alpha-m_\ell/2}(-1)^{m_\ell+1-j}}.$$

When $m_\ell = 1$, $J_\ell \stackrel{d}{=} \Sigma_{\alpha-1/2, S_{\ell,M}-1}$. It follows in this case the chance that customer $M+1$ chooses the existing dish ω_ℓ, which has been selected by only one of the M previous customers, is

$$\mathbb{P}(b_{M+1,\ell} = 1 | m_\ell = 1) := \mathbb{P}(A_{M+1,\ell} \neq 0 | m_\ell = 1) = 1 - \frac{\tilde{\psi}_{\alpha-1/2}(S_{\ell,M})}{\tilde{\psi}_{\alpha-1/2}(S_{\ell,M} - 1)}.$$

The density of $J_\ell | A_{M+1,\ell} \neq 0, m_\ell = 1$ is

$$\frac{\tilde{\psi}_{\alpha-1/2}(S_{\ell,M} - 1)(1 - \mathrm{e}^{-\lambda}) f_{\Sigma_{\alpha-1/2, S_{\ell,M}-1}}(\lambda)}{[\tilde{\psi}_{\alpha-1/2}(S_{\ell,M} - 1) - \tilde{\psi}_{\alpha-1/2}(S_{\ell,M})]}.$$

Integrating $\phi(y|\eta_\ell, \lambda)$ over λ with respect to this density leads to the density of $A_{M+1,\ell} | A_{M+1,\ell} \neq 0, m_\ell = 1$,

$$\frac{\Gamma(1-\alpha) \sum_{j=0}^{2} \binom{2}{j} [S_{\ell,M} + 1 - j + \frac{1}{2}(y - \eta_\ell)^2]^{\alpha-1} (-1)^{2-j}}{\sqrt{2\pi} [\tilde{\psi}_{\alpha-1/2}(S_{\ell,M} - 1) - \tilde{\psi}_{\alpha-1/2}(S_{\ell,M})]}.$$

This is an unfamiliar density, however again one can sample this much more easily using the fact that the conditional distribution of $A_{M+1,\ell} | J_\ell, A_{M+1,\ell} \neq 0, m_\ell = 1$ is Normal$(\eta_\ell, 1/J_\ell)$.

5. Constrained Multivariate Priors and IBP generalizations. Possible multivariate extensions of the generalized IBP presents new challenges in terms of construction of processes, selection of multivariate versions of μ, their analysis and interpretations. In this section, we will show how to do this in great generality. This, similar to our results for the univariate case, allows one to focus on modeling issues and interpretations that arise in a more intricate multidimensional setting rather than be encumbered by the calculus of random measures. Since the joint distributions are similar in form to the univariate cases, the arguments we employed previously carry over to this setting in a rather transparent fashion. We present results for a very general multivariate setting and then specialize these to a simple multinomial setting. We also discuss the case of multivariate CRMs which have suitable constraints to handle, for instance, multinomial extensions of the IBP.

Let us first say a few words about multivariate Lévy processes with positive jumps. For $\mathbf{x}_q = (x_1, x_2, \dots, x_q)$, let $\rho_q(\mathbf{x}_q)$ denote a sigma-finite function concentrated on $\mathbb{R}_+^q / \{\mathbf{0}\}$ then following Barndorff-Nielsen, Pedersen and Sato [3], Proposition 3.1, there exists a multivariate CRM $\mu_0 := (\mu_j, j = 1, \dots, q)$ with jumps specified by ρ_q if $\int_{\mathbb{R}_+^q} \min(x_\cdot, 1) \rho_q(\mathbf{x})\, d\mathbf{x} < \infty$, where $x_\cdot = \sum_{j=1}^{q} x_j$. We shall assume a common base measure $B_0(d\omega)$ representing again features. So it follows that one can write $\mu_j = \sum_{k=1}^{\infty} \tau_{j,k} \delta_{\tilde{\omega}_k}$. Furthermore, $\mu_\cdot = \sum_{j=1}^{q} \mu_j = \sum_{k=1}^{\infty} \tau_{\cdot,k} \delta_{\tilde{\omega}_k}$, for $\tau_{\cdot,k} = \sum_{j=1}^{q} \tau_{j,k}$. The multivariate CRM μ_0 is constructed from N a Poisson random measure on $(\mathbb{R}_+^q, \Omega)$ with mean intensity $\mathbb{E}[N(d\mathbf{x}_q, d\omega)] = \nu(d\mathbf{x}_q, d\omega) = \rho_q(d\mathbf{x}_q|\omega)B_0(d\omega)$, and can be represented as $N := \sum_{k=1}^{\infty} \delta_{\tilde{\tau}_{q,k},\tilde{\omega}_k}$, where $\tilde{\tau}_{q,k} = (\tau_{1,k}, \dots, \tau_{q,k})$. Hence, for $j = 1, \dots, q$, $\mu_j(d\omega) := \int_{\mathbb{R}_+^q} h_j(\mathbf{x}_q) N(d\mathbf{x}_q, d\omega)$, where $h_j(\mathbf{x}_q) = x_j$. We use the notation $\rho_q(d\mathbf{x}_q|\omega) := \rho_q(\mathbf{x}_q|\omega)\,d\mathbf{x}_q$. In order to extend definitions to $\mathbb{R}^q$, choose ρ_q such that $\rho_q(\{\mathbf{0}\}) = 0$, and $\int_{\mathbb{R}^q} \min(|x.|, 1)\rho_q(\mathbf{x})\,d\mathbf{x} < \infty$.

