IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A characterization of product-form exchangeable feature probability functions

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1698801 since 2019-04-17T16:10:11Z

Published version:

DOI:10.1214/17-AAP1333

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

arXiv:1607.02066v1 [math.PR] 7 Jul 2016

A characterization of product-form exchangeable feature probability functions

Marco Battiston
Bocconi University, Italy
email: marco.battiston@phd.unibocconi.it

Stefano Favaro
University of Torino and Collegio Carlo Alberto, Italy
email: stefano.favaro@unito.it

Daniel M. Roy
University of Toronto, Canada
email: droy@utstat.toronto.edu

Yee Whye Teh
University of Oxford, UK
email: y.w.teh@stats.ox.ac.uk

Abstract

We characterize the class of exchangeable feature allocations assigning probability $V_{n,k} \prod_{l=1}^{k} W_{m_l} U_{n-m_l}$ to a feature allocation of n individuals, displaying k features with counts $(m_1, \ldots, m_k)$ for these features. Each element of this class is parametrized by a countable matrix V and two sequences U and W of non-negative weights. Moreover, a consistency condition is imposed to guarantee that the distribution for feature allocations of $n-1$ individuals is recovered from that of n individuals, when the last individual is integrated out. In Theorem 1.1, we prove that the only members of this class satisfying the consistency condition are mixtures of the Indian Buffet Process over its mass parameter γ and mixtures of the Beta–Bernoulli model over its dimensionality parameter N. Hence, we provide a characterization of these two models as the only, up to randomization of the parameters, consistent exchangeable feature allocations having the required product form.

Keywords: Exchangeable Feature Allocations, Gibbs-type partitions, Indian Buffet process.

1 Introduction

Feature allocations are popular models in machine learning. In these models, we consider a set of n individuals, each displaying a (possibly empty) set of features. Specifically, let $(\mathcal{X}, \mathcal{B})$ be measurable space, representing the collection of all possible features. Each individual is described by a random finite subset X_i of $\mathcal{X}$, collecting his features. Each feature $x \in \mathcal{X}$ can be shared by many individuals. Given a set of n individuals, a *feature allocation* describes the sharing of features among these individuals. A way of describing this sharing is to associate to each of the k points in $\cup_{1 \leq i \leq n} X_i$ a subset of $[n] := \{1, \ldots, n\}$, summarizing the individuals having that particular feature. We denote by $(f_{n,1}, \ldots, f_{n,k})$ the subsets of $[n]$ representing each of the k features and by $(m_1, \ldots, m_k)$ the cardinalities of these sets.

A feature allocation is *exchangeable* when its distribution is invariant under permutation of the indexes of the individual, i.e. the feature allocation induced by the random sets $(X_1, \ldots, X_n)$ is equal in distribution to that induced by $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$, for all permutation σ of $[n]$. Moreover, as pointed out in Broderick et al. (2013), it is usually convenient to assign an order to the k features present in a feature allocation of n individuals. A way of achieving this purpose is drawing k values from a continuous distribution and ordering the k features accordingly. The resulting feature allocation is said to be a *randomly ordered feature allocation*. In Broderick et al. (2013), the authors study the class of exchangeable randomly ordered feature allocations admitting as a sufficient statistics the vector $(m_1, \ldots, m_k)$, i.e., the class of randomly ordered exchangeable feature allocations of the form

$$P(f_{n,1}, \ldots, f_{n,k}) = \pi_n(m_1, \ldots, m_k), \tag{1}$$

for a symmetric function π_n, called an *exchangeable feature probability punction* (EFPF), Broderick et al. (2013).

