Bayesian Analysis (2020) **15**, Number 3, pp. 683–710

Gibbs-type Indian Buffet Processes

Creighton Heaukulani[*] and Daniel M. Roy[†]

Abstract. We investigate a class of feature allocation models that generalize the Indian buffet process and are parameterized by Gibbs-type random measures. Two existing classes are contained as special cases: the original two-parameter Indian buffet process, corresponding to the Dirichlet process, and the stable (or three-parameter) Indian buffet process, corresponding to the Pitman–Yor process. Asymptotic behavior of the Gibbs-type partitions, such as power laws holding for the number of latent clusters, translates into analogous characteristics for this class of Gibbs-type feature allocation models. Despite containing several different distinct subclasses, the properties of Gibbs-type partitions allow us to develop a black-box procedure for posterior inference within any subclass of models. Through numerical experiments, we compare and contrast a few of these subclasses and highlight the utility of varying power-law behaviors in the latent features.

Keywords: feature allocation, partition, combinatorial stochastic processes, completely random measure, Bayesian nonparametrics.

1 Introduction

Feature allocation models (Ghahramani et al., 2007; Broderick et al., 2013) assume that data are grouped into a collection of possibly overlapping subsets, called *features*. The best known example is the *Indian buffet process* (IBP) (Griffiths and Ghahramani, 2006; Ghahramani et al., 2007), which has been successfully applied to a number of unsupervised learning problems in which the features represent unobserved/latent factors underlying the data. While the IBP provides a nonparametric distribution suited to learning an appropriate number of features from the data, additional modeling flexibility—like heavy-tailed (i.e., power law) behavior in the number of latent features—is desirable in many applications. Recent generalizations of the IBP addressing these needs parallel existing developments in the theory of random partitions. Indeed, random feature allocations may be viewed as a generalization of random partitions where the subsets of the partition are allowed to overlap. In recent work, Roy (2014) defines a broad class of random feature allocations called the *generalized Indian buffet process*, each member of which corresponds to the law of an exchangeable partition. In this article, we study the subclass corresponding to the random *Gibbs-type partitions* (Gnedin and Pitman, 2006), which we call the *Gibbs-type Indian buffet process* or simply *Gibbs-type IBP*. The Gibbs-type IBP inherits many useful properties from the Gibbs-type partitions (which include many of the partitioning models studied in the literature), and the special form of these models will allow us to develop practical black-box algorithms for simulation and posterior inference.

[*] University of Cambridge, Cambridge, United Kingdom, c.k.heaukulani@gmail.com

[†] University of Toronto, Toronto, Canada, droy@utstat.toronto.edu

https://doi.org/10.1214/19-BA1166

1.1 Exchangeable feature allocations and the IBP

In the terminology introduced by Broderick et al. (2013), a *feature allocation* of a set A is a multiset of nonempty subsets of A, called *features*, with the further restriction that no element of A belongs to infinitely many features. In statistical applications, we usually take A to be the set $[n] := \{1, \dots, n\}$ for some $n \geq 1$, where A then indexes a sequence of n data points. Intuitively, a random feature allocation of $[n]$ can be used to model n data points in terms of latent features the data points share. Note that a data point may have multiple features, and so this notion generalizes clustering.

In many applications, we do not know the number of latent features necessary to adequately model a data set. In such cases, one requires a nonparametric model in which the number of latent features is a random variable to be inferred from the data. The canonical example of such a model is the Indian buffet process (IBP), introduced by Griffiths and Ghahramani (2006); Ghahramani et al. (2007). An IBP is a random feature allocation with an a priori unbounded number of potential latent features whose construction can be explained with the following culinary analogy: Imagine a sequence of customers entering an Indian buffet restaurant. Each customer selects a finite number of dishes, chosen from a limitless supply of potential dishes to taste. The first customer enters the buffet and takes Poisson(γ) dishes, where $\gamma > 0$ is called the *mass parameter*. For every $n \geq 1$, the $n+1$-st customer enters the buffet and decides to take each previously tasted dish k with probability $n_k/(n+\theta)$, where n_k is the number of previous customers that took dish k, and where $\theta > 0$ is called the *concentration parameter*. The customer then takes Poisson($\theta\gamma/(\theta + n)$) new (previously untasted) dishes. For every $n \geq 1$, let K_n denote the number of distinct dishes tasted among the first n customers, and let $F_n := \{F_{n,1}, \dots, F_{n,K_n}\}$, where $F_{n,1}, F_{n,2}, \dots$ are random subsets of $[n]$ such that $i \in F_{n,k}$ if and only if the i-th customer took the k-th dish, for every $i \leq n$ and $k \leq K_n$. By construction, for every $n \geq 1$, F_n is a random feature allocation of $[n]$, and the sequence $F := (F_n)_{n \geq 1}$ defines a random feature allocation of $\mathbb{N} := \{1, 2, \dots\}$. We call F an *Indian buffet process with mass parameter* γ *and concentration parameter* θ.

Ghahramani et al. (2007) show that, for every $n \geq 1$, the distribution of F_n is invariant to every permutation of $[n]$, i.e., the order of the customers does not influence the distribution of the resulting feature allocation. A feature allocation with this property is called *exchangeable* in analogy to exchangeable sequences, which satisfy a related family of distributional invariance properties. Indeed, much of the recent work on exchangeable feature allocations has been inspired by analogous work in the theory of exchangeable partitions. In statistical applications, random feature allocations have been applied to many of the same clustering problems as random partitions, where the extra flexibility of overlapping cluster assignments has often resulted in improved modeling power.

1.2 The Gibbs-type IBP

The Gibbs-type Indian buffet process, or Gibbs-type IBP, defines a class of exchangeable feature allocations that generalizes the IBP. Let $\alpha < 1$, which we will call the *discount parameter*, and let $\overline{V} := (V_{n,k} \colon n \geq k \geq 1)$ be a triangular array of non-negative weights satisfying $V_{1,1} = 1$ and the recursive equations

$$V_{n,k} = (n - \alpha k) V_{n+1,k} + V_{n+1,k+1}, \qquad n \geq k \geq 1. \tag{1.1}$$

(In the cases we will study, the weights $\overline{V}$ themselves are determined by a finite set of parameters, which we will denote by Θ.) Define the primitives

$$Q^n_{\alpha,\Theta}(z_1, z_2) := \sum_{k=1}^{n} \frac{V_{n+z_1,k+z_2}}{\alpha^k} \mathscr{C}(n,k;\alpha), \qquad n \geq z_1 \geq 1,\ z_2 \in \{0,1\}, \tag{1.2}$$

where $\mathscr{C}(n,k;\alpha)$ denotes the generalized factorial coefficient

$$\mathscr{C}(n,k;\alpha) := \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (-i\alpha)_n, \qquad n \geq k \geq 1, \tag{1.3}$$

and $(a)_n := \Gamma(a+n)/\Gamma(a)$. (See Charalambides (2005) for a background on the generalized factorial coefficients.) Then the Gibbs-type IBP may be described as follows: Let $\gamma > 0$, and imagine a sequence of customers entering an Indian buffet restaurant.

- The first customer tries Poisson(γ) dishes from the buffet.
- For every $n \geq 1$, the $n+1$-st customer
  - tries each previously tasted dish k independently with probability

    $$(S_{n,k} - \alpha) Q^n_{\alpha,\Theta}(1,0),$$

    where $S_{n,k}$ is the number among the first n customers that tried dish k;
  - and tries Poisson($\gamma Q^n_{\alpha,\Theta}(1,1)$) new dishes from the buffet.

Construct a random feature allocation F of $\mathbb{N}$ from the actions of the customers, as described in Section 1.1. We call F a *Gibbs-type Indian buffet process with parameters* $(\gamma, \alpha, \overline{V})$. Like the original IBP, F is exchangeable, a property that will become clear in Section 3 when we provide an alternative construction via exchangeable sequences of random measures.

The reader familiar with the theory of Gibbs-type partitions (which we review in Section 2) will recognize the recursive set of weights $\overline{V}$ appearing in Equation (1.1), which along with the discount parameter α determines the law of a *Gibbs-type partition* (Gnedin and Pitman, 2006). In what follows, we will see that every such choice $(\alpha, \overline{V})$ defining a subclass of the Gibbs-type partitions will determine a subclass of the Gibbs-type IBP. In fact, some subclasses of Gibbs-type IBPs have already appeared in the literature, although they have not been presented from this perspective. For example, the *stable* (or *three-parameter*) IBP introduced by Teh and Görür (2009) and further studied by Broderick et al. (2012) is a Gibbs-type IBP with the weights

$$V_{n,k} = \frac{\prod_{\ell=1}^{k-1} (\theta + \ell\alpha)}{(\theta+1)_{n-1}}, \qquad n \geq k \geq 1, \tag{1.4}$$

for some parameter θ satisfying

$$\begin{cases} \theta > -\alpha, & \text{when } \alpha \in [0,1), \\ \theta = m|\alpha| \text{ for some } m \in \{1,2,\dots\}, & \text{when } \alpha < 0. \end{cases} \tag{1.5}$$

This setting of $(\alpha, \overline{V})$ corresponds to a subclass of the Gibbs-type partitions known as the *two-parameter Chinese Restaurant processes*, i.e., the random partitions induced by the pattern of ties in exchangeable sequences sampled from a *Pitman–Yor process* (Perman et al., 1992; Pitman and Yor, 1997). (We will discuss the connection between exchangeable partitions and random probability measures in Section 2.) In this case, we have $\Theta = \{\theta\}$ and the quantities $Q^n_{\alpha,\Theta}(1,0)$ and $Q^n_{\alpha,\Theta}(1,1)$ reduce to

$$Q^n_{\alpha,\Theta}(1,0) = \frac{1}{\theta+n} \quad \text{and} \quad Q^n_{\alpha,\Theta}(1,1) = \frac{\Gamma(\theta+1)\Gamma(\theta+\alpha+n)}{\Gamma(\theta+n+1)\Gamma(\theta+\alpha)}, \tag{1.6}$$

respectively. For $\alpha = 0$ and $\theta > 0$, we obtain the (two-parameter) IBP reviewed in Section 1.1, and for $\alpha = 0$ and $\theta = 1$, the corresponding Gibbs-type IBP reduces to a more restrictive one-parameter variant of the IBP originally presented by Griffiths and Ghahramani (2006). In short, the stable IBP is the feature allocation analogue to the two-parameter Chinese Restaurant process, and the two-parameter IBP is the analogue to the one-parameter Chinese Restaurant process.

1.3 Outline and summary of results

In Section 2, we review the theory of exchangeable Gibbs-type partitions, focusing on a few important subclasses. In Section 3, we derive the Gibbs-type IBP from a construction with completely random measures. As an intermediate step, we define the *Gibbs-type beta process*, a completely random measure that generalizes the *beta process* introduced by Hjort (1990). We present stick-breaking constructions for the Gibbs-type beta process that generalize similar representations in the literature for the beta and stable beta processes (Teh et al., 2007; Paisley et al., 2010; Teh and Görür, 2009; Broderick et al., 2012; Paisley et al., 2011). While these constructions are special cases of the *generalized beta process* and corresponding *generalized IBP* defined by Roy (2014), the special form of the Gibbs-type partitions will allow us to additionally derive practical algorithms for simulation and posterior inference with the Gibbs-type IBP.