REMARK 5.1. Within the IBP context, the points $\mathbf{x}_q$ will generally relate to the possible parameter values of a distribution, for instance, a multinomial distribution. As such, these points will often take values in a constrained sub-space of $\mathbb{R}^q$. A challenge is to construct random measures such that their prior and posterior distributions have points that take values in the same constrained subspace of $\mathbb{R}^q$, with probability one. We will illustrate this for a multivariate IBP based on a simple multinomial distribution.

5.1. *A general multivariate IBP process.* Let now $A_0 := (A_1, \dots, A_\upsilon)$, for υ a positive integer not necessarily equal to q, denote a random vector taking values in $\mathcal{A} \subseteq \mathbb{R}^\upsilon$, with conditional distribution $G_{A_0}(\cdot|\mathbf{s}_q)$, where $\mathbf{s}_q \in \mathbb{R}^q$, and such that $1 - \pi_{A_o}(\mathbf{s}_q) := \mathbb{P}(A_1 = 0, \dots, A_\upsilon = 0|\mathbf{s}_q) > 0$. We require that $\varphi := \int_\Omega \int_{\mathbb{R}^q} \pi_{A_0}(\mathbf{s}_q)\rho_q(d\mathbf{s}_q|\omega)B_0(d\omega) < \infty$. One can decompose A_0 in terms of a Bernoulli $(\pi_{A_0}(\mathbf{s}_q))$ spike variable and a generalized multivariate slab vector $A'_0 := (A'_1, \dots, A'_\upsilon)$, with distribution:

$$\tilde{G}_{A'_0}(d\mathbf{a}_0|\mathbf{s}_q) = \left[\frac{\mathbb{I}\{\mathbf{a}_0 \notin \mathbf{0}\}G_{A_0}(d\mathbf{a}_0|\mathbf{s}_q)}{\pi_{A_0}(\mathbf{s}_q)}\right].$$

This means that at least one of the components of A'_0 is nonzero with probability one. Then we can define a vector valued process $Z_0^{(i)} := (Z_1^{(i)}, \dots, Z_\upsilon^{(i)})$, where $Z_j^{(i)} = \sum_{k=1}^{\infty} A_{j,k}^{(i)}\delta_{\tilde{\omega}_k}$, and $Z_\cdot^{(i)} = \sum_{j=1}^{\upsilon} Z_j^{(i)} = \sum_{k=1}^{\infty} A_{\cdot,k}^{(i)}\delta_{\tilde{\omega}_k}$, where conditional on $\mu_0 := (\mu_1, \dots \mu_q)$, for each fixed (i, k), $A_{0,k}^{(i)} := (A_{1,k}^{(i)}, \dots, A_{\upsilon,k}^{(i)})$ is independent $G_{A_0}(\cdot|\mathbf{s}_{q,k})$, where $\mathbf{s}_{q,k} = (s_{1,k}, \dots, s_{q,k})$ are vector valued points of a PRM with intensity $\rho_q(\cdot|\omega)$. We say that $Z_0^{(1)}, \dots, Z_0^{(M)}|\mu_0$ are i.i.d. IBP$(G_{A_0}|\mu_0)$. Let $\mathbf{J}_{q,\ell} = (J_{1,\ell}, \dots, J_{q,\ell})$ denote a random vector of latent jumps that have been picked from N, along with ω_ℓ, with arguments $\mathbf{s}_{q,\ell}$. Set $\mathcal{B}_\ell := \{i : Z_0^{(i)}(\omega_\ell) = A_{0,\ell}^{(i)} \notin \mathbf{0}\}$, and denote the arguments of $A_{0,j}^{(i)}$ as $\mathbf{a}_{0,j}^{(i)}$. Despite the extension to the multivariate setting, the form of the likelihood structure shares many similarities with the univariate case and one may conclude that the relevant joint distribution of $(\mathbf{J}_{q,\ell})$, $(Z_0^{(1)}, \dots, Z_0^{(M)})$ is given by

$$\left[\prod_{k=1}^{M} \mathrm{e}^{-\varphi_k}\right] \prod_{\ell=1}^{K} \mathrm{S}_{\mathrm{spike}\times\mathrm{slab}}(\mathbf{s}_{q,\ell}|\mathcal{A}_{0,\ell})\rho_q(\mathbf{s}_{q,\ell}|\omega_\ell)B_0(d\omega_\ell), \tag{5.1}$$

where now, for $\mathcal{A}_{0,\ell} = \{A_{0,\ell}^{\prime(i)} : i \in \mathcal{B}_\ell\}$,

$$\mathrm{S}_{\text{spike}\times\text{slab}}(\mathbf{s}_q|\mathcal{A}_{0,\ell}) = [1-\pi_{A_0}(\mathbf{s}_q)]^{M-m_\ell}\pi_{A_0}^{m_\ell}(\mathbf{s}_q)\prod_{i\in\mathcal{B}_\ell}\tilde{G}_{A_0'}(d\mathbf{a}_{0,\ell}^{(i)}|\mathbf{s}_q), \tag{5.2}$$

$\nu_M(d\mathbf{s}_q, \omega) = [1-\pi_{A_0}(\mathbf{s}_q)]^M \rho_q(d\mathbf{s}_q|\omega)B_0(d\omega) := \rho_{q,M}(d\mathbf{s}_q|\omega)B_0(d\omega)$, and for $k = 1, 2, \ldots$; $\varphi_k = \int_\Omega\int_{\mathbb{R}^q}\pi_{A_0}(\mathbf{s}_q)\rho_{q,k-1}(d\mathbf{s}_q|\omega)B_0(d\omega)$. We now summarize the results.