When dealing with random exchangeable feature allocations, we also require consistency conditions that guarantee that the distribution of a feature allocation of n individuals coincides with that of $n-1$ individuals, when the last individual is integrated out. When considering randomly ordered exchangeable feature allocations with EFPF, this consistency notion specializes to the condition

$$\pi_n\left(m_1, \ldots, m_k\right) = \sum_{j=0}^{\infty} \binom{k+j}{j} \sum_{\underline{z} \in \{0,1\}^k} \pi_{n+1}(m_1 + z_1, \ldots, m_k + z_k, \underbrace{1, \ldots, 1}_{j}). \tag{2}$$

Feature allocations satisfying this condition are said to be *consisent*.

The most remarkable example of exchangeable consistent feature allocation with EFPF is the *Indian Buffet Process* (IBP), initially introduced in Griffiths and Ghahramani (2006), in its one parameter version, and then extended to its two, Ghahramani et al. (2007), and three parameters versions, Teh and Görür (2010). The EFPF of a 3-parameter (γ, θ, α) IBP has the following form

$$\frac{1}{k!}\Big(\frac{\gamma}{(\theta+1)_{n-1\uparrow}}\Big)^k \exp\Big(-\sum_{i=1}^{n} \gamma \frac{(\alpha+\theta)_{i-1\uparrow}}{(1+\theta)_{i-1\uparrow}}\Big) \prod_{l=1}^{k} (1-\alpha)_{m_l-1\uparrow} (\theta+\alpha)_{n-m_l\uparrow},$$

where $(x)_{m\uparrow}$ denotes the rising factorial and the parameters must satisfy the conditions $\gamma > 0$,

$0 \leq \alpha < 1$, and $\theta > -\alpha$. The 2-parameter IBP is recovered when α is set equal to zero, and the 1 parameter IBP when we also impose $\theta = 1$. For a review of the IBP and its applications in machine learning, the reader is refered to Griffiths and Ghahramani (2011).

The IBP is derived as the limit of a Beta–Bernoulli model, in Griffiths and Ghahramani (2006). This latter model is the counterpart of the IBP when the set of all possible features $\mathcal{X}$ has finite cardinality, N. The EFPF of a Beta–Bernoulli model with parameters (N, α, θ) is

$$\binom{N}{k}\left(\frac{-\alpha}{(\theta+\alpha)_{n\uparrow}}\right)^k\left(\frac{(\theta+\alpha)_{n\uparrow}}{(\theta)_{n\uparrow}}\right)^N\prod_{i=1}^k(1-\alpha)_{m_i-1\uparrow}(\theta+\alpha)_{n-m_i\uparrow}, \tag{3}$$

where $\alpha < 0$ and $\theta > -\alpha$. In Appendix A.1, we provide a brief description of the Beta–Bernoulli model and a derivation of its EFPF.

Feature allocations are generalizations of partitions. Indeed, a random partition is the particular case of a random feature allocation in which each random set X_i is a singleton with probability one. All notions just introduced (consistent, exchangeable, ordered feature allocation and EFPF) were first introduced for partitions and only recently extended to the feature allocation case. The reader is referred to Pitman (2006) for a complete review of exchangeable random partitions. One the most important distribution for random partitions is the Ewens–Pitman formula, which is generalisation of the famous Ewens formula. Starting from this distribution, Gnedin and Pitman (2006) considers a larger class of random partitions, having an exchangeable partition probability function (see Pitman (1995) for a definition) with the same product form as the Ewens–Pitman formula, but allowing a more general parametrization, depending on a triangular array and on a sequence of non-negative weights. Theorem 12 of Gnedin and Pitman (2006) characterizes all elements of this class of distributions for random partitions satisfying a consistency condition similar to (2). The resulting class of distributions is termed *Gibbs-type partitions*.

Motivated by the work Gnedin and Pitman (2006) in the partition context and by the product form of the EFPF of the IBP and of the Beta–Bernoulli, we consider the class of distributions for consistent exchangeable feature allocations with EFPF of the form

$$\pi_n(m_1, \ldots, m_k) = V_{n,k}\prod_{l=1}^k W_{m_l}U_{n-m_l}, \tag{4}$$

for an infinite array $V = (V_{n,k} : (n,k) \in \mathbb{N} \times \mathbb{N}_0)$ and two sequences $W = (W_j : j \in \mathbb{N})$ and $U = (U_j : j \in \mathbb{N}_0)$ of non-negative weights, where $\mathbb{N}$ denotes the set of positive natural numbers and $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$.