Partitions with Gibbs-type structure exhibit many properties that are useful for applications. For example, when the discount parameter α is in $(0,1)$, a Gibbs-type partition exhibits heavy-tailed (i.e., power law) behavior in the asymptotic distribution of the number of clusters induced by the partition. Latent features in the stable IBP were shown to exhibit analogous power-law behavior (Teh and Görür, 2009; Broderick et al., 2012), and in Section 5 we show that these characteristics are in a sense inherited from the two-parameter Chinese Restaurant Process or, equivalently, the Pitman–Yor process (with $\alpha \in (0,1)$). More generally, our results show that the Gibbs-type IBP inherits these power-law properties for any such class of partitioning models. Similarly, when $\alpha < 0$, the Gibbs-type partitions correspond to models with a random but finite number of clusters, and in Section 5.3 we show that the Gibbs-type IBP in this case corresponds to models with a random but finite number of features.

Many computations of interest with Gibbs-type partitions are expressed only through the parameters $(\alpha, \overline{V})$. Likewise, the primitives $Q^n_{\alpha,\Theta}(z_1, z_2)$ in Equation (1.2) only depend on these quantities. Note that the description of the Gibbs-type IBP in the previous section only requires the arguments $Q^n_{\alpha,\Theta}(1,1)$ and $Q^n_{\alpha,\Theta}(1,0)$. These quantities have probabilistic interpretations and are related to the well-studied probabilities of sampling a new and previous *color* (or *species*) under the law of a Gibbs-type partition. (These concepts will be made clear in Section 2.) A likelihood function for the Gibbs-type IBP will be presented in Section 3, which additionally requires the arguments $Q^{n-s}_{\alpha,\Theta}(s,1)$ for $s \leq n$. These terms also have probabilistic interpretations related to events in a Gibbs-type partition, which are all discussed in the supplementary material (Heaukulani and Roy, 2019). In Section 6, we derive a black-box posterior inference procedure that only requires these $n+1$ values of the primitives as input. Finally, in Section 7 we demonstrate some of the practical differences between a few subclasses of the Gibbs-type IBP in a Bayesian nonparametric latent feature model applied to synthetic data and the classic MNIST digits dataset.

2 Exchangeable Gibbs-type partitions

We briefly review the theory of Gibbs-type partitions; the reader should consult Gnedin and Pitman (2006) for a more thorough treatment and Pitman (2002, Chs. 2 & 3) for background on exchangeable partitions more generally. Let Π be a random partition of $\mathbb{N} := \{1, 2, \dots\}$ into disjoint subsets, called *blocks*. We may write $\Pi = \{A_1, A_2, \dots\}$, where A_1 is the block containing 1 and A_{k+1}, for every $k \geq 1$, is the (possibly empty) block containing the least integer not in $A_1 \cup \dots \cup A_k$. For every $n \geq 1$, let Π_n be the restriction of Π to $[n] := \{1, \dots, n\}$. For every $n \geq k \geq 1$, let $N_{n,k}$ be the number of elements in $A_k \cap [n]$, and let B_n be the number of (nonempty) blocks in Π_n. The partition Π_n is said to be *exchangeable* when its distribution is invariant under every permutation of the underlying set $[n]$ and Π is said to be exchangeable when every restriction Π_n, for $n \geq 1$, is exchangeable.

The random partition Π is of *Gibbs-type* when it is exchangeable and, for some $\alpha < 1$ and $V_{n,k} \geq 0$, $n \geq k \geq 1$ satisfying Equation (1.1), we have

$$f_\Pi(n_1, \dots, n_k) := \mathbb{P}(B_n = k, N_{n,1} = n_1, \dots, N_{n,k} = n_k) \tag{2.1}$$

$$= V_{n,k} \prod_{\ell=1}^{k} (1-\alpha)_{n_\ell - 1}, \tag{2.2}$$

for every $n \geq k \geq 1$ and $n_1, \dots, n_k \geq 1$ satisfying $\sum_j n_j = n$. The function $f_\Pi(n_1, \dots, n_k)$, which is symmetric by exchangeability, is called the *exchangeable partition probability function*, or EPPF. The class of Gibbs-type partitions was introduced by Gnedin and Pitman (2006) and has since been the subject of intense study due, in part, to the fact that the product form of the Gibbs-type EPPF in Equation (2.2) admits closed-form solutions for many quantities of interest.

An exchangeable partition can be related to the pattern of colored balls drawn from an urn in a sequence of rounds as follows: On each round, we may either (1) draw a ball from the urn at random, record the color, and place the ball back into the urn with another ball of the same color, or (2) we may place a ball of a new, previously unseen color into the urn. The distinct colors of the balls correspond to the blocks in Π, and the indices of the rounds during which a particular color was drawn indicates the members of the corresponding block. In particular, on the first round the urn is empty and a ball of a new color is placed into the urn creating $B_1 = 1$ block. We see from Equation (2.2) that during the $n+1$-st round, we draw a ball of the k'th previously seen color from the urn with probability

$$\begin{aligned}\mathbb{P}[N_{n+1,k} > N_{n,k} | B_n, N_{n,1}, N_{n,2}, \dots] &= \frac{f_\Pi(N_{n,1}, \dots, N_{n,k}+1, \dots, N_{n,B_n})}{f_\Pi(N_{n,1}, \dots, N_{n,B_n})} \\ &= \frac{V_{n+1,B_n}}{V_{n,B_n}}(N_{n,k} - \alpha),\end{aligned} \tag{2.3}$$

for every $k \leq B_n$, where $N_{n,k}$ denotes the size of the k-th block at the end of the n-th round. We draw a ball of a new color with probability

$$\begin{aligned}\mathbb{P}[B_{n+1} > B_n | B_n, N_{n,1}, N_{n,2}, \dots] &= \frac{f_\Pi(N_{n,1}, \dots, N_{n,B_n}, 1)}{f_\Pi(N_{n,1}, \dots, N_{n,B_n})} \\ &= \frac{V_{n+1,B_n+1}}{V_{n,B_n}}.\end{aligned} \tag{2.4}$$

Gnedin and Pitman (2006, Section 2) show that the distribution of the number of blocks after the n'th round is given by

$$\mathbb{P}(B_n = k) = \frac{V_{n,k}}{\alpha^k}\mathscr{C}(n, k; \alpha), \qquad k \leq n, \tag{2.5}$$

where $\mathscr{C}(n, k; \alpha)$ is the generalized factorial coefficient given in Equation (1.3).

The theory of exchangeable partitions is intimately connected to the theory of random probability measures. In particular, by a representation theorem due to Kingman (1978), every exchangeable partition may be obtained from the ties among an exchangeable sequence sampled from a random probability measure, and the laws of the partition and measure are one-to-one. The measures inducing the Gibbs-type partitions are called *Gibbs-type random measures*. (The reader should consult Kingman (1975) for background on random probability measures.) We will focus on subclasses of the Gibbs-type partitions induced by several random probability measures that have been well-studied in the literature. For example, the class of Gibbs-type random measures include the Dirichlet and Pitman–Yor processes already mentioned in the introduction. Another subclass we will refer to frequently are those induced by the *normalized generalized gamma processes* (Pitman, 2003), which have the weights

$$V_{n,k} = \frac{e^\beta \alpha^{k-1}}{\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k - i/\alpha; \beta), \tag{2.6}$$

where $\alpha \in (0,1)$, $\beta > 0$, and $\Gamma(x;a) := \int_x^\infty s^{a-1}e^{-s}\mathrm{d}s$ is the incomplete gamma function. Special cases include the partitions induced by the *normalized inverse Gaussian processes* (Lijoi et al., 2005) when $\alpha = 1/2$; the *normalized α-stable processes* (Kingman, 1975) in the limit $\beta \to 0$; and the Dirichlet processes, again, in the limit $\alpha \to 0$.

More generally, Gnedin and Pitman (2006, Theorem 12) showed that the law of every Gibbs-type partition with fixed discount parameter $\alpha < 1$ is a unique probability mixture over one of three classes of extreme partitions, depending on the value of α. When $\alpha \in (0,1)$, the extreme partitions are induced by the *Poisson–Kingman random measures* (Pitman, 2003); in this case, it follows from Pitman (2003, Proposition 9), that

$$V_{n,k} = \frac{\alpha^k}{\Gamma(n-k\alpha)\tau^{k\alpha}f_\alpha(\tau)} \int_0^1 p^{n-k\alpha-1} f_\alpha(\tau(1-p))\mathrm{d}p, \tag{2.7}$$

for a parameter $\tau > 0$, where f_α is the density of a positive α-stable random variable. Members of this subclass are obtained by mixing over τ with respect to a probability distribution on the positive real numbers. Particular attention has been paid to the cases when τ has density function $h(t)f_\alpha(t)$, for some measureable function $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. For example, when $h(t) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)}t^{-\theta}$ for some $\theta > -\alpha$, then Equation (2.7) reduces to Equation (1.4) and we obtain the partitions induced by the Pitman–Yor processes (with $\alpha \in (0,1)$). When $h(t) = e^{\beta^\alpha - \beta t}$ for some $\beta > 0$, then Equation (2.7) reduces to Equation (2.6) and we obtain the partitions induced by the normalized generalized gamma processes. See Pitman (2003, Section 5) for a further treatment. When $\alpha = 0$, the extreme partitions are induced by the Dirichlet processes with concentration parameter $\theta > 0$. Members of the subclass are obtained by mixing over a random θ with respect to a probability distribution on $\mathbb{R}_{>0}$. Finally, when $\alpha < 0$, the extreme partitions are induced by the Pitman–Yor processes with concentration parameter $\theta = m|\alpha|$, for some positive integer m. In this case, the weights in Equation (1.4) may be rewritten as

$$V_{n,k} = \frac{|\alpha|^{k-1}\prod_{\ell=1}^{k-1}(m-\ell)}{(m|\alpha|+1)_{n-1}} 1_{\{1,\dots,\min(n,m)\}}(k), \qquad n \ge k \ge 1, \tag{2.8}$$

highlighting the restriction on the weights $\overline{V}$ to be non-negative. This is equivalent to an urn scheme with a finite number m of different colors (Pitman, 2002, Chapter 3, Section 2). Members of the subclass are obtained by mixing over a random m with respect to a probability distribution on the positive integers.

In summary, each Gibbs-type partition with fixed $\alpha < 1$ is a unique probability mixture over the extreme partitions induced by either:

1. The Pitman–Yor processes with discount parameter α and concentration parameter $\theta = m|\alpha|$ for m in $\mathbb{N}$, when $\alpha < 0$;
2. The Dirichlet processes with concentration parameter $\theta > 0$, when $\alpha = 0$; or
3. The Poisson–Kingman processes with parameter $\tau > 0$, when $\alpha \in (0,1)$.

It should be clear that each Gibbs-type partition defines a Gibbs-type IBP via the construction in Section 1.2, and so it will suffice to characterize the Gibbs-type IBP in each of these regimes. In order to perform posterior inference within each subclass of the Gibbs-type IBP, we will place prior distributions on any parameters defining the weights $\overline{V}$ and infer their values from data (see Section 6).

3 Constructions from random measures

Thibaux and Jordan (2007) connected exchangeable feature allocations with the theory of completely random measures by showing that the IBP captures the combinatorial structure of an exchangeable sequence of Bernoulli processes directed by a beta process (Hjort, 1990). Generalizations of this approach have appeared in the literature, which, from our perspective, include generalizations of the IBP that are parameterized by the law of the Pitman–Yor processes (Teh and Görür, 2009; Broderick et al., 2012) and, more generally, by the law of any exchangeable partition (Roy, 2014). Here we describe the case corresponding to the Gibbs-type partitions.