PROPOSITION 5.1. *Suppose that* $Z_0^{(1)}, \ldots, Z_0^{(M)}|\mu_0$ *are i.i.d.* $\mathrm{IBP}(G_{A_0}|\mu_0)$, *where* μ_0 *is a multivariate* $\mathrm{CRM}(\rho_q B_0)$:

(i) *Then it follows that the posterior distribution of* $N|(Z_0^{(1)}, \ldots, Z_0^{(M)})$ *is equivalent to the distribution of* $N_M + \sum_{\ell=1}^K \delta_{\mathbf{J}_{q,\ell},\omega_\ell}$, *where* N_M *is a* $\mathrm{PRM}(\rho_{q,M}B_0)$ *and the vector* $\mathbf{J}_{q,\ell} = (J_{1,\ell}, \ldots, J_{q,\ell})$ *has joint density, with argument* $\mathbf{s}_{q,\ell}$, *proportional to* $\mathrm{S}_{\text{spike}\times\text{slab}}(\mathbf{s}_{q,\ell}|\mathcal{A}_{0,\ell})\rho_q(\mathbf{s}_{q,\ell}|\omega_\ell)$ *described in* (5.2).

(ii) *Let* $\mu_{0,M} = (\mu_{1,M}, \ldots, \mu_{q,M})$ *denote a multivariate* $\mathrm{CRM}(\rho_{q,M}B_0)$, *then the posterior distribution of* $\mu_0|Z_0^{(1)}, \ldots, Z_0^{(M)}$, *is equivalent to that of a multivariate process whose* j*th component is equivalent in distribution to* $\mu_{j,M} + \sum_{\ell=1}^K J_{j,\ell}\delta_{\omega_\ell}$.

(iii) *The corresponding distribution of* $Z_0^{M+1}|(Z_0^{(1)}, \ldots, Z_0^{(M)})$ *can be represented in terms of* $\tilde{Z}_0^{(M+1)} := (\tilde{Z}_1^{(M+1)}, \ldots, \tilde{Z}_\upsilon^{(M+1)})$, *which is determined by the points of* N_M, *call it an* $\mathrm{IBP}(A_0, \rho_{q.M}B_0)$ *vector, and* $\ell = 1, \ldots, K$ *vectors* $(A_{0,M+1}^{(\ell)})$, *where for each fixed* ℓ, $A_{0,\ell}^{(M+1)}|\mathbf{J}_{q,\ell} = \mathbf{s}_q$ *has distribution* $G_{A_0}(\cdot|\mathbf{s}_q)$. *In other words component-wise* $Z_0^{(M+1)}$ *can be represented in distribution as* $\tilde{Z}_j^{(M+1)} + \sum_{\ell=1}^K A_{j,\ell}^{(M+1)}\delta_{\omega_\ell}$, *where* $A_{j,\ell}^{(M+1)}$ *is the* j*th component of* $A_{0,\ell}^{(M+1)}$.

We close with a generalization of Proposition 3.3, which can be used to sample $\tilde{Z}_0^{(M+1)}$ by replacing the general ρ_q below with $\rho_{q,M}$.

PROPOSITION 5.2. *Let* $Z_0 = (Z_1, \ldots, Z_\upsilon)$ *have an* $\mathrm{IBP}(A_0, \rho_q B_0)$ *distribution, where it is assumed that* ρ_q *is homogeneous. If* $\varphi := \int_{\mathbb{R}^q}\pi_{A_0}(\mathbf{s}_q)\rho_q(d\mathbf{s}_q) < \infty$, *then for* $j = 1, \ldots, \upsilon$, $Z_j \overset{d}{=} \sum_{k=1}^{\xi(\varphi)} X_{j,k}\delta_{\tilde{\omega}_k}$ *where* $(\tilde{\omega}_k)$ *are i.i.d.* B_0, $X_{0,k} := (X_{1,k}, \ldots X_{\upsilon,k})$ *are i.i.d. across* k, *and independent of* $\xi(\varphi)$, *a Poisson random variable with mean* φ. *Furthermore, setting* $H_{0,k} = (H_{1,k}, \ldots H_{q,k})$, *there are the i.i.d. pairs* $((H_{0,k}, X_{0,k}))$ *with slab distribution*:

$$\mathbb{P}(X_{0,k} \in d\mathbf{a}_0|H_{0,k} = \mathbf{s}_q) = \tilde{G}_{A_0'}(d\mathbf{a}_0|\mathbf{s}_q) = \left[\frac{\mathbb{I}\{\mathbf{a}_0 \notin \mathbf{0}\}G_{A_0}(d\mathbf{a}_0|\mathbf{s}_q)}{\pi_{A_0}(\mathbf{s}_q)}\right]$$

and $\mathbb{P}(H_{0,k} \in d\mathbf{s}_q) = \pi_{A_0}(\mathbf{s}_q)\rho_q(d\mathbf{s}_q)/\varphi$. *Replace* $\rho_q(d\mathbf{s}_q)$ *with* $\rho_q(d\mathbf{s}_q|\omega)$ *to obtain the in-homogeneous case.*

The proof proceeds in the same manner as the univariate case using the characteristic functional of $\sum_{j=1}^{\upsilon} t_j Z_j(f)$, for general t_j.

REMARK 5.2. The multivariate IBP generative schemes proceed in the same fashion as described in Section 3.2 with $((H_{0,i}, X_{0,i}))$ playing the role of $((H_i, X_i))$. The in-homogeneous scheme just involves $((H_{0,\ell}, X_{0,\ell}))|(w_\ell)$.