In the feature context, we show that the IBP and the Beta–Bernoulli are the only consistent exchangeable feature allocations with form (4), up to randomization of their γ and N parameters respectively. Consistency and exchangeability also imply that the two sequences of weights, W and U, must have the same form as in the IBP and in the Beta–Bernoulli, for two constants α and θ, satisfying $\alpha \leq 1$ and $\theta > -\alpha$. In addition, V must satisfy a recursion with coefficients depending on α and θ and the set of solutions of this recursion forms convex set. For each fixed α and θ, we describe the extreme points of this convex set. Their form only depends on the value of α. For $0 < \alpha < 1$, the

set of extreme points coincide with the family of V of a 3-parameter IBP. For $\alpha = 0$, this set is still continuous and coincides with the family of V of the 2-parameter IBP. For $\alpha < 0$ the set of extreme points is countably infinite and each extreme point corresponds to the V of a Beta–Bernoulli model. To sum up, we will prove the following theorem.

THEOREM 1.1 A consistent exchangeable feature allocation has EPFP of the form (4) iff, for some $\alpha < 1$ and $\theta > -\alpha$, $W_m = (1-\alpha)_{m-1\uparrow}$ and $U_m = (\theta+\alpha)_{m\uparrow}$ and the elements of V satisfies the recursion

$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \left((\theta+\alpha)_{n\uparrow}\right)^j (\theta+n)^k V_{n+1,k+j}. \tag{5}$$

Moreover, for fixed (α, θ), the set of solutions of (5) are

1. for $0 < \alpha < 1$, mixtures over γ of the V of a 3-parameter IBP;
2. for $\alpha = 0$, mixtures over γ of the V of a 2-parameter IBP;
3. for $\alpha < 0$, mixtures over N of the V of a Beta–Bernoulli model with N features.

In the next section, we prove Theorem 1.1. Specifically, in subsection 2.1, we prove the first part, characterizing U and W and finding the recursion for V. In subsection 2.2, we describe how to derive the extreme solutions of this recursion. Finally, in subsection 2.3, we study the three cases $0 < \alpha < 1$, $\alpha = 0$, and $\alpha < 0$.

2 Proof of Theorem 1.1

The problem is to describe all distributions for exchangeable feature allocations with EFPF (4) subject to the consistency constraint (2), which becomes

$$V_{n,k} \prod_{i=1}^{k} W_{m_l} U_{n-m_l} = \sum_{j=0}^{\infty} \binom{k+j}{j} U_n^j W_1^j \sum_{\underline{z} \in \{0,1\}^k} V_{n+1,k+j} \prod_{i=1}^{k} W_{m_i+z_i} U_{n+1-m_i-z_i}, \tag{6}$$

for all $n \in \mathbb{N}$, $k \in \mathbb{N}_0$, and for all $m_i \leq n$ with $i \leq k$. We start by noting that the representation (4) is not unique. Specifically, we can scale the weights in the following ways, for $\kappa > 0$, and obtain the same distribution:

1. $\tilde{V}_{n,k} = \kappa^{-k} V_{n,k}$ and $\tilde{W}_j = \kappa W_j$;
2. $\tilde{V}_{n,k} = \kappa^{-k} V_{n,k}$ and $\tilde{U}_j = \kappa U_j$;
3. $\tilde{V}_{n,k} = \kappa^{-nk} V_{n,k}$, $\tilde{W}_j = \kappa^j W_j$ and $\tilde{U}_j = \kappa^j U_j$;
4. $\tilde{V}_{n,k} = \kappa^{-k(n-1)} V_{n,k}$, $\tilde{W}_j = \kappa^{j-1} W_j$ and $\tilde{U}_j = \kappa^j U_j$.