3.1 Gibbs-type beta processes

Let Π be the exchangeable Gibbs-type partition defined by Equation (2.2), whose restriction Π_n to $[n]$ has block sizes (in order of appearance; see Section 2) denoted by $N_{n,1}, N_{n,2}, \ldots$, for every $n \geq 1$. By Kingman's paint-box construction (Kingman, 1978), the limiting relative frequencies of the blocks

$$P_k := \lim_{n \to \infty} \frac{N_{n,k}}{n} \tag{3.1}$$

exist almost surely for every $k \in \mathbb{N}$. Let μ_1 be the distribution of P_1, which is called the *structural distribution*. The structural distribution reveals quite a bit about the exchangeable partition, but does not necessarily characterize it (Pitman, 2002, Chapter 2.3); (Pitman, 1995). The structural distribution will entirely determine the law of the corresponding Gibbs-type IBP. Let Ω be a complete, separable metric space and let $\mathcal{A}$ be its Borel σ-algebra. Following Roy (2014, Theorem 1.2), define a purely atomic random measure B on $(\Omega, \mathcal{A})$ by

$$B := \sum_{k \geq 1} \tilde{b}_k \delta_{\tilde{\omega}_k}, \tag{3.2}$$

where $(\tilde{\omega}_1, \tilde{b}_1), (\tilde{\omega}_2, \tilde{b}_2), \ldots$ are the points of a Poisson process on $\Omega \times (0, 1]$ with (σ-finite) intensity measure

$$\nu_\Pi(\mathrm{d}\omega \times \mathrm{d}p) := B_0(\mathrm{d}\omega)\, p^{-1} \mu_1(\mathrm{d}p), \tag{3.3}$$

for some non-atomic σ-finite measure B_0 on $(\Omega, \mathcal{A})$. Note that, because ν_Π is not a finite measure, B will have a countably infinite number of atoms, almost surely. We call B a *Gibbs-type beta process with EPPF* f_Π *and base measure* B_0. Also note that the construction of B ensures that the random variables $B(A_1), \ldots, B(A_k)$ are independent for every finite, disjoint collection $A_1, \ldots, A_k \in \mathcal{A}$, and B is therefore said to be *completely* *random* or have *independent increments*. (See Kallenberg (2002, Chapter 12) for a background on completely random measures.) Following Thibaux and Jordan (2007), define a sequence $(Z_n)_{n\in\mathbb{N}} := (Z_1, Z_2, \dots)$ of random measures on $(\Omega, \mathcal{A})$ that are conditionally i.i.d., given B, with

$$Z_n = \sum_{k\geq 1} 1_{\{U_{n,k}<\tilde{b}_k\}} \delta_{\tilde{\omega}_k}, \qquad n \in \mathbb{N}, \tag{3.4}$$

where $(U_{n,k})_{n,k\in\mathbb{N}}$ is an independent collection of i.i.d. Uniform$(0,1)$ random variables. Then $(Z_n)_{n\in\mathbb{N}}$ is an exchangeable sequence of Bernoulli processes. By construction, because B is completely random, the elements of $(Z_n)_{n\in\mathbb{N}}$ are completely random, both conditionally on B, and unconditionally.

Fix $n \geq 1$. We now describe the conditional distribution of Z_{n+1} given $Z_{[n]} := (Z_1, \dots, Z_n)$. A rigorous derivation can be found in Roy (2014) and James (2017, Proposition 3.1). The following exposition emphasizes intuition, and follows the approach of Teh and Görür (2009) and Thibaux and Jordan (2007), who built on the work of Kim (1999, Theorem 3.3). By the complete randomness of Z_{n+1}, we may first analyze the conditional distribution of its *fixed atoms* (that is, any atoms that have also appeared among $Z_{[n]}$), followed by the conditional distribution of its *ordinary component* (which consists of atoms that have not appeared among $Z_{[n]}$). Consider the fixed atoms: That is, let $(\omega_1, \dots, \omega_{K_n})$ be the K_n distinct atoms among $Z_{[n]}$, listed in order of appearance (i.e., the order in which they first appear in the sequence, with ties broken uniformly at random and independently). We can relate these distinct atoms to the atoms in B: we have $(\omega_1, \dots, \omega_{K_n}) = (\tilde{\omega}_{j_1}, \dots, \tilde{\omega}_{j_{K_n}})$ for some random integers $(j_1, \dots, j_{K_n})$. For $k \leq K_n$, the measure Z_{n+1} takes atom ω_k with some probability $b_k = \tilde{b}_{j_k}$, and the conditional distribution of b_k, given $Z_{[n]}$, is

$$\mathbb{P}[b_k \in \mathrm{d}p \mid Z_1, \dots, Z_n] = \frac{p^{S_{n,k}-1}(1-p)^{n-S_{n,k}}\mu_1(\mathrm{d}p)}{g(n, S_{n,k})}, \tag{3.5}$$

where $S_{n,k} := \sum_{j=1}^{n} Z_j(\{\omega_k\})$, for $k \leq K_n$, and

$$g(n,s) := \int_{(0,1]} p^{s-1}(1-p)^{n-s}\mu_1(\mathrm{d}p), \qquad n \geq s \geq 1. \tag{3.6}$$

Therefore, for every $k \leq K_n$, we have

$$\begin{align} \mathbb{P}[Z_{n+1}(\{\omega_k\}) = 1 \mid Z_1, \dots, Z_n] &= \mathbb{E}[b_k \mid Z_1, \dots, Z_n] \tag{3.7} \\ &= \frac{g(n+1, S_{n,k}+1)}{g(n, S_{n,k})}. \tag{3.8} \end{align}$$

Now consider the ordinary component: Informally speaking, the distribution of the atoms of Z_{n+1} that have not appeared among $Z_{[n]}$ may be described as follows: for some infinitesimal set $\mathrm{d}\omega \subseteq \Omega \setminus \{\omega_1, \dots, \omega_{K_n}\}$,

$$\begin{align} \mathbb{P}[Z_{n+1}(\mathrm{d}\omega) = 1 \mid Z_1, \dots, Z_n] &= \int_{(0,1]} p(1-p)^n \nu_\Pi(\mathrm{d}\omega \times \mathrm{d}p) \tag{3.9} \\ &= B_0(\mathrm{d}\omega) g(n+1, 1). \tag{3.10} \end{align}$$

More precisely, on $\Omega \setminus \{\omega_1, \dots, \omega_{K_n}\}$, the measure Z_{n+1} is a Poisson process with intensity measure $g(n+1, 1)B_0$, and the number of new atoms in Z_{n+1} is therefore Poisson distributed with rate $\gamma g(n+1, 1)$, where $\gamma := B_0(\Omega) < \infty$.

3.2 Exchangeable feature allocations of Gibbs-type

We now construct an exchangeable feature allocation from the exchangeable sequence $(Z_n)_{n\in\mathbb{N}}$. Recall the buffet process analogy introduced in Sections 1.1 and 1.2. Let $n \geq 1$. For every $i \leq n$, associate the Bernoulli process Z_i with the i-th customer entering the Indian buffet restaurant, and associate the K_n distinct atoms $(\omega_1, \dots, \omega_{K_n})$ among $Z_{[n]}$ with the distinct dishes sampled among the first n customers, where the dishes are listed in order of appearance, as described earlier. Then K_n represents the total number of dishes taken by the first n customers, and $S_{n,k}$ is the number of customers, among the first n customers, that sampled dish k. Let $F_{n,1}, F_{n,2}, \dots$ be random subsets of $[n]$ such that, for every $i \leq n$ and $k \leq K_n$, we have $i \in F_{n,k}$ if and only if $Z_i(\{\omega_k\}) = 1$. It is easy to verify that $F_n := \{F_{n,1}, \dots, F_{n,K_n}\}$ is a random feature allocation of $[n]$, and $F := (F_n)_{n\in\mathbb{N}}$ is a random feature allocation of $\mathbb{N}$. Because the sequence $(Z_n)_{n\in\mathbb{N}}$ is exchangeable, it follows that F is an exchangeable feature allocation of $\mathbb{N}$. Note that F captures only the *combinatorial structure* of the sequence $(Z_n)_{n\in\mathbb{N}}$—that is, the pattern of shared atoms among the elements of the sequence $(Z_n)_{n\in\mathbb{N}}$, ignoring the locations of the atoms themselves—analogously to the way exchangeable partitions only capture the combinatorial structure of exchangeable sequences of random variables directed by a random probability measure.

We will now show that the law of $(Z_n)_{n\in\mathbb{N}}$, and therefore the induced feature allocation F, is characterized by the Gibbs-type IBP presented in Section 1.2. In particular, we will show the probability the $n+1$-st customer takes a previously sampled dish agrees with the probability that each atom in $Z_{[n]}$ appears in Z_{n+1} (in Equation (3.8)), and the mean of the Poisson distributed number of new dishes taken by the $n+1$-st customer agrees with the mean of the Poisson distributed number of new atoms in Z_{n+1}. To that end, it suffices to study the triangular array of integrals $g(n, s)$, for $n \geq s \geq 1$. The structural distribution μ_1 relates the Gibbs-type beta process B to the probabilities of combinatorial events in the exchangeable partition Π_n. In particular, we have

$$g(n, s) = \mathbb{P}(N_{s,1} = s \cap N_{n,1} = s) = \mathbb{P}(N_{n,B_{n-s+1}} = s). \tag{3.11}$$

To understand these identities, we return to the urn scheme interpretation, described in Section 2, which we recall is initialized by placing a colored ball into the urn. From Equation (3.1), we may informally interpret the structural distribution μ_1 as the (asymptotic) probability of drawing this color from the urn in subsequent rounds of the scheme. We may therefore interpret the definition of $g(n, s)$ in Equation (3.6) as the probability of drawing this color in the first s rounds of the urn scheme, followed by not drawing it again in the following $n - s$ rounds, resulting in the first equality in Equation (3.11). The second equality follows by exchangeability, i.e., we may reorder the first s draws from the urn scheme to instead be the last s draws without affecting this probability. Formal derivations of such formulae can be obtained with properties of the structural distribution, as discussed by Pitman (1995); Pitman (2002, Section 2.3).