5.2. *The simple multinomial case*: *Indian Buffet Process with a condiment*. For $q = 1, 2, \ldots$ consider the vector of probabilities $\mathbf{p}_q = (p_1, \ldots, p_q)$ taking values in the simplex $\mathcal{S}_q = \{\mathbf{s}_q : s_j > 0, s. = \sum_{j=1}^{q} s_j < 1\}$. A basic multivariate model is the case where A_0 has a simple multinomial $(1, \mathbf{p}_q, 1 - \sum_{i=1}^{q} p_i) := \mathsf{M}(1, \mathbf{p}_q)$ distribution, with probability mass function:

$$\mathbb{P}(A_1 = a_1, \ldots, A_q = a_q | \mathbf{p}_q) = \left[\prod_{j=1}^{q} p_j^{a_j}\right](1 - p.)^{1 - \sum_{k=1}^{q} a_k},$$

where at most one of the terms in $\mathbf{a}_0 := (a_1, \ldots, a_q)$ is one and the others are 0. It follows that $A. := \sum_{j=1}^{q} A_j$ has a Bernoulli$(\sum_{j=1}^{q} p_j)$ distribution. The spike is determined by $1 - \pi_{A_0}(\mathbf{p}_q) = (1 - p.)$ and the multivariate slab distribution $\tilde{G}_{A_0'}(\cdot|\mathbf{p}_q)$ is a simple multinomial with probability mass function, for $a. := \sum_{i=1}^{q} a_i$,

$$\mathbb{P}(A_1' = a_1, \ldots, A_q' = a_q | \mathbf{p}_q) = \mathbb{I}_{\{a.=1\}} \prod_{j=1}^{q} \left(\frac{p_j}{\sum_{i=1}^{q} p_i}\right)^{a_j}, \tag{5.3}$$

which is the distribution of $A_0|A. = 1$. Now for the matrix entries for each fixed (i, k), $A_{0,k}^{(i)}|\mathbf{p}_{q,k}$ is a $\mathsf{M}(1, \mathbf{p}_{q,k})$ vector. It follows that $A_{\cdot,k}^{(i)} := \sum_{j=1}^{q} A_{j,k}^{(i)}$ is Bernoulli$(\sum_{j=1}^{q} p_{j,k})$, and $A_{0,k}^{(i)}|A_{\cdot,k}^{(i)} = 1$ has distribution $\tilde{G}_{A_0'}(\cdot|\mathbf{p}_{q,k})$ specified by (5.3). Hence, $Z_{\cdot}^{(i)} = \sum_{j=1}^{\upsilon} Z_j^{(i)} = \sum_{k=1}^{\infty} A_{\cdot,k}^{(i)} \delta_{\tilde{\omega}_k}$, is a Bernoulli$(\sum_{j=1}^{q} p_j)$-IBP process based on the univariate CRM $\mu. = \sum_{k=1}^{\infty} [\sum_{j=1}^{q} p_{j,k}] \delta_{\tilde{\omega}_k}$. The points of $\mu.$ are specified by ρ defined by ρ_q through the transformation $p = p. = \sum_{j=1}^{q} p_j$. Each nonzero entry in the matrix contains a vector of length q where one entry takes the value 1 and the other entries in the vector take the value 0. This can perhaps be interpreted as the case where a particular feature or topic is selected and then further segmented into one of q categories. The q categories could refer to geographic regions, colors, genres, etc. Returning to the culinary metaphor, this could be described in terms of a customer selecting a certain dish but along with that choosing one particular condiment to go along with that dish. Perhaps a mango chutney or simply salt. This means that 2 or more individuals may have selected the same dish (have the same basic trait) but might differ in terms of the accompanying condiment. Notice that there is the following decomposition: $\mathcal{B}_\ell = \bigcup_{j=1}^{q} \mathcal{B}_{j,\ell}$, where $\mathcal{B}_{j,\ell} = \{i : Z_j^{(i)}(\omega_\ell) := A_{j,\ell}^{(i)} = 1\}$.

We now show how our results lead to a formal execution in this case. First, $\rho_{q,M}(\mathbf{p}_q|\omega) = (1-p_\cdot)^M \rho_q(\mathbf{p}_q|\omega)$ and

$$\mathrm{S}_{\text{spike}\times\text{slab}}(\mathbf{p}_{q,\ell}|\mathcal{A}_{0,\ell}) = \left[\prod_{j=1}^{q} [p_{j,\ell}]^{m_{j,\ell}}\right](1-p_{\cdot,\ell})^{M-m_\ell},$$

where in a sample of size M, $m_{j,\ell} = |\mathcal{B}_{j,\ell}|$ and $m_\ell = \sum_{j=1}^{q} m_{j,\ell}$

These facts determine the distributions of N_M, $\mu_{0,M}$ and $\tilde{Z}_0^{(M+1)}$, and the random vectors $\mathbf{J}_{q,\ell} = (J_{1,\ell}, \dots, J_{q,\ell})$. Marginally, for each ℓ, $A_{0,\ell}^{(M+1)}$ is $\mathsf{M}(1, \mathbf{r}_{q,\ell})$, where $\mathbf{r}_{q,\ell} = (r_{1,\ell}, \dots, r_{q,\ell})$ is defined for $k = 1, \dots, q$, as $r_{k,\ell} = \mathbb{E}[J_{k,\ell}]$. Hence, $A_{\cdot,\ell}^{(M+1)} = \sum_{j=1}^{q} A_{j,\ell}^{(M+1)}$ is Bernoulli$(\sum_{j=1}^{q} r_{j,\ell})$, and given $A_{\cdot,\ell}^{(M+1)} = 1$, $A_{0,\ell}^{(M+1)}$ becomes a simple multinomial distribution $\tilde{G}_{A'_0}(\cdot|\mathbf{u}_{q,\ell})$ specified by (5.3) where $\mathbf{u}_{q,\ell} = (u_{1,\ell}, \dots, u_{q,\ell})$ with $u_{j,\ell} = r_{j,\ell}/\sum_{k=1}^{q} r_{k,\ell}$.