By imposing $W_1 = 1$, we avoid the first ambiguity and with $U_0 = 1$ we fix the second one. These conditions also exclude the third ambiguity, but do not exclude the last one, which will be fixed following Proposition 2.1.

2.1 Characterization of W and U

The following Proposition shows that W and U must have a form akin to that of the IBP and V is constrained to satisfy a particular recursion. In the statement of Proposition 2.1, $(x)_{n\uparrow\tau}$ denotes the generalized rising factorial.

PROPOSITION 2.1. The weights V, W, and U, with the normalizations $W_1 = U_0 = 1$, define a consistent random feature allocation of form (4) iff for some $a, b > 0$ and $\tau \geq 0$
(i) $W_m = (a)_{m-1\uparrow\tau}$, for all $m \in \mathbb{N}$;
(ii) $U_m = (b)_{m\uparrow\tau}$, for all $m \in \mathbb{N}_0$;
(iii) For all $(n, k) \in \mathbb{N} \times \mathbb{N}_0$, V satisfies

$$\begin{aligned} \sum_{j\geq 0} V_{1,j} &= 1, \\ V_{n,k} &= \sum_{j=0}^{\infty} \binom{k+j}{j} \left((b)_{n\uparrow\tau}\right)^j (a + b + \tau(n-1))^k V_{n+1,k+j}. \end{aligned} \tag{7}$$

PROOF. The consistent exchangeable feature allocation with no features with probability one can be represented as in (4), with $V_{n,0} = 1$ for all $n \in \mathbb{N}$ and $V_{n,k} = 0$ for $k \geq 1$. The consistency condition (6) for $k = 1$ gives

$$V_{n,1} W_{m_1} U_{n-m_1} = \sum_{j=0}^{\infty} (j+1) V_{n+1,j+1} U_n^j \left(W_{m_1+1} U_{n-m_1} + W_{m_1} U_{n+1-m_1}\right).$$

This condition implies that, for all $n \in \mathbb{N}$ and for all $m_1 \leq n$,

$$\frac{W_{m_1+1}}{W_{m_1}} + \frac{U_{n+1-m_1}}{U_{n-m_1}} = \frac{V_{n,1}}{\sum_{j=0}^{\infty} (j+1) V_{n+1,j+1} U_n^j}. \tag{8}$$

Since the right hand side of (8) does not depend on m_1, it follows that, for all n and for all $i, j \leq n$,

$$\frac{W_{i+1}}{W_i} - \frac{W_{j+1}}{W_j} = \frac{U_{n+1-j}}{U_{n-j}} - \frac{U_{n+1-i}}{U_{n-i}}.$$

In particular, considering $n = 2$, $i = 2$, and $j = 1$, we find

$$\frac{W_3}{W_2} - W_2 = \frac{U_2}{U_1} - U_1 =: \tau.$$

For $n > 1$, $i = n$, and $j = n - 1$, we also obtain

$$\frac{W_{n+1}}{W_n} - \frac{W_n}{W_{n-1}} = \frac{U_2}{U_1} - U_1 = \tau,$$

which implies, for all $n > 1$,

$$\frac{W_{n+1}}{W_n} = \tau(n-1) + W_2,$$

hence $W_n = (W_2)_{n-1\uparrow\tau}$. In a similar manner, we consider $n > 1$, $i = 2$, and $j = 1$ and obtain

$$\frac{U_n}{U_{n-1}} - \frac{U_{n-1}}{U_{n-2}} = \frac{W_3}{W_2} - W_2 = \tau.$$

As before, this formula implies $U_n = (U_1)_{n\uparrow\tau}$. The recursion (7) follows, by rewriting (6) as