Clearly $g(1,1) = 1$. Consider $g(n+1,1) = \mathbb{P}(N_{n+1,B_{n+1}} = 1) = \mathbb{P}(B_{n+1} > B_n)$. This is the probability that a new color is drawn on the $n+1$-st round, which conditioned on B_n is given by $V_{n+1,B_n+1}/V_{n,B_n}$ (see Equation (2.4)). Then by taking an expectation over B_n (with respect to Equation (2.5)), we have for every $n \geq 1$,

$$\mathbb{P}(B_{n+1} > B_n) = \mathbb{E}\Big[\frac{V_{n+1,B_n+1}}{V_{n,B_n}}\Big] = \sum_{k=1}^{n}\Big(\frac{V_{n+1,k+1}}{\alpha^k}\mathscr{C}(n,k;\alpha)\Big) = Q^{\,n}_{\alpha,\Theta}(1,1),$$

where we recall that $Q^{\,n}_{\alpha,\Theta}(\cdot,\cdot)$ was given by Equation (1.2). This is the mean number of new dishes tasted by the $n+1$-st customer in the Gibbs-type IBP, as desired. In general, $g(n,s) = \mathbb{P}\{N_{n,B_{n-s+1}} = s\}$ is the probability that a new color is drawn on the $(n-s+1)$-st iteration and then drawn again $s-1$ times in a row. Conditioned on B_{n-s}, sampling a new color occurs with probability $V_{n-s+1,B_{n-s}+1}/V_{n-s,B_{n-s}}$, and drawing this color $s-1$ additional times occurs with probability

$$\begin{aligned}&\frac{V_{n-s+2,B_{n-s}+1}}{V_{n-s+1,B_{n-s}+1}}(1-\alpha)\frac{V_{n-s+3,B_{n-s}+1}}{V_{n-s+2,B_{n-s}+1}}(2-\alpha)\cdots\frac{V_{n,B_{n-s}+1}}{V_{n-1,B_{n-s}+1}}(s-1-\alpha)\\&\qquad=\frac{V_{n,B_{n-s}+1}}{V_{n-s+1,B_{n-s}+1}}(1-\alpha)_{s-1}.\end{aligned}\tag{3.12}$$

Multiplying, we have

$$\mathbb{P}[N_{n,B_{n-s+1}} = s \mid B_{n-s}] = \frac{V_{n,B_{n-s}+1}}{V_{n-s,B_{n-s}}}(1-\alpha)_{s-1}.\tag{3.13}$$

With an iterated expectation and the form of Equation (3.13), we may write

$$\begin{aligned}\frac{g(n+1,s+1)}{g(n,s)} &= \mathbb{E}\Big[\frac{\mathbb{P}[N_{n+1,B_{n-s+1}} = s+1 \mid B_{n-s}]}{\mathbb{P}[N_{n,B_{n-s+1}} = s \mid B_{n-s}]}\Big] &(3.14)\\ &= (s-\alpha)\mathbb{E}\Big[\frac{V_{n+1,B_{n-s}+1}}{V_{n,B_{n-s}+1}}\Big]. &(3.15)\end{aligned}$$

Recall that, on the event $\{N_{n,B_{n-s+1}} = s\}$, we have $B_{n-s}+1 = B_{n-s+1} = B_n$. Therefore,

$$(s-\alpha)\mathbb{E}\Big[\frac{V_{n+1,B_n}}{V_{n,B_n}}\Big] = (s-\alpha)\sum_{k=1}^{n}\frac{V_{n+1,k}}{\alpha^k}\mathscr{C}(n,k;\alpha) = (s-\alpha)Q^{\,n}_{\alpha,\Theta}(1,0),\tag{3.16}$$

which shows that Equation (3.8) is the probability the $n+1$-st customer in the Gibbs-type IBP tastes a dish that has been tasted s times previously. Finally, let b_0 denote a probability density function for the normalized base measure $\gamma^{-1}B_0$. In the supplementary material, we show that $(Z_1,\dots,Z_n)$ has a probability density function p_n given by

$$\begin{aligned}p_n(Z_1,\dots,Z_n) = \gamma^{K_n}\exp\Big(-\gamma\sum_{j=1}^{n}Q^{\,j-1}_{\alpha,\Theta}(1,1)\Big)&\\ \times\prod_{k=1}^{K_n}\big[(1-\alpha)_{S_{n,k}-1}Q^{\,n-S_{n,k}}_{\alpha,\Theta}(S_{n,k},1)b_0(\omega_k)\big],&\end{aligned}\tag{3.17}$$

where $Q^{\,0}_{\alpha,\Theta}(n,1) := (1-\alpha)_{n-1}V_{n,1}$ for every $n \geq 1$.

3.3 Special cases

Clearly, any EPPF of the Gibbs-type form in Equation (2.2) will induce a Gibbs-type IBP. Some special cases of these constructions are already known in the literature. We have already discussed the Gibbs-type IBPs corresponding to partitions induced by the Pitman–Yor (and, thus, Dirichlet) processes. Indeed, in the Pitman–Yor process case the structural distribution is $\mu_1 = \text{beta}(1-\alpha, \theta+\alpha)$ for $\alpha \in [0,1)$ and $\theta > -\alpha$. In this case, the Gibbs-type beta process specializes to the *stable* (or *three-parameter*) *beta process* (Teh and Görür, 2009), which contains the original beta process when $\alpha = 0$. Despite those authors not studying the case when $\alpha < 0$ and $\theta = m|\alpha|$, for some m in $\mathbb{N}$, we may just as well define this extension of the stable beta process and stable IBP, and, indeed, the structural distribution (and so the construction of B and $(Z_n)_{n\in\mathbb{N}}$) are of the same form. See (Pitman, 1995, Proposition 9 and the text following) for references on the structural distributions in all of these cases.

As described at the end of Section 2, the only remaining case of the Gibbs-type IBPs to consider are those corresponding to the Gibbs-type partitions with $\alpha \in (0,1)$, which are the partitions induced by the Poisson–Kingman processes with parameter $\tau > 0$. In this case, Pitman (2003, Section 5.4) shows that the structural distribution μ_1 admits a probability density function on $(0,1)$ given by

$$p(v) = \frac{\alpha}{\Gamma(1-\alpha)} v^{-\alpha} \tau^{-\alpha} \frac{f_\alpha(\tau(1-v))}{f_\alpha(\tau)}, \tag{3.18}$$

which was also derived by Favaro and Walker (2013) with an application of Perman et al. (1992, Theorem 2.1). For the remainder, we will refer to any subclass of the Gibbs-type beta process or IBP by the name of the random measures inducing the corresponding Gibbs-type partition. For example, we will say *Pitman–Yor-type beta process* and *Pitman–Yor-type IBP* instead of stable beta process and stable IBP, etc.

4 Stick-breaking representations

So-called *stick-breaking* representations for the beta process are analogous to the stick-breaking constructions for random probability measures such as the Dirichlet and Pitman–Yor processes. (See Sethuraman (1994); Ishwaran and James (2001) for background on stick-breaking representations for random probability measures.) These representations are useful for applications because they lead to practical inference procedures. With an application of Roy (2014, Theorem 1.3), we may obtain an analogous stick-breaking representation of the Gibbs-type beta process as follows: Recall from Equation (3.1) that P_i is the limiting frequency of the i-th block in a Gibbs-type partition, whose distribution is denoted by μ_i. Then $(P_i)_{i\in\mathbb{N}}$ are the size-biased frequencies. Let B_0 be a non-atomic measure on $(\Omega, \mathcal{A})$, and define

$$B := \sum_{i=1}^{\infty} \sum_{j=1}^{C_i} P_{i,j} \delta_{\omega_{i,j}}, \tag{4.1}$$

where $(C_i)_{i\in\mathbb{N}}$, $(\omega_{i,j})_{i,j\in\mathbb{N}}$, and $(P_{i,j})_{j\in\mathbb{N}}$ are independent processes and

1. $(C_i)_{i\in\mathbb{N}}$ are i.i.d. Poisson(γ) random variables with $\gamma := B_0(\Omega)$;
2. $(\omega_{i,j})_{i,j\in\mathbb{N}}$ are i.i.d. random elements in Ω with distribution $\gamma^{-1}B_0$; and
3. For every $i \in \mathbb{N}$, the random variables in the collection $(P_{i,j})_{j\in\mathbb{N}}$ are i.i.d. copies of P_i.

The problem of constructing B then amounts to that of constructing the size-biased frequencies $(P_i)_{i\in\mathbb{N}}$ specific to the underlying Gibbs-type partition. Efficient constructions for these size-biased frequencies are available for many subclasses of the Gibbs-type partitions; in these cases, we obtain efficient stick-breaking constructions for the corresponding subclasses of the Gibbs-type beta process. Here we summarize these results.

For every $i \in \mathbb{N}$, let

$$P_i = W_i \prod_{j=1}^{i-1}(1 - W_j), \tag{4.2}$$

with $P_1 = W_1$, for some random elements $W := (W_j)_{j\in\mathbb{N}}$ in $(0, 1]$. If $W_j \sim \text{beta}(1, \theta)$, i.i.d. for every $j \in \mathbb{N}$ and $\theta > 0$, then Equation (4.2) is the i-th stick of a Dirichlet process (Sethuraman, 1994). In our terminology, Paisley et al. (2010) showed that B is then a Dirichlet-type beta process (with concentration parameter θ and base measure B_0). If the random variables W are merely independent with $W_j \sim \text{beta}(1 - \alpha, \theta + j\alpha)$, for every $j \in \mathbb{N}$ and some $\alpha \in (0, 1)$ and $\theta > -\alpha$, then Equation (4.2) is the i-th stick of a Pitman–Yor process (Perman et al., 1992). In our terminology, Broderick et al. (2012) showed that B is a Pitman–Yor-type beta process (with discount parameter α, concentration parameter θ, and base measure B_0). As with the Pitman–Yor-type IBP, these authors did not consider a stick-breaking construction for the Pitman–Yor-type beta process with $\alpha < 0$ and $\theta = m|\alpha|$ for some m in $\mathbb{N}$. However, the sticks of the Pitman–Yor processes in this case are still independent and distributed as $W_j \sim \text{beta}(1 - \alpha, m|\alpha| + j\alpha)$, for every $j \in \mathbb{N}$ (Pitman, 1995, Proposition 9), and so this extension does indeed arise from the construction in Equation (4.1).

In order to complete the stick-breaking representations for the Gibbs-type beta processes, all that remains is to describe the distribution of W in the case when $\alpha \in (0, 1)$. Favaro and Walker (2013) applied (Perman et al., 1992, Theorem 2.1) to show that the sequence $(W_j)_{j\in\mathbb{N}}$ is composed of dependent random variables that may be characterized sequentially as follows: The first stick $P_1 = W_1$ has distribution μ_1 given by Equation (3.18). For every $j \geq 2$, conditioned on $W_1, \ldots, W_{j-1}$, the random variable W_j admits a conditional density on $(0, 1]$ with density function

$$p(w_j \mid w_1, \ldots, w_{j-1}) = \frac{\alpha}{\Gamma(1-\alpha)}\Big[\tau w_j \prod_{k=1}^{j-1}(1 - w_k)\Big]^{-\alpha} \frac{f_\alpha(\tau \prod_{k=1}^{j}(1 - w_k))}{f_\alpha(\tau \prod_{k=1}^{j-1}(1 - w_k))}, \tag{4.3}$$

where $\tau > 0$ is the parameter of the Poisson–Kingman model (see Equation (2.7)). An algorithm for slice sampling the sequence W was provided therein, and Favaro et al. (2014) showed that, under certain assumptions on the parameter α, these sticks can be directly constructed with beta and gamma random variables.

We present an alternative stick-breaking representation for the Gibbs-type beta process that represents the measures $\sum_{j=1}^{C_i} P_{i,j}\delta_{\omega_{i,j}}$, for every $i \geq 1$, in Equation (4.1) with independent Poisson processes. This representation results from an application of Roy (2014, Theorem 1.4). Let

$$B = \sum_{n=0}^{\infty} \sum_{(\omega,p)\in\eta_n} p\,\delta_{\omega}, \tag{4.4}$$

where $\eta_0, \eta_1, \eta_2, \ldots$ are independent Poisson processes on $\Omega \times (0,1]$ with finite intensity measures

$$(\mathbb{E}\eta_n)(\mathrm{d}\omega \times \mathrm{d}p) = B_0(\mathrm{d}\omega)(1-p)^n \mu_1(\mathrm{d}p), \qquad n \in \{0,1,2,\ldots\}. \tag{4.5}$$

One may verify that B in Equation (4.4) is indeed the Gibbs-type beta process given by Equation (3.3) using a Poisson process superposition argument and the identity $p^{-1} = \sum_{n=0}^{\infty}(1-p)^n$. For $\alpha < 0$, we have

$$(\mathbb{E}\eta_n)(\mathrm{d}\omega \times \mathrm{d}p) = B_0(\mathrm{d}\omega)\frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)\Gamma(\theta+\alpha)} p^{-\alpha}(1-p)^{\theta+\alpha+n-1}\mathrm{d}p, \tag{4.6}$$

where $\theta = m|\alpha|$ for some m in $\mathbb{N}$. This same form characterizes the case $\alpha = 0$ by setting $\theta > 0$. When $\alpha \in (0,1)$ and $\theta > -\alpha$, Equation (4.6) characterizes the rest of the Pitman–Yor-type beta processes. More generally, when $\alpha \in (0,1)$ we have

$$(\mathbb{E}\eta_n)(\mathrm{d}\omega \times \mathrm{d}p) = B_0(\mathrm{d}\omega)\frac{\alpha}{\Gamma(1-\alpha)}(1-p)^n p^{-\alpha}\tau^{-\alpha}\frac{f_\alpha(\tau(1-p))}{f_\alpha(\tau)}\mathrm{d}p, \tag{4.7}$$

where $\tau > 0$.