Since $Z_\cdot^{(i)} = \sum_{j=1}^{q} Z_j^{(i)}$ is a Bernoulli process, it is evident that customers choose new dishes, ω_ℓ, according to the Indian buffet process arising in the univariate case. Specifically, via the transformation $p = \sum_{j=1}^{q} p_j$, one has for $k = 1, 2, \dots, \varphi_k = \sum_{j=1}^{q} \int_{\mathcal{S}_q} p_j \rho_{q,k-1}(d\mathbf{p}_q) = \int_0^1 p(1-p)^{k-1}\rho(dp)$. In addition, customer $M+1$ chooses new dishes and accompanying condiments according to $\tilde{Z}_0^{(M+1)}$, where for each k, $X_{0,k} := (X_{1,k}, \dots, X_{q,k})$ given $H_{0.k} = (p_1, \dots, p_q)$, has distribution $\tilde{G}_{A'_0}(\cdot|\mathbf{p}_q)$ specified by (5.3). $H_{0,k}$ is a random vector with joint density $[\sum_{j=1}^{q} p_j]\rho_{q,M}(p_1, \dots, p_q)/\varphi_{M+1}$. Hence, $X_{0,k}$ is marginally a simple multinomial with probability mass function $\mathbb{I}_{\{\sum_{i=1}^{q} x_i = 1\}} \prod_{j=1}^{q} \pi_{j,k}^{x_j}$, where $\pi_{j,k} = \mathbb{E}[H_{j,k}/H_{\cdot,k}]$ for $j = 1, \dots, q$. That is to say given each new dish chosen, ω_k, customer $M+1$ selects with it the jth of q possible condiments with probability $\pi_{j,k}$. Customer $M+1$ also will possibly choose a previously selected dish ω_ℓ, with probability $\sum_{j=1}^{q} r_{j,\ell}$ and given this will choose one of q condiments, say j, (possibly different than what has previously been chosen by others) with probability $u_{j,\ell}$. That is to say according to a multinomial$(1, (r_{1,\ell}, \dots, r_{q,\ell}))$ distribution. The customer will do this for each of the $\ell = 1, \dots, K$ previously selected dishes.

5.3. *Multinomial case*: *Stable-Beta-Dirichlet process priors*. One of the tasks in the multivariate case is to find convenient priors for $\mu_0 = (\mu_1, \dots, \mu_q)$ such that both its prior and posterior jump values are confined to a constrained space such as $\mathcal{S}_q$, with probability one. Kim, James and Weissbach [32] show, within a different data context but generally applicable, that this is not true if the components μ_j, $j = 1, \dots, q$ are specified to be independent.

In the simple multinomial case, we now show how the class of Beta-Dirichlet priors introduced in [32] leads to explicit results while maintaining values in $\mathcal{S}_q$. We introduce a slight modification of this model which allows for power-law behavior in the sense of Teh and Görür [48]. Specify μ_0 to be a stable-Beta-Dirichlet process with parameters $(\alpha, \beta+\alpha; \gamma_1, \dots, \gamma_q; \theta)$ by setting

$$\rho_q(\mathbf{p}_q) = \frac{\theta\Gamma(\sum_{j=1}^q \gamma_j)}{\prod_{j=1}^q \Gamma(\gamma_j)} p.^{-\alpha-\sum_{j=1}^q \gamma_j}(1-p.)^{\beta+\alpha-1}\prod_{j=1}^q p_j^{\gamma_j-1}\mathbb{I}_{\{0<p.<1\}}, \tag{5.4}$$

for $0 \le \alpha < 1$, $\beta > -\alpha$, $\theta > 0$ and $\gamma_j > 0$ for $j = 1, \dots, q$. When $\alpha = 0$, this is the Beta-Dirichlet process given in [32]. Making the change of variable $s = p.$, it is easy to check that $\mu. = \sum_{j=1}^q \mu_j$ is a stable-Beta process in the sense of [48] with Lévy density $\rho(s) = \theta s^{-\alpha-1}(1-s)^{\beta+\alpha-1}$. It follows that for each ℓ, $\mathbf{J}_{q,\ell} = (J_{1,\ell}, \dots, J_{q,\ell})$ is such that $\sum_{j=1}^q J_{j,\ell}$ has a Beta$(m_\ell - \alpha, M+\beta+\alpha-m_\ell)$ distribution just as the univariate case. Furthermore, $\mathrm{D}_{q,\ell} := (D_{1,\ell}, \dots, D_{q,\ell})$, where $D_{j,\ell} = J_{j,\ell}/\sum_{k=1}^q J_{k,\ell}$ is independent of $\sum_{k=1}^q J_{k,\ell}$ and is a Dirichlet$(m_{1,\ell}+\gamma_1, \dots, m_{q,\ell}+\gamma_q)$ vector. In addition, $\mu_{0,M}$ is a stable-Beta-Dirichlet process with parameters $(\alpha, M+\beta+\alpha; \gamma_1, \dots, \gamma_q; \theta)$.