$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} U_n^j V_{n+1,k+j} \sum_{\underline{z}\in\{0,1\}^k} \prod_{i=1}^{k} \frac{W_{m_i+z_i}}{W_{m_l}} \frac{U_{n+1-m_i-z_i}}{U_{n-m_l}},$$

and by noticing that

$$\sum_{\underline{z}\in\{0,1\}^k} \prod_{i=1}^{k} \frac{W_{m_i+z_i}}{W_{m_l}} \frac{U_{n+1-m_i-z_i}}{U_{n-m_l}} = \left(U_1 + W_2 + \tau\,(n-1)\right)^k.$$

Also, $\sum_{j=0}^{\infty} V_{1,j} = 1$ comes from $\sum_{j=0}^{\infty} V_{1,j} W_1^j = 1$ and $W_1 = 1$. Finally, the reverse implication easily follows by checking that the probability distribution with form (4) and V, W and U as in the statement of the proposition satisfies the consistency condition (6).

□

The last possible rescaling can now be fixed by imposing $\tau = 1$. Indeed, let W a sequence of weights parametrizing a feature allocation of form (4). From Proposition 2.1, $W_j = (a)_{j-1\uparrow\tau}$ for some $a > 0$ and some $\tau \geq 0$. If we consider the rescaling $\tilde{W}_j = \kappa^{j-1} W_j$, from $\tilde{W}_j = \kappa^{j-1}\,(a)_{j-1\uparrow\tau} = (\kappa a)_{j-1\uparrow\kappa\tau}$ and Proposition 2.1, we obtain $\tilde{W}_j = (\tilde{a})_{j-1\uparrow\tilde{\tau}}$, where $\tilde{a} = \kappa a$ and $\tilde{\tau} = \kappa\tau$. Therefore, by imposing $\tau = 1$, we avoid the last ambiguity on the rescaling since κ must be equal to 1 and $\tilde{W} = W$.

We now introduce the parametrization, $\alpha := 1 - a$ and $\theta := a + b$, for $\alpha < 1$ and $\theta > -\alpha$. Then, $W_{m_l} = (1-\alpha)_{m_l-1\uparrow}$ and $U_{n-m_l} = (\theta+\alpha)_{n-m_l\uparrow}$, which matches the form of the W and U for the IBP.

2.2 General tools to derive the extreme V

Let $\mathcal{V}_{\alpha,\theta}$ be the set of those elements $V \in \mathbb{R}_+^{\mathbb{N}\times\mathbb{N}_0}$ satisfying (7). Endow this set with the smallest σ-algebra $\mathcal{B}_{\mathcal{V}}$ that makes the maps $V \mapsto V_{n,k}$ measurable and define the barycenter V^μ of each measure μ on $\mathcal{B}_{\mathcal{V}}$ as the pointwise average,

$$V_{n,k}^{\mu} = \int_{\mathcal{V}_{\alpha,\theta}} V_{n,k}\,\mu(\mathrm{d}V). \tag{9}$$

It is easy to check that $\mathcal{V}_{\alpha,\theta}$ is a convex set, i.e., for all probability measures μ on $\mathcal{B}_{\mathcal{V}}$, $V^\mu \in \mathcal{V}_{\alpha,\theta}$ (see Appendix A.2). The goal of this section is to check that this set is also a simplex and to describe its extreme elements.

Given a measurable space of functions with the convex structure just defined, Dynkin Dynkin (1978) describes a general theory which can be applied to show that this set is also a simplex and to determine its extreme points. Similar results have been studied or rediscovered by many works, see references in Gnedin and Pitman (2006). To apply the results of Dynkin (1978) to our problem, we

will follow the same strategy used by Gnedin and Pitman (2006). Rather than studying $\mathcal{V}_{\alpha,\theta}$ directly, we consider another space, isomorphic to $\mathcal{V}_{\alpha,\theta}$ and easier to study, and we find its extreme points applying the results by Dynkin (1978).