These stick-breaking representations are useful for applications because inference procedures may be obtained in which the sticks are auxiliary variables. Though only a finite number of the sticks may be represented in practice, these representations yield error bounds when we truncate the outer sums in either Equation (4.1) or Equation (4.4) to a finite number of terms (Doshi-Velez et al., 2009, Section 3.6); (Paisley et al., 2011, Theorem 1); (Roy, 2014, Theorem 1.5). Additionally, a Markov chain Monte Carlo routine including an auxiliary variable may be used to numerically integrate over the number of represented sticks, which removes the approximation error in the asymptotic regime of the Markov chain (Ishwaran and James, 2001).

5 Controlling the statistics of latent features

In statistical applications, it is important to tailor the assumptions that a model encodes about the structure and complexity of the data. In this section, we characterize the asymptotic behavior of the distribution of the latent features in the Gibbs-type IBP. As before, let K_n denote the number of dishes sampled among the first n customers in the Gibbs-type IBP. Additionally, let $K_{n,j}$ denote the number of dishes sampled exactly j times among the first n customers, for every $n \geq j \geq 1$.

5.1 Power-law behavior when $\alpha \in (0,1)$

As we saw in Section 2, when $\alpha \in (0,1)$ the underlying Gibbs-type partitions correspond to the class of partitions induced by the Poisson–Kingman measures with parameter $\tau > 0$, which includes the normalized generalized gamma processes and a subclass of the Pitman–Yor processes. These models have been shown to exhibit *power-law* (i.e., heavy-tailed) behavior in the asymptotic distribution on the number of blocks in the partition (Pitman, 2003). Empirical measurements in a variety of domains have been shown to exhibit power-law behavior. For example, the occurrence of unique words in a document, the degrees of interactions in a protein network, and the number of citations of an academic article all exhibit power law behavior. An appropriate model for data that may depend on these factors should be able to capture this behavior in its latent structure. It was shown by Teh and Görür (2009) and Broderick et al. (2012) that the Pitman–Yor IBP exhibits power-law behavior in the asymptotic distributions of K_n and $K_{n,j}$. We will now see that this behavior is, in a sense, inherited from the partitions induced by the Pitman–Yor processes, and that power-law behavior for any partition induced by a Poisson–Kingman measure translates into power-law behavior in the corresponding Gibbs-type IBP.

Let $\alpha \in (0,1)$, let $\tau > 0$, and let ν_Π be the Lévy intensity of the Gibbs-type beta process defined in Equation (3.3), parameterized by the structural distribution for the Poisson–Kingman measures in Equation (3.18). In this case, it follows analogously to the results by Broderick et al. (2012, p. 459) that ν_Π satisfies the limiting behavior

$$\int_{\Omega\times(0,x]} p\,\nu_\Pi(\mathrm{d}\omega \times \mathrm{d}p) \sim \frac{\alpha}{1-\alpha} C x^{1-\alpha}, \quad \text{as } x \to 0, \tag{5.1}$$

for a constant $C := \tau^{-\alpha}$, where $\sim$ indicates that the ratio of the left and right hand sides tends to one in the specified limit. With derivations analogous to those by Broderick et al. (2012, Proposition 6.1, Lemma 6.2, Lemma 6.3, & Proposition 6.4), it is straightforward to verify that, with probability one,

$$K_n \sim \gamma C n^\alpha \quad \text{and} \quad K_{n,j} \sim \gamma \frac{\alpha \Gamma(j-\alpha)}{j!\,\Gamma(1-\alpha)} C n^\alpha, \quad \text{as } n \to \infty, \tag{5.2}$$

where $\gamma > 0$ is the mass parameter of the Gibbs-type IBP. These statistics therefore exhibit power law behavior controlled by the value of $\alpha \in (0,1)$; the closer α is to one, the heavier the tails of these distributions.

Recall from Section 2 that members of the Gibbs-type partitions (and thus the Gibbs-type IBP) are obtained by mixing over a random parameter τ. Consider the case when τ has a density function on $\mathbb{R}_{>0}$ given by $h(t) f_\alpha(t)$, for a measurable function $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. In this case, the constant C in Equation (5.2) becomes

$$C := \int_0^\infty t^{-\alpha} h(t) f_\alpha(t) \mathrm{d}t. \tag{5.3}$$

By choosing $h(t) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} t^{-\theta}$ for some $\theta > -\alpha$, we have that ν_Π is the Lévy intensity of the Pitman–Yor-type beta process, and $C = \alpha^{-1}\Gamma(\theta+1)/\Gamma(\theta+\alpha)$, which was previously derived by Broderick et al. (2012). By choosing $h(t) = e^{\beta^\alpha - \beta t}$, for some $\beta > 0$, then ν_Π is the Lévy intensity of a normalized generalized gamma-type beta process, and we find that $C = e^{\beta^\alpha} \int_0^\infty t^{-\alpha} e^{-\beta t} f_\alpha(t) \mathrm{d}t$. In this case, if $\alpha = 1/2$, then ν_Π is the Lévy intensity of a normalized inverse Gaussian-type beta process, and C has a closed form solution given by $C = \frac{2}{\pi} \beta^{1/2} e^{\beta^{1/2}} \phi_1(\beta^{1/2})$, where ϕ_ν is the modified Bessel function of the third type.

In order to compare the power-law behaviors of different Gibbs-type partitions, De Blasi et al. (2014) chose hyperparameters for the Pitman–Yor and normalized generalized gamma processes such that the expected number of blocks in the corresponding partitions satisfy $\mathbb{E}[B_{50}] \approx 25$. By plotting statistics such as the expected number of blocks B_n in the partition as n varies, one may visualize differences in the asymptotic behaviors between the models. As one should expect, these same hyperparameter settings also provide an appropriate comparison for their corresponding Gibbs-type IBPs. In particular, recall that in the Gibbs-type IBP the j-th customer samples a Poisson($\gamma Q^{j-1}_{\alpha,\Theta}(1,1)$) number of new dishes. Then the total number of dishes K_n sampled by n customers has a Poisson distribution with mean $\gamma \sum_{j=1}^n Q^{j-1}_{\alpha,\Theta}(1,1)$, where we recall that $Q^0_{\alpha,\Theta}(1,1) := 1$. Setting $(\alpha, \theta) = (0.25, 12.22)$ and $(\alpha, \beta) = (0.74, 1)$ for the Pitman–Yor- and normalized generalized gamma-type IBPs, respectively, we then have that $\mathbb{E}[K_{50}] \approx 25\gamma$ for both models. In Figure 1, we plot the behavior of K_n and $K_{n,1}$ as n increases for these two Gibbs-type IBP subclasses, with the additional choice of $\gamma = 1$.

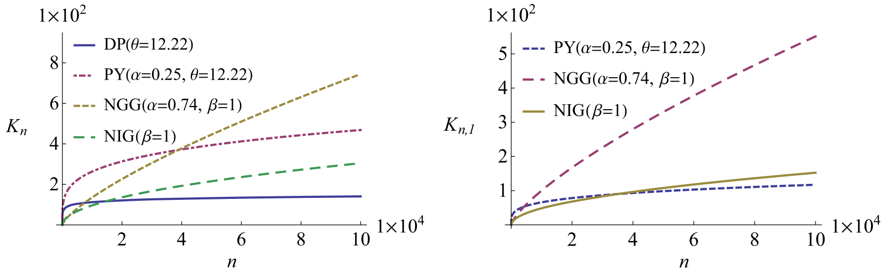
We can see that, for this comparable set of hyperparameters, the normalized generalized gamma-type IBP exhibits heavier tails than the Pitman–Yor-type IBP on both statistics, though in smaller n regimes the reverse holds. The normalized inverse Gaussian-type IBP, at the same setting of $\beta = 1$, exhibits similar tail behavior in $K_{n,1}$ to the Pitman–Yor-type IBP. For comparison, the asymptotic behavior of K_n for the Dirichlet-type IBP at the same hyperparameter setting as the Pitman–Yor-type IBP is also displayed, which does not exhibit power-law behavior (K_n grows proportionally with $\log n$ in this case (Ghahramani et al., 2007)). These characteristics distinguish the subclasses of Poisson–Kingman-type IBPs and provide a variety of power-law modeling options to a practitioner.

5.2 Logarithmic growth when $\alpha = 0$

Recall that the Gibbs-type partitions with $\alpha = 0$ coincide with the random partitions induced by the Dirichlet processes with concentration parameter θ. With probability one, the number of blocks in the partition of $[n]$ satisfies $B_n \sim \theta \log n$ as $n \to \infty$ (Korwar and Hollander, 1973). Similarly, with probability one, the number of features in the corresponding Gibbs-type IBP (i.e., the original IBP) satisfies $K_n \sim \gamma\theta \log n$ as $n \to \infty$ (Ghahramani et al., 2007), where γ is the mass parameter of the IBP.

5.3 Finite feature models when $\alpha < 0$

Finally, recall that the Gibbs-type partitions with $\alpha < 0$ coincide with the random partitions induced by the Pitman–Yor processes with discount parameter $\alpha < 0$ and



(a) Asymptotic behavior of K_n.

(b) Asymptotic behavior of $K_{n,1}$.

Figure 1: The behavior of K_n (the number of features) and $K_{n,1}$ (the number of features with exactly one assignment) for several subclasses of the Gibbs-type IBP, as n increases. Heavy-tailed behavior demonstrates power-law properties.

concentration parameter $\theta = m|\alpha|$ (see Equation (1.5)), where m is a *random* element in $\mathbb{N}$ (Pitman, 2002, Chapter 3, Section 2); (Gnedin and Pitman, 2006, Theorem 12). This subclass may be interpreted as an urn scheme with a finite—but random—number of colors m, and a number of specific examples have been investigated in the literature (Gnedin, 2010; De Blasi et al., 2014). In this case, with probability one, $B_n = m$ for all sufficiently large n. That is, there are a finite number of blocks that are eventually exhausted.

As one may anticipate, the corresponding Gibbs-type IBP in this regime may be analogously interpreted as a feature allocation with a random finite number of features. In particular, when $\alpha < 0$, the Gibbs-type IBP will have a finite number of features if and only if $\mathbb{E}[m] < \infty$. Informally, recall from Section 3.1 that the number of new dishes K^+_{n+1} sampled by the $n+1$-st customer in the Gibbs-type IBP is Poisson distributed with rate $\gamma\mathbb{P}(B_{n+1} > B_n)$ (see Equations (3.9) and (3.11)). Using two Borel–Cantelli arguments, we show in the supplementary material that $\mathbb{E}[m] < \infty$ if and only if the sequence $\mathbb{P}(B_{n+1} > B_n)$ is summable if and only if $K^+_{n+1} = 0$ for all sufficiently large n a.s.