Hence, customer $M+1$ chooses an existing dish ω_ℓ and accompanying condiment j with probability:

$$r_{j,\ell} = \frac{m_{j,\ell}+\gamma_j}{m_\ell + \sum_{k=1}^q \gamma_k} \times \frac{m_\ell - \alpha}{M+\beta}.$$

For $\tilde{Z}_0^{(M+1)}$, it follows that for $j = 1, \dots, q$, $\tilde{Z}_j^{(M+1)} \overset{d}{=} \sum_{k=1}^{\xi(\varphi_{M+1})} X_{j,k}\delta_{\tilde{\omega}_k}$, where

$$\varphi_{M+1} = \theta\int_0^1 s^{1-\alpha-1}(1-s)^{M+\beta+\alpha-1}\,ds = \frac{\theta\Gamma(1-\alpha)\Gamma(M+\beta+\alpha)}{\Gamma(M+\beta+1)}$$

and for each k, $X_{0,k} = (X_{1,k}, \dots, X_{q,k})$ is a simple multinomial with probability mass function:

$$\mathbb{I}_{\{\sum_{i=1}^q x_i = 1\}}\prod_{j=1}^q \left(\frac{\gamma_j}{\sum_{i=1}^q \gamma_i}\right)^{x_j}. \tag{5.5}$$

Thus, customer $M+1$ chooses a Poisson(φ_{M+1}) number of new dishes exactly as in the univariate case and also for each new dish chosen selects a single condiment j with probability $\gamma_j/\sum_{i=1}^q \gamma_i$, for $j = 1, \dots, q$.

REMARK 5.3. It follows that for each (M, k), $H_{0,k}$ is marginally a Beta-Dirichlet random vector with parameters $(1-\alpha, M+\beta+\alpha; \gamma_1, \dots, \gamma_q)$. This means that $H_{\cdot,k} := \sum_{j=1}^q H_{j,k}$ is a Beta$(1-\alpha, M+\beta+\alpha)$ random variable and the vector $(D_{1,k}, \dots, D_{q,k})$, for $D_{j,k} = H_{j,k}/H_{\cdot,k}$, is Dirichlet$(\gamma_1, \dots, \gamma_q)$, which leads to (5.5).

REMARK 5.4. Our analysis allows for feature specific specifications for $(q, (\gamma_j))$ by replacing them with $(q(\omega), (\gamma_j(\omega)))$ in (5.4). For selected features, one now uses $(q(\omega_\ell), (\gamma_j(\omega_\ell)))$ in the calculations. This does not alter φ_{M+1}. See also the recent work of [35] for an application of the Beta-Dricihet prior in a multivariate feature model setting.

APPENDIX: POSTERIOR ANALYSIS OF GENERALIZED INDIAN BUFFET PROCESSES

We will now apply Propositions 2.1 and 2.2 of James [25], pages 6–8 (see also [23], pages 7–8), to provide a formal derivation of the results in Propositions 3.1, 3.2 and Theorem 3.1. The result for Proposition 5.1 follows by similar arguments.

Let (J_ℓ, ω_ℓ) for $\ell = 1, \dots, K$ denote the $K := K_M$ unique points among $((\tau_k, \tilde{\omega}_k))$ picked from the Poisson random measure N, based on the conditionally i.i.d. processes $(Z_1, \dots, Z_M)$. Then one can write the joint distribution of $((Z_1, \dots, Z_M), (J_1, \dots, J_K), N)$, where $(J_1, \dots, J_K)$ are the unique jumps picked from μ with arguments $(s_1, \dots, s_K)$, as

$$\mathrm{e}^{-N(f_M)}\mathcal{P}(dN|\nu)\prod_{\ell=1}^{K} N(ds_\ell, d\omega_\ell)\prod_{i\in\mathcal{B}_\ell}\left[\frac{\pi_A(s_\ell)\tilde{G}_{A'}(da_{i,\ell}|s_\ell)}{1-\pi_A(s_\ell)}\right], \tag{A.1}$$

where $f_M(s,\omega) = -M\log(1-\pi_A(s))$, and hence $N(f_M) = -M\sum_{k=1}^{\infty}\log(1-\pi_A(\tau_k))$. Similar to the main text, for $j = 0, 1, 2, \dots$. define $\nu_j(ds, d\omega) = \mathrm{e}^{-f_j(s,\omega)}\nu(ds, d\omega) = [1-\pi_A(s)]^j\rho(s|\omega)B_0(d\omega)\,ds$. Now applying the exponential change of measure result in Proposition 2.1 of James [25], it follows that

$$\mathrm{e}^{-N(f_M)}\mathcal{P}(dN|\nu) = \mathcal{P}(dN|\nu_M)\mathrm{e}^{-\Psi(f_M)},$$

where $\mathbb{E}[\mathrm{e}^{-N(f_M)}] = \mathrm{e}^{-\Psi(f_M)}$, and

$$\Psi(f_M) = \int_\Omega\int_0^\infty \big(1-[1-\pi_A(s)]^M\big)\rho(s|\omega)\,ds\,B_0(d\omega). \tag{A.2}$$