Let $(\mathbb{N}_0^\infty, \mathcal{C}(\mathbb{N}_0^\infty))$ be the infinite product space of $\mathbb{N}_0$, endowed with its cylinder σ-algebra. To each $V \in \mathcal{V}_{\alpha,\theta}$ we associate a Markov law, P_V, on this space. Specifically, writing $K_n : \mathbb{N}_0^\infty \to \mathbb{N}_0$ for the n-th coordinate projection on the product space, the Markov law associated to V has the initial distribution given by

$$P_V\left(K_1 = j\right) = V_{1,j}, \tag{10}$$

and transition probabilities

$$P_V\left(K_{n+1} = j + k | K_n = k\right) = \binom{k+j}{j} \left((\alpha+\theta)_{n\uparrow}\right)^j (\theta+n)^k \frac{V_{n+1,k+j}}{V_{n,k}}, \tag{11}$$

if $j \geq 0$ and 0 otherwise. Let $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}} = \{P_V : V \in \mathcal{V}_{\alpha,\theta}\}$ be the set of Markov laws. The map $T : \mathcal{V}_{\alpha,\theta} \to \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, defined by $T(V) = P_V$ is a convex isomorphism (see Appendix A.2 for a proof). Hence, if P_V is extreme in $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, so is V in $\mathcal{V}_{\alpha,\theta}$. We now describe how to find the extreme points of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$. Before that, we remark that, given an EFPF with form (4) parametrized by V, it is straightforward to show that K_n corresponds to the number of features in the corresponding random feature allocation of n individuals, i.e., K_n is the cardinality of $\bigcup_{1\leq i\leq n} X_i$.

As we will see from Proposition 2.2, for every $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma(K_n, K_{n+1}, \ldots)$ is a sufficient σ-algebra for $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$. Hence, for each $n \in \mathbb{N}$, there exists a common $\mathcal{F}_n$-measurable regular conditional probability $Q_n : \mathbb{N}_0^\infty \times \mathcal{C}(\mathbb{N}_0^\infty) \to [0,1]$ for $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ given $\mathcal{F}_n$, such that, for all $P_V \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ and $A \in \mathcal{C}(\mathbb{N}_0^\infty)$,

$$Q_n(\omega, A) = P_V((K_m)_{m\in\mathbb{N}} \in A | \mathcal{F}_n)(\omega), \tag{12}$$

for P_V-almost all $\omega \in \mathbb{N}_0^\infty$. In order to avoid having to repeat uninteresting measure-theoretic details, when $A' \in \sigma(K_1, \ldots, K_n)$, we will take advantage of the Markov property of $(K_m)_{m\in\mathbb{N}}$ to assume that

$$Q_n(\omega, A') = P((K_m)_{m\in\mathbb{N}} \in A' | \mathcal{F}_n)(\omega) = P((K_m)_{m\in\mathbb{N}} \in A' | K_n)(\omega) \tag{13}$$

for all $\omega \in \mathbb{N}_0^\infty$, where we have dropped the V from the notation P_V in order to highlight the independence of the cotransition probabilities under P_V from V itself. This is justified because the equality holds for all $P \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$.

Associated to each Markov kernel Q_n, there is a Markov operator Π_n given by

$$\Pi_n f(\omega) = \int f(\omega') Q_n(\omega, \mathrm{d}\omega'), \tag{14}$$

for all f bounded $\mathcal{F}_n$-measurable real functions. Henceforth, for every σ-algebra $\mathcal{F}$, we will simply write $f \in \mathcal{F}$ to denote that f is bounded and $\mathcal{F}$-measurable. The sequence $(\mathcal{F}_n, \Pi_n)_{n\in\mathbb{N}}$ forms a *specification* in $(\mathbb{N}_0^\infty, \mathcal{C}(\mathbb{N}_0^\infty))$ (see Appendix A.2 for a proof). We can apply Theorem 5.1 of Dynkin (1978), which states that $(\Pi_n)_{n\in\mathbb{N}}$ is an asymptotically H-sufficient statistic, which in turn means (see

also Section 4.4 of Dynkin (1978)) that, for all P_V that are extreme,

$$P_V(\{\omega \in \mathbb{N}_0^\infty : \forall f \in \mathcal{C}(\mathbb{N}_0^\infty),\ \lim_{n\to\infty} \Pi_n f(\omega) = \int f \,\mathrm{d}P_V\}) = 1. \tag{15}$$