6 Black-box posterior inference

We propose a Markov chain Monte Carlo algorithm generalizing the procedure for posterior inference with the IBP, originally developed by Ghahramani et al. (2007) and Meeds et al. (2007). We will see how these inference methods may be treated as a black-box, where implementing any subclass of the Gibbs-type IBP requires only several evaluations of the primitives $Q^n_{\alpha,\Theta}(z_1, z_2)$, given by Equation (1.2).

Fix $n \geq 1$, and let $(\omega_1, \ldots, \omega_{K_n})$ denote the K_n distinct atoms among the sequence $Z_1, \ldots, Z_n$, where, in this section, we assume that the ordering is chosen uniformly at random, conditioned on K_n. For every $i \leq n$ and $k \leq K_n$, define $Z_{i,k} := Z_i(\{\omega_k\})$, and let $Z := (Z_{i,k})_{i \leq n,\, k \leq K_n}$. Latent feature models have been applied to a variety of statistical problems (as discussed in Section 1.1). In most of these applications, the features (associated with the atoms) represent latent clusters or factors underlying a data set comprised of n observations $Y := (Y_1, \ldots, Y_n)$. Informally, observation Y_i is associated with every latent component ω_k for which $Z_{i,k} = 1$. More carefully, let $\Omega = (\omega_1, \ldots, \omega_{K_n})$ and recall that Ω is an i.i.d. sequence (drawn from the normalized base measure) and independent of Z, conditioned on K_n. Let ψ be a latent variable independent of Ω and Z, and define $\Phi = (\psi, \Omega)$. We then fix a likelihood $p(Y|Z, \Phi) = \prod_{i=1}^n f(Y_i; \psi, Z_i)$ for some density f. In other words, the numbering of the features is irrelevant to the likelihood.

Consider resampling an element of Z from its conditional distribution given Y, Φ, and the remainder of Z. Fix a data point $i \leq n$. For every $k \leq K_n$, let $Z_{-(i,k)}$ be the elements of Z excluding $Z_{i,k}$, let $Z^z_{-(i,k)}$ be the elements of Z with $Z_{i,k}$ replaced by z, and let $S_k^{(-i)} := \sum_{j \neq i, j \leq n} Z_{j,k}$ be the number of datapoints, other than i, that have feature k. For $k \leq K_n$ and $S_k^{(-i)} > 0$, Bayes's rule implies that

$$\mathbb{P}[Z_{i,k} = z \mid Y, Z_{-(i,k)}, \Phi] \propto p(Y \mid Z^z_{-(i,k)}, \Phi) \times \mathbb{P}[Z_{i,k} = z \mid Z_{-(i,k)}], \quad z \in \{0, 1\}, \tag{6.1}$$

where $p(Y \mid Z, \Phi)$ is the likelihood defined above. Recall that we have associated the i-th customer in the buffet analogy with Z_i. By exchangeability, we may treat this as the last customer to enter the buffet, and so

$$\mathbb{P}[Z_{i,k} = 1 \mid Z_{-(i,k)}] \propto (S_k^{(-i)} - \alpha) Q^{n-1}_{\alpha,\Theta}(1, 0). \tag{6.2}$$

Therefore, conditioned on K_n, for every $i \leq n$ and $k \leq K_n$ where $S_k^{(-i)} > 0$, we may resample $Z_{i,k}$ according to Equations (6.1) and (6.2).

We can resample the remaining elements of Z using the Metropolis–Hastings proposal proposed by Meeds et al. (2007). In particular, for every $i \leq n$, we propose removing those features possessed by only Z_i, that is, those atoms ω_k in $\{\omega_1, \ldots, \omega_{K_n}\}$ with $Z_{i,k} = 1$ and $S_k^{(-i)} = 0$. We propose replacing these atoms with K_i^+ new atoms (possessed only by Z_i). Recall that K_i^+ is interpreted as the number of dishes taken by only the i-th customer. Because we may treat the i-th customer as if they were the last to enter the buffet, the distribution of K_i^+ is the same as the distribution of the number of new dishes sampled by the last customer, and so

$$K_i^+ \;\sim\; \text{Poisson}(\gamma Q^{n-1}_{\alpha,\Theta}(1, 1)). \tag{6.3}$$

The Markov proposal replaces those entries in Φ associated with the removed atoms with a set of new parameters associated with the new atoms, sampled from the normalized base measure. (In order to get a simple acceptance probability, the numbering of the features, and thus the column ordering of the array Z, can be resampled uniformly at random. Alternatively, one can ignore the ordering and work implicitly in the space of equivalence classes up to ordering, as the columns are already uniquely identified by their latent parameters, assuming the base measure is non-atomic.) Let Z^* and Φ^* denote the proposed feature assignments and parameters. It is straightforward to show that the Metropolis–Hastings acceptance probability for this proposal is $\min\{1, p(Y|Z^*, \Phi^*)/p(Y|Z, \Phi)\}$ (Meeds et al., 2007). This move potentially changes the number of atoms K_n among $Z_1, \ldots, Z_n$ and thus the number of latent features in the feature allocation. We then proceed to the next process Z_{i+1} and repeat this procedure. Iterating these steps, along with standard Gibbs sampling moves that resample the latent parameters Φ, results in a Markov chain that targets the posterior distribution of Z and Φ, conditioned on the data Y, as its steady state distribution.

Without good prior knowledge of what the parameters γ, α and Θ governing the IBP model should be for a particular application and data set, we may give them broad prior distributions and infer their values during posterior inference. See Section 7 for further details. Note that the inference procedure we have described may be treated as a black-box for any subclass of Gibbs-type IBPs, where the user only needs to supply several evaluations of the primitives $Q^n_{\alpha,\Theta}(\cdot,\cdot)$. In particular, resampling Z only requires the two values $Q^{n-1}_{\alpha,\Theta}(1,1)$ and $Q^{n-1}_{\alpha,\Theta}(1,0)$ (in order to evaluate Equations (6.1) and (6.3)) for a dataset of size n. In order to resample the hyperparameters γ, α and Θ for the IBP model, one needs to supply $n-1$ additional evaluations to obtain $Q^{n-s}_{\alpha,\Theta}(s,1)$, for $n \geq s \geq 1$, required by Equation (3.17). These $n+1$ values may be precomputed and stored for given values of α and Θ. See the supplementary material for some notes on computing these primitives, the required generalized factorial coefficients $\mathscr{C}(n,k;\alpha)$ in Equation (1.3), and the Gibbs-type weights $\overline{V}$ for different models.

7 Experiments

We now demonstrate the differences between several subclasses of the Gibbs-type IBP. We do not implement models with $\alpha < 0$ here due to computational difficulties (as discussed in the supplementary material). This section will therefore focus on subclasses of the Gibbs-type IBP with $\alpha \in [0,1)$. See Section 8 for a further discussion.

For every $i \leq n$, assume that data point Y_i is composed of p measurements $Y_i := (Y_{i,1}, \ldots, Y_{i,p})$. Consider the following factor analysis model for Y:

$$Y_{i,j} = \sum_{k=1}^{K_n} W_{i,k} Z_{i,k} A_{k,j} + \varepsilon_{i,j}, \qquad i \leq n,\, j \leq p, \tag{7.1}$$

where $W := (W_{i,k})_{k \leq K_n, i \leq n}$ are $\mathbb{R}$-valued modulating weights, $A := (A_{k,j})_{k \leq K_n, j \leq p}$ are $\mathbb{R}$-valued factor loadings, and $\varepsilon := (\varepsilon_{i,j})_{j \leq p, i \leq n}$ are $\mathbb{R}$-valued additive noise terms. Let

$$\begin{aligned}
W_{i,k} \mid \sigma_W \ &\sim\ \mathcal{N}(0, \sigma_W^2), && i \leq n,\, k \leq K_n, && (7.2)\\
A_{k,j} \mid \sigma_{A,j} \ &\sim\ \mathcal{N}(0, \sigma_{A,d}^2), && j \leq p,\, k \leq K_n, && (7.3)\\
\varepsilon_{i,j} \mid \sigma_Y \ &\sim\ \mathcal{N}(0, \sigma_Y^2), && i \leq n,\, j \leq p, && (7.4)
\end{aligned}$$

where $\sigma_Y, \sigma_W, \sigma_{A,1}, \ldots, \sigma_{A,p}$ are positive-valued hyperparameters. Viewing Y, Z, W, A, and ε as matrices in the obvious way, we may write $Y = (W \circ Z)A + \varepsilon$ where $\circ$ represents element-wise multiplication. Then the data Y is conditionally matrix Gaussian and admits the conditional density

$$p(Y \mid Z, W, A, \sigma_X) = \frac{1}{(2\pi)^{np/2} \sigma_X^{np}} \exp\Big\{ -\frac{1}{2\sigma_X^2} \operatorname{tr}\Big[(Y-M)^T (Y-M) \Big] \Big\}, \tag{7.5}$$

where $M = (W \circ Z)A$. Note that, in practice, W or A may be analytically marginalized out of this likelihood expression, in which case Y is still conditionally Gaussian.

In the experiments below, we give all hyperparameters broad prior distributions and resample their values during inference with *slice sampling* (Neal, 2003). Where relevant, the discount parameter α is given a beta$(1,1)$ prior distribution. All other parameters in Θ (i.e., the Gibbs-type hyperparameters, which are all positive-valued) are given independent gamma prior distributions, whose hyperparameters are themselves given independent exponential(1) prior distributions. For the noise parameter σ_Y, we let $\sigma_Y^{-2} \mid a_Y, b_Y \sim \text{gamma}(a_Y, b_Y)$, where $a_Y, b_Y \sim \text{exponential}(1)$ are independent. We give σ_W a similar prior specification. Independently of σ_Y and σ_W, we couple the factor variance parameters $(\sigma_{A,j})_{j \le p}$ with a similar model: let $\sigma_{A,j}^{-2} \mid a_A, b_A \sim \text{gamma}(a_A, b_A)$, for all $j \le p$, where $a_A, b_A \sim \text{exponential}(1)$ are independent. Finally, for the IBP mass parameter, let $\gamma \mid a_\gamma, b_\gamma \sim \text{gamma}(a_\gamma, b_\gamma)$, and let $a_\gamma, b_\gamma \sim \text{exponential}(1)$ be independent. In this case, Equation (3.17) implies the conditional distribution of γ remains in the family of gamma distributions, with conditional density

$$p(\gamma \mid Z, \alpha, \Theta, a_\gamma, b_\gamma) \propto \gamma^{K_n} \exp\Big(-\gamma \sum_{j=1}^{n} Q_{\alpha,\Theta}^{j-1}(1,1)\Big) \times \text{gamma}(\gamma; a_\gamma, b_\gamma) \tag{7.6}$$

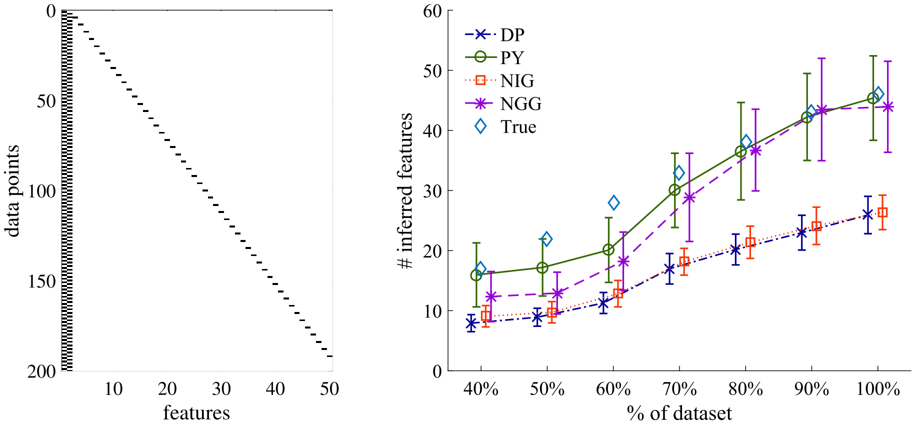
$$= \text{gamma}\Big(\gamma; a_\gamma + K_n, b_\gamma + \sum_{j=1}^{n} Q_{\alpha,\Theta}^{j-1}(1,1)\Big). \tag{7.7}$$

7.1 Synthetic data

First consider a synthetic latent feature allocation, displayed as a 200×50 binary matrix in Figure 2(a). The rows correspond to the $n = 200$ data points and the columns correspond to the $K_n = 50$ latent features, that is, the i-th row and k-th column is shaded black if $Z_{i,k} = 1$ (in the notation of Section 6). In this example, every data point possesses one of the first two features, and the remaining 48 features are each only possessed by one data point. We simulate a dataset Y of $n = 200$ measurements in $p = 50$ variables from the model in Equations (7.1) to (7.4) with $\sigma_X = \sigma_W = 1$, and $\sigma_{A,j} = 1$ for $j \le p$.