Now since the pairs (s_ℓ, ω_ℓ) are distinct for $\ell = 1, \dots, K$, an application of Proposition 2.2 in James [25] leads to the equivalence:

$$\mathcal{P}(dN|\nu_M)\prod_{\ell=1}^{K} N(ds_\ell, d\omega_\ell) = \mathcal{P}(dN|\nu_M, \mathbf{s}, \omega)\prod_{\ell=1}^{K}\nu_M(ds_\ell, d\omega_\ell),$$

where $\nu_M(ds_\ell, d\omega_\ell) = [1-\pi_A(s_\ell)]^M\rho(s_\ell|\omega_\ell)B_0(d\omega_\ell)\,ds_\ell$, and $\mathcal{P}(dN|\nu_M, \mathbf{s}, \omega)$ denotes the distribution of the random measure $N_M + \sum_{\ell=1}^{K}\delta_{s_\ell,\omega_\ell}$, where N_M has a PRM(ν_M) distribution. Combining these results leads to an equivalent expression for (A.1), which reveals the relevant posterior and marginal quantities:

$$\mathcal{P}(dN|\nu_M, \mathbf{s}, \omega)\mathrm{e}^{-\Psi(f_M)}\prod_{\ell=1}^{K}\mathrm{S}_{\text{spike}\times\text{slab}}(s_\ell|\mathcal{A}_\ell)\rho(s_\ell|\omega_\ell)B_0(d\omega_\ell), \tag{A.3}$$

where, for $\mathcal{A}_\ell = \{A'_{i,\ell} : i \in \mathcal{B}_\ell\}$,

$$\mathrm{S}_{\text{spike}\times\text{slab}}(s|\mathcal{A}_\ell) = \big[1-\pi_A(s)\big]^{M-m_\ell}\pi_A^{m_\ell}(s)\prod_{i\in\mathcal{B}_\ell}\tilde{G}_{A'}(da_{i,\ell}|s).$$

It remains to obtain an equivalent form of $\Psi(f_M)$ in (A.2). Recall that $\rho_j(s|\omega) = [1-\pi_A(s)]^j \rho(s|\omega)$ for $j = 0, 1, 2, \ldots$. Hence, $\rho_0(s|\omega) := \rho(s|\omega)$. Define for $k = 1, 2, \ldots,$

$$\varphi_k := \int_\Omega \int_0^\infty \pi_A(s)\rho_{k-1}(s|\omega)\, ds\, B_0(d\omega). \tag{A.4}$$

Noting the expressions in (A.2) and (A.4), it follows that $\varphi_1 = \Psi(f_1)$, and it is not difficult to see that

$$\Psi(f_M) = \sum_{k=1}^M \varphi_k, \quad \text{hence} \quad \mathrm{e}^{-\Psi(f_M)} = \prod_{k=1}^M \mathrm{e}^{-\varphi_k}.$$

Substituting this expression in (A.3), leads to our final form of the joint distribution:

$$\mathcal{P}(dN|\nu_M, \mathbf{s}, \omega)\left[\prod_{k=1}^M \mathrm{e}^{-\varphi_k}\right] \prod_{\ell=1}^K \mathrm{S}_{\text{spike}\times\text{slab}}(s_\ell|\mathcal{A}_\ell)\rho(s_\ell|\omega_\ell)B_0(d\omega_\ell).$$

Integrating out N, leads to Proposition 3.1, and then Proposition 3.2 and Theorem 3.1 follow from applications of Bayes' rule. Note in the multivariate setting of Section 5, similar to (A.1), it follows that the joint distribution of N, $(\mathbf{J}_{q,\ell})$, $(Z_0^{(1)}, \ldots, Z_0^{(M)})$ can be expressed as

$$\mathrm{e}^{-N(f_M)}\mathcal{P}(dN|\nu)\prod_{\ell=1}^K N(d\mathbf{s}_{q,\ell}, d\omega_\ell) \prod_{i\in\mathcal{B}_\ell}\left[\frac{\pi_{A_0}(\mathbf{s}_{q,\ell})\tilde{G}_{A_0'}(d\mathbf{a}_{0,\ell}^{(i)}|\mathbf{s}_{q,\ell})}{1-\pi_{A_0}(\mathbf{s}_{q,\ell})}\right],$$

where now $\mathcal{B}_\ell := \{i : Z_0^{(i)}(\omega_\ell) = A_{0,\ell}^{(i)} \notin \mathbf{0}\}$ and $f_M(\mathbf{s}_q, \omega) = -M\log[1-\pi_{A_0}(\mathbf{s}_q)]$. It is evident that arguments similar to the univariate case lead to Proposition 5.1.

Acknowledgements. I would like to thank the editors and three referees for their helpful comments and assistance. The idea that the representation $A_{i,k} = b_{i,k}A'_{i,k}$ can be thought of in terms of spike and slab priors was gained from a presentation of a previous version of this work [27], by Piyush Rai in Lawrence Carin's reading group at Duke. My thanks to Wray Buntine, Stefano Favaro, Nicolas Privault and Xinghua Zheng for their comments, technical assistance and encouragement.

REFERENCES

[1] ARCHAMBEAU, C., LAKSHMINARAYANAN, B. and BOUCHARD, G. (2015). Latent IBP compound Dirichlet allocation. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 321–333.

[2] ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Mathematical Society (EMS), Zürich. MR2032426

[3] BARNDORFF-NIELSEN, O. E., PEDERSEN, J. and SATO, K. (2001). Multivariate subordination, self-decomposability and stability. *Adv. in Appl. Probab.* **33** 160–187. MR1825321

[4] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2015). Central limit theorems for an Indian buffet model with random weights. *Ann. Appl. Probab.* **25** 523–547. MR3313747

[5] BRODERICK, T., JORDAN, M. I. and PITMAN, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statist. Sci.* **28** 289–312. MR3135534

[6] 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.

[7] BRODERICK, T., PITMAN, J. and JORDAN, M. I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836.

[8] BRODERICK, T., WILSON, A. C. and JORDAN, M. I. (2017). Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli*. To appear. DOI: 10.3150/16-BEJ855.

[9] CARON, F. (2012). Bayesian nonparametric models for bipartite graphs. In *Neural Information Processing Systems* (*NIPS* 2012), Lake Tahoe, CA.

[10] DOKSUM, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201. MR0373081

[11] GERSHMAN, S. J., FRAZIER, P. I. and BLEI, D. M. (2015). Distance dependent infinite latent feature models. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 334–345.