We implemented the posterior inference procedure described in Section 6 for 6,000 burn-in iterations. In Figure 2(b) we display the number of features inferred by the Dirichlet, Pitman–Yor, normalized inverse Gaussian, and normalized inverse gamma—denoted DP, PY, NIG, and NGG, respectively—subclasses of the Gibbs-type IBP on different subsets of the data. In particular, we ran the inference procedure on 40% of the data points, then on 50%, and so on, indicated by the horizontal axis from left to right. The mean number of inferred features (along with ± one standard deviation) over 3,000 samples following the burn-in period are displayed for each model. The true number of features in each subset of the data are also displayed for reference.

We note that all models attained approximately the same training loglikelihood given each data subset (averaged over the samples). However, the more flexible PY and NGG-IBP variants were able to more accurately infer the number of features underlying



(a) Latent feature matrix.

(b) Inferred number of features.

Figure 2: (a) A synthetic latent feature matrix for $n = 200$ data points with $K_{200} = 50$ features. The simulated data was in $p = 50$ variables. (b) The number of features inferred by different subclasses of the Gibbs-type IBP as we sequentially include more of the data. For each subset of the data, we plot the mean number of features over 3,000 samples following a burn-in period. Bars at ± one standard deviation are also displayed. The true number of features in each subset of the data is plotted for reference.

the data compared to the less expressive subclasses, the DP- and NIG-IBPs. We recall that the DP-IBP is an extreme point of both the PY- and NGG-IBP subclasses. The discount parameter α differentiates these models, and as we saw in Section 5, inferring this parameter allows these models to detect the power law structure present in the latent feature allocation displayed in Figure 2(a). In the supplementary material, we provide trace plots of the Gibbs-type hyperparameters over the burn-in period, along with histograms over samples repeatedly drawn following the burn-in.

7.2 MNIST digits

We also applied the model in Section 6 to $n = 1000$ examples of the digit ‘3’ from the MNIST handwritten digits dataset. We projected the data onto its first $p = 64$ principal components in order to replicate the experiment performed by Teh et al. (2007) with the DP-IBP (and a more restrictive setting of the hyperparameters). Here we present the same qualitative analyses for different subclasses of the Gibbs-type IBP. The reader can see Paisley et al. (2010); Broderick et al. (2012) for similar experiments. We ran our posterior inference procedure for 20,000 iterations, which was sufficient for every model to burn-in. We collected 1,000 samples (thinned from 10,000 samples) of all latent variables in the model following the burn-in period, and we display boxplots of the number of inferred features over the collected samples in Figure 3. (In the supplementary material, we provide visualizations of the inferred values of the Gibbs-type hyperparameters.) In Figure 4, we find the maximum a posteriori (MAP) sample (of all latent variables and

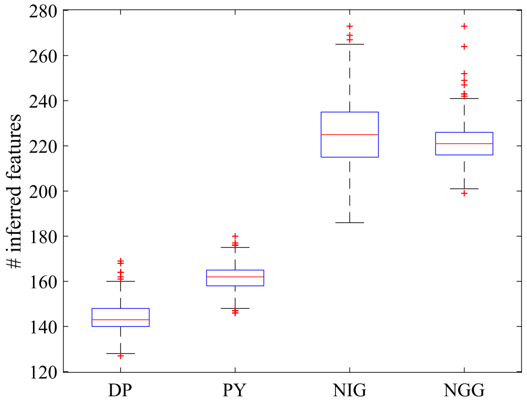
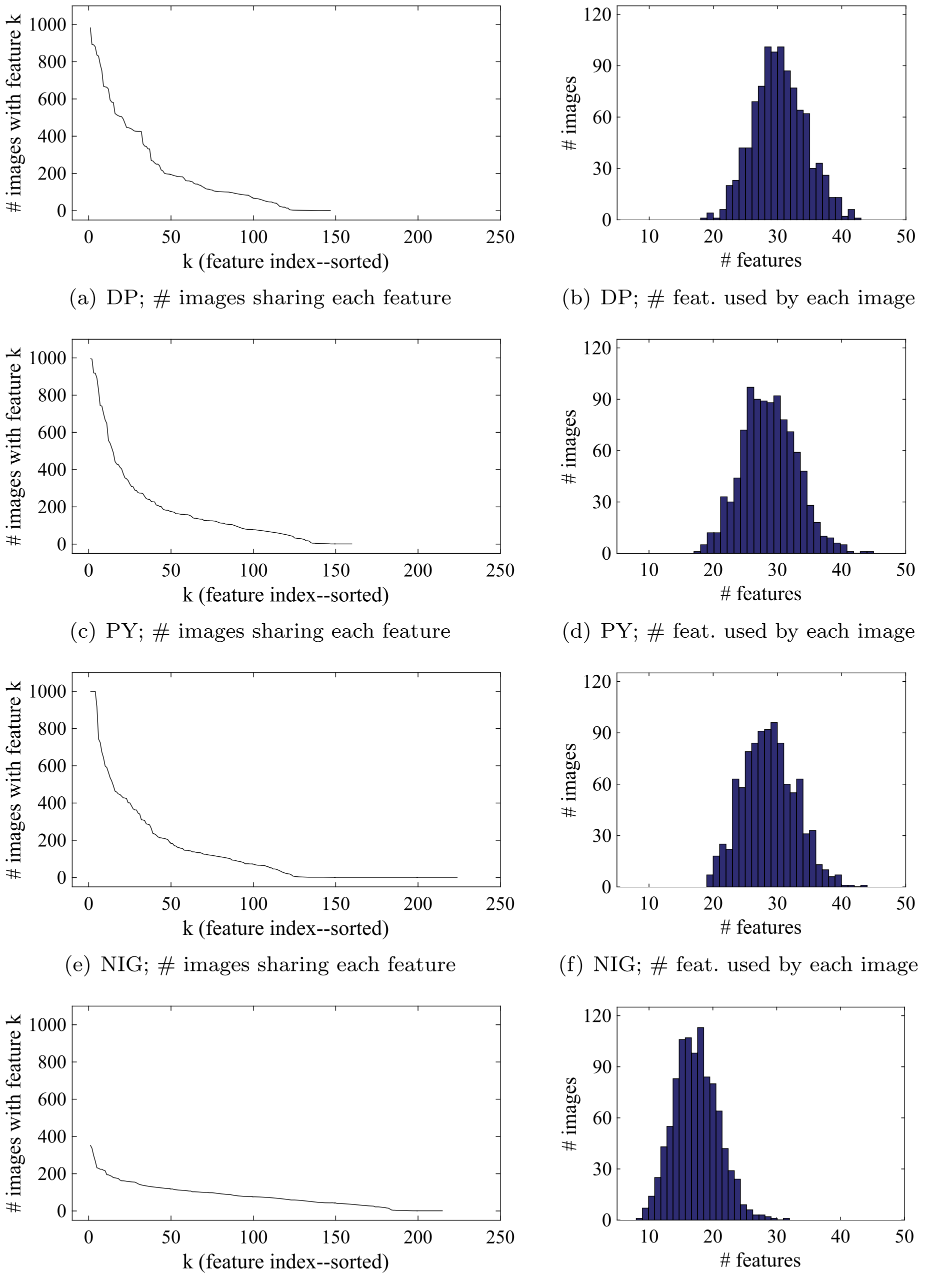


Figure 3: Number of features inferred by the different subclasses on the MNIST dataset. Boxplots over 1,000 samples (thinned from 10,000 samples) collected following a burn-in period of 20,000 iterations.

parameters, from among the collected samples) for each model, and for that sample we plot (1) the number of images sharing each feature and (2) a histogram of the number of features used by each image. For visualization, the features in the former plots are ordered according to the number of images assigned to them. The scale of the axes in the subfigures are held fixed for comparison.

We find that the heavy-tailed models, i.e., the PY-, NIG-, and NGG-IBPs, exhibit different extents of power-law behaviors achieved by tailoring the total number of inferred features and the number of features with relatively few assignments. In particular, Figure 3 shows that the PY-IBP infers more features than the DP-IBP (based on an unpaired t-test at a 0.05 significance level). Moreover, both the NIG- and NGG-IBP models infer significantly higher numbers of features than the PY-IBP, but do not themselves differ significantly. Figure 4 shows that these differences are due to varying power-law behaviors between the models. In particular, the PY-, NIG-, and NGG-IBP models display increasingly heavier tail behavior in the (distribution of the) number of images sharing each feature. The NGG-IBP model is notable as clearly having dramatically heavier tails than all other models in this distribution. This additionally results in a noticeably lower average number of features per image (visible in the histogram), which does not appear to differ significantly between the other three subclasses.

This experiment demonstrates important variations between the Gibbs-type IBP subclasses. Compare the latent feature distributions between the three heavy-tailed variants. On one hand, the NIG-IBP has heavier tails than the PY-IBP, accomplished by creating many features to which very few images are assigned, resulting in a significantly larger number of features. On the other hand, the NGG-IBP has much heavier tails than the NIG-IBP, accomplished by heavily skewing the distribution towards the (right) tail, yet maintaining approximately the same total number of features. It is particularly interesting to compare the PY- and NGG-IBP models in this respect, as the DP-IBP may be approximated by both of these subclasses. As discussed in Section 5,



(a) DP; # images sharing each feature

(b) DP; # feat. used by each image

(c) PY; # images sharing each feature

(d) PY; # feat. used by each image

(e) NIG; # images sharing each feature

(f) NIG; # feat. used by each image

(g) NGG; # images sharing each feature

(h) NGG; # feat. used by each image

Figure 4: Latent feature statistics inferred by each model on the MNIST dataset. For each model, the number of images assigned to each feature is displayed as a plot (sorted for visualization), and the number of features used by an image is displayed as a histogram.

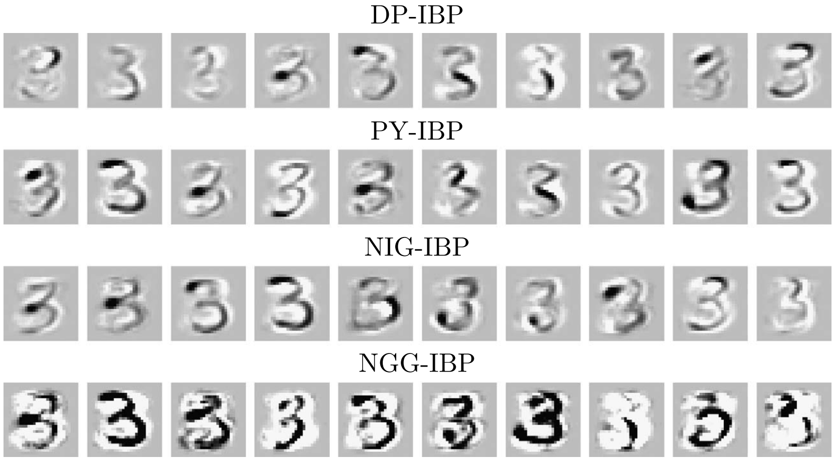


Figure 5: Top 10 (according to the weight matrix W) important features (represented by the factors in A) for the digit ‘3’ inferred by each subclass of the Gibbs-type IBP. Darker pixel values correspond to larger values in (the corresponding factor in) A.

these differing properties provide several different options to a practitioner, which are accessible through our black-box constructions and posterior inference procedures.

Finally, we can visualize the effect that the different latent feature distributions have on this particular application by investigating some of the latent features inferred by each model. In Figure 5, we display the top 10 (according to the weight matrix W) most important features (represented by the factors in A) from the MAP sample collected for each model. The features inferred by the DP-, PY-, and NIG-IBP models do not appear to differ, however, the NGG-IBP clearly places the heaviest weight on its features (resulting in darker pixel values). Moreover, a few of these features appear to capture distinct parts of the digits.

8 Conclusion

The Gibbs-type IBPs are a broad class of feature allocation models, parameterized by the law of a Gibbs-type random partition. We showed how the Gibbs-type IBP can be constructed from exchangeable sequences of completely random measures and gave several stick-breaking representations. We also characterized the asymptotic behavior of the number of latent features in a Gibbs-type IBP, which was seen to mimic the asymptotic behavior of the underlying random partition. We described black-box routines for simulation and performing posterior inference with Gibbs-type IBPs that only require a set of precomputed constants that are specific to the corresponding partition law. Our numerical experiments demonstrated differences between the Gibbs-type IBP subclasses, where we saw that different extents of heavy tailed latent feature behavior could be attained beyond the PY-IBP.

Many models that use the beta process as a basic building block can be generalized by instead using the Gibbs-type beta process, which could benefit many applications of the IBP. Further applications of the beta process beyond the IBP should also be considered. For example, Roy (2014) provides a finitary construction for exchangeable sequences of Bernoulli processes (as in Equation (3.4)) rendered conditionally i.i.d. by a *hierarchical beta process* (Thibaux and Jordan, 2007). Such processes are used as admixture models, in which a collection of feature allocations share features, analogously to (collections of) random partitions induced by a hierarchy of partitioning schemes. Feature allocations induced by hierarchies of Gibbs-type beta processes would be a natural generalization of this framework, providing flexible properties (such as power law behavior) to the admixture model.

Finally, we cannot practically apply the simulation or inference procedures described in this article to Gibbs-type IBPs for $\alpha < 0$, because we cannot robustly compute the required primitives $Q^n_{\alpha,\Theta}(\cdot,\cdot)$ in this case (as described in the supplementary material). Constructions by Roy (2014, Definition 6.1) provide alternative simulation procedures, however, posterior inference algorithms have yet to be developed. The stick-breaking representations in Section 4 do not depend on these primitives, and so they may suggest an approach for inference.

Supplementary Material

Supplementary Material: Gibbs-type Indian buffet processes
(DOI: 10.1214/19-BA1166SUPP; .pdf).

References

Broderick, T., Jordan, M. I., and Pitman, J. (2012). "Beta Processes, Stick-Breaking and Power Laws." *Bayesian Analysis*, 7(2): 439–476. MR2934958. doi: https://doi.org/10.1214/12-BA715. 685, 686, 690, 695, 697, 703

Broderick, T., Pitman, J., and Jordan, M. I. (2013). "Feature allocations, probability functions, and paintboxes." *Bayesian Analysis*, 8(4): 801–836. MR3150470. doi: https://doi.org/10.1214/13-BA823. 683, 684

Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*. John Wiley & Sons. MR2131068. doi: https://doi.org/10.1002/0471733180. 685

De Blasi, P., Favaro, S., Lijoi, A., Mena, R., Prünster, I., and Ruggiero, M. (2014). "Are Gibbs-type priors the most natural generalization of the Dirichlet process?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Special issue on Bayesian nonparametrics. 698, 699

Doshi-Velez, F., Miller, K. T., Gael, J. V., and Teh, Y. W. (2009). "Variational inference for the Indian buffet process." In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*. 696

Favaro, S., Lomeli, M., Nipoti, B., and Teh, Y. W. (2014). "On the stick-breaking representation of σ-stable Poisson–Kingman models." *Electronic Journal of Statistics*, 8(1): 1063–1085. MR3263112. doi: https://doi.org/10.1214/14-EJS921. 695

Favaro, S. and Walker, S. G. (2013). "Slice sampling σ-stable Poisson–Kingman mixture models." *Journal of Computational and Graphical Statistics*, 22(4): 830–847. MR3173745. doi: https://doi.org/10.1080/10618600.2012.681211. 694, 695

Ghahramani, Z., Griffiths, T. L., and Sollich, P. (2007). "Bayesian nonparametric latent feature models." *Bayesian Statistics*, 8: 201–226. See also the discussion and rejoinder. MR2433194. 683, 684, 698, 699

Gnedin, A. (2010). "A species sampling model with finitely many types." *Electronic Communications in Probability*, 15: 79–88. MR2606505. doi: https://doi.org/10.1214/ECP.v15-1532. 699

Gnedin, A. and Pitman, J. (2006). "Exchangeable Gibbs partitions and Stirling triangles." *Journal of Mathematical Sciences*, 138(3): 5674–5685. MR2160320. doi: https://doi.org/10.1007/s10958-006-0335-z. 683, 685, 687, 688, 689, 699

Griffiths, T. L. and Ghahramani, Z. (2006). "Infinite latent feature models and the Indian buffet process." In *Advances in Neural Information Processing Systems 19*. 683, 684, 686
Creighton Heaukulani and Daniel M. Roy

Heaukulani, C. and Roy, D. M. (2019). "Supplementary Material: Gibbs-type Indian buffet processes." *Bayesian Analysis*. doi: https://doi.org/10.1214/19-BA1166SUPP. 687

Hjort, N. L. (1990). "Nonparametric Bayes estimators based on beta processes in models for life history data." *The Annals of Statistics*, 18(3): 1259–1294. MR1062708. doi: https://doi.org/10.1214/aos/1176347749. 686, 690

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453). MR1952729. doi: https://doi.org/10.1198/016214501750332758. 694, 696

James, L. F. (2017). "Bayesian Poisson calculus for latent feature modeling via generalized Indian Buffet Process priors." *The Annals of Statistics*, 45(5). MR3718160. doi: https://doi.org/10.1214/16-AOS1517. 691

Kallenberg, O. (2002). *Foundations of Modern Probability*. New York: Springer, 2nd edition. MR1876169. doi: https://doi.org/10.1007/978-1-4757-4015-8. 691

Kim, Y. (1999). "Nonparametric Bayesian estimators for counting processes." *The Annals of Statistics*, 27(2): 562–588. MR1714717. doi: https://doi.org/10.1214/aos/1018031207. 691

Kingman, J. F. C. (1975). "Random discrete distributions." *Journal of the Royal Statistical Society, Series B*, 37(1): 1–22. MR0368264. 688, 689

Kingman, J. F. C. (1978). "The representation of partition structures." *Journal of the London Mathematical Society*, 2(2): 374–380. 688, 690

Korwar, R. M. and Hollander, M. (1973). "Contributions to the theory of Dirichlet processes." *The Annals of Probability*, 1(4): 705–711. MR0350950. doi: `https://doi.org/10.1214/aop/1176996898`. 698

Lijoi, A., Mena, R. H., and Prünster, I. (2005). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: `https://doi.org/10.1198/016214505000000132`. 689

Meeds, E., Ghahramani, Z., Neal, R. M., and Roweis, S. T. (2007). "Modeling dyadic data with binary latent factors." In *Advances in Neural Information Processing Systems 20*. 699, 700

Neal, R. M. (2003). "Slice Sampling." *The Annals of Statistics*, 31(3): 705–741. MR1994729. doi: `https://doi.org/10.1214/aos/1056562461`. 702

Paisley, J., Carin, L., and Blei, D. M. (2011). "Variational inference for stick-breaking beta process priors." In *Proceedings of the 28th International Conference on Machine Learning*. 686, 696

Paisley, J., Zaas, A., Woods, C. W., Ginsburg, G. S., and Carin, L. (2010). "A stick-breaking construction of the beta process." In *Proceedings of the 27th International Conference on Machine Learning*. 686, 695, 703

Perman, M., Pitman, J., and Yor, M. (1992). "Size-biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: `https://doi.org/10.1007/BF01205234`. 686, 694, 695

Pitman, J. (1995). "Exchangeable and partially exchangeable random partitions." *Probability theory and related fields*, 102(2): 145–158. MR1337249. doi: `https://doi.org/10.1007/BF01213386`. 690, 692, 694, 695

Pitman, J. (2002). *Combinatorial stochastic processes*. Springer. Presented as a lecture course at the 32nd Summer School on Probability Theory held in Saint-Flour, July 2002. Available online. MR2245368. 687, 689, 690, 692, 699

Pitman, J. (2003). "Poisson–Kingman partitions." In *Statistics and science: a Festschrift for Terry Speed*, 1–34. Institute of Mathematical Statistics. 688, 689, 694, 697

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 25(2): 855–900. MR1434129. doi: `https://doi.org/10.1214/aop/1024404422`. 686

Roy, D. M. (2014). "The continuum-of-urns scheme, generalized beta and Indian buffet processes, and hierarchies thereof." *arXiv preprint* 1501.00208 [math.PR] (version 1). 683, 686, 690, 691, 694, 696, 707

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4: 639–650. MR1309433. 694, 695

Teh, Y. W. and Görür, D. (2009). "Indian Buffet Processes with Power-law Behavior." In *Advances in Neural Information Processing Systems 22*. 685, 686, 690, 691, 694, 697

Teh, Y. W., Görür, D., and Ghahramani, Z. (2007). "Stick-breaking Construction for the Indian Buffet Process." In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*. 686, 703

Thibaux, R. and Jordan, M. I. (2007). "Hierarchical Beta Processes and the Indian Buffet Process." In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*. 690, 691, 707

Acknowledgments

We thank Stefano Favaro for help with (and code for) implementing various Gibbs-type models and for communicating most of the material in Section S.2 of the supplementary material to us. We thank Lancelot F. James and Igor Prünster for helpful advice and pointers to references. Finally, we thank Annalisa Cerquetti, Hong Ge, Konstantina Palla, Christian Steinruecken, and anonymous reviewers for feedback on drafts. CH was supported in part by the Stephen Thomas studentship at Queens' College, Cambridge, with funding also from the Cambridge Trusts. DMH was supported in part while a research fellow of Emmanuel College, Cambridge, with funding also from a Newton International Fellowship and AFOSR grant #FA9550-15-1-0074.