[12] GHAHRAMANI, Z., GRIFFITHS, T. L. and SOLLICH, P. (2007). Bayesian nonparametric latent feature models. In *Bayesian Statistics* 8. 201–226. Oxford Univ. Press, Oxford. MR2433194

[13] GNEDIN, A. and PITMAN, J. (2005). Regenerative composition structures. *Ann. Probab.* **33** 445–479.

[14] GÖRUR, D., JÄKEL, F. and RASMUSSEN, C. E. (2006). A choice model with infinitely many latent features. In *Proceedings of the* 23*rd International Conference on Machine Learning* 361–368. ACM, New York.

[15] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2006). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems* 18 (*NIPS*-2005).

[16] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2011). The Indian buffet process: An introduction and review. *J. Mach. Learn. Res.* **12** 1185–1224.

[17] HEAUKULANI, C. and ROY, D. M. (2016). The combinatorial structure of beta negative binomial processes. *Bernoulli* **22** 2301–2324.

[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] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

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

[21] ISHWARAN, H. and JAMES, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *J. Amer. Statist. Assoc.* **99** 175–190. MR2054297

[22] ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158

[23] JAMES, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Unpublished manuscript. Available at arXiv:math.PR/0205093.

[24] JAMES, L. F. (2003). Bayesian calculus for gamma processes with applications to semiparametric intensity models. *Sankhyā* **65** 179–206.

[25] JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799.

[26] JAMES, L. F. (2006). Poisson calculus for spatial neutral to the right processes. *Ann. Statist.* **34** 416–440. MR2275248

[27] JAMES, L. F. (2014). Poisson latent feature calculus for generalized Indian buffet processes. Preprint. Available at arXiv:1411.2936.

[28] JAMES, L. F., ORBANZ, P. and TEH, Y. W. (2015). Scaled subordinators and generalizations of the Indian buffet process. Preprint. Available at arXiv:1510.07309.

[29] JAMES, L. F., ROYNETTE, B. and YOR, M. (2008). Generalized gamma convolutions, Dirichlet means, Thorin measures, with explicit examples. *Probab. Surv.* **5** 346–415. MR2476736

[30] KANG, H. and CHOI, S. (2013). Bayesian multi-subject common spatial patterns with Indian buffet process priors. In *ICASSP* 3347–3351. IEEE, New York.

[31] KIM, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27** 562–588. MR1714717

[32] KIM, Y., JAMES, L. and WEISSBACH, R. (2012). Bayesian analysis of multistate event history data: Beta-Dirichlet process prior. *Biometrika* **99** 127–140. MR2899668

[33] KIM, Y. and LEE, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29** 666–686. MR1865336

[34] KNOWLES, D. and GHAHRAMANI, Z. (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *Ann. Appl. Stat.* **5** 1534–1552. MR2849785

[35] LEE, J., MÜLLER, P., SENGUPTA, S., GULUKOTA, K. and JI, Y. (2014). Bayesian inference for tumor subclones accounting for sequencing and structural variants. Preprint. Available at arXiv:1409.7158 [stat.ME].

[36] LO, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. Verw. Gebiete* **59** 55–66. MR0643788

[37] LO, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. MR0733519

[38] LO, A. Y. and WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245.

[39] MILLER, K. T., GRIFFITHS, T. L. and JORDAN, M. I. (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems* (*NIPS*) 22.

[40] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

[41] PAISLEY, J. and CARIN, L. (2009). Nonparametric factor analysis with beta process priors. In *International Conference on Machine Learning* (*ICML*), Montreal, Canada.

[42] PALLA, K., KNOWLES, D. and GHAHRAMANI, Z. (2012). An infinite latent attribute model for network data. In *ICML* 2012.

[43] PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784

[44] PITMAN, J. (1997). Partition structures derived from Brownian motion and stable subordinators. *Bernoulli* **3** 79–96. MR1466546

[45] 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. IMS, Beachwood, OH. MR2004330

[46] PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002; With a foreword by Jean Picard. MR2245368

[47] RAI, P. and DAUME, H. (2009). The infinite hierarchical factor regression model. In *Advances in Neural Information Processing Systems* 1321–1328.

[48] TEH, Y. W. and GÖRÜR, D. (2009). Indian buffet processes with power-law behavior. In *NIPS* 2009.

[49] THIBAUX, R. (2008). Nonparametric Bayesian Models for Machine Learning. Doctoral dissertation, Univ. California, Berkeley.

[50] THIBAUX, R. and JORDAN, M. I. (2007). Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics* 564–571.

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

[52] WILLIAMSON, S., WANG, C., HELLER, K. and BLEI, D. (2010). The IBP compound Dirichlet process and its application to focused topic modeling. In *The* 27*th International Conference on Machine Learning* (*ICML* 2010).

[53] ZHOU, M., HANNAH, L., DUNSON, D. and CARIN, L. (2012). Beta-negative binomial process and Poisson factor analysis. AISTATS. Preprint. Available at arXiv:1112.3605.

[54] ZHOU, M., MADRID-PADILLA, O-H. and SCOTT, J. G. (2016). Priors for random count matrices derived from a family of negative binomial processes. *J. Amer. Statist. Assoc.* **111** 1144–1156.

DEPARTMENT OF INFORMATION SYSTEMS,
BUSINESS STATISTICS AND
OPERATIONS MANAGEMENT
HONG KONG UNIVERSITY OF SCIENCE
AND TECHNOLOGY
CLEAR WATER BAY, KOWLOON,
HONG KONG
E-MAIL: lancelot@ust.hk