A path $\omega \in \mathbb{N}_0^\infty$ induces a Markov law $P_V \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ and is said to be *regular* iff for all $f \in \mathcal{C}(\mathbb{N}_0^\infty)$, $\lim_{n\to\infty} \Pi_n f(\omega) = \int f \,\mathrm{d}P_V$. The set of point in $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ that are induced by regular paths is called the *maximal boundary*. The set of extreme points, also called the *minimal boundary*, is the subset of the maximal boundary, corresponding to those points P_V that also satisfy (15), i.e., they assign probability 1 to the set of regular paths inducing them.

In our context, to identify the maximal boundary, it is enough to check (15) for all functions $f \in \mathcal{C}(\mathbb{N}_0^\infty)$ that are indicators of cylinder sets of the form $K_n^{-1}\{k\}$ for $n \in \mathbb{N}$ and $k \in \mathbb{N}_0$. That is, the elements belonging to the maximal boundary are those $P_{\bar{V}} \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ such that, for some $\omega \in \mathbb{N}_0^\infty$,

$$\lim_{m\to\infty} P(K_n = k|\mathcal{F}_m)(\omega) = \lim_{m\to\infty} P(K_n = k|K_m)(\omega) = P_{\bar{V}}(K_n = k),$$

for all $(n,k) \in \mathbb{N} \times \mathbb{N}_0$. To find the extremes measures of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, we compute the cotransition (backwards) probabilities of $(K_n)_{n\in\mathbb{N}}$.

Proposition 2.2. The cotransition probabilities are

$$P(K_n = k|K_m = l) = \frac{d_{n,k}^{m,l}}{d^{m,l}} d^{n,k}, \tag{16}$$

for $n < m$ and $k \le l$, while the distribution of K_n under $P_V \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ is

$$P_V(K_n = k) = V_{n,k} d^{n,k}, \tag{17}$$

where

$$d^{m,l} = ((\theta+1)_{m-1\uparrow} + \sum_{j=1}^{m-1} (\theta+\alpha)_{m-j\uparrow}(\theta+1+m-j)_{j-1\uparrow})^l, \tag{18}$$

and

$$d_{n,k}^{m,l} = \binom{l}{k} ((\theta+n)_{m-n\uparrow})^k (\sum_{j=1}^{m-n} (\alpha+\theta)_{m-j\uparrow}(\theta+m-1)_{j-1\downarrow})^{l-k}. \tag{19}$$

proof. See Appendix A.3.

□

Note that the cotransition probabilities are independent of V.

2.3 Characterization of $V_{n,k}$

In this section, we study the three cases $0 < \alpha < 1$, $\alpha = 0$, and $\alpha < 0$ separately. Recall that a path $\omega = (\omega_1, \omega_2, \ldots) \in \mathbb{N}_0^\infty$ is regular and induces $\bar{V} \in \mathcal{V}_{\alpha,\theta}$ if the limit

$$\lim_{m\to\infty} P(K_n = k | K_m = \omega_m) = \lim_{m\to\infty} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} d^{n,k} = \bar{V}_{n,k} d^{n,k} \tag{20}$$

exists for all (n,k). In this case, $P_{\bar{V}}$ belongs to the maximal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$. If $P_{\bar{V}}$ also assigns probability one to the set of regular paths inducing it, then $P_{\bar{V}}$ is extreme.

2.3.1 Case $0 < \alpha < 1$

For (α, θ) fixed, s.t. $0 < \alpha < 1$ and $\theta > -\alpha$, let $V^{3IBP,\alpha,\theta}(\gamma)$ be the V of the 3-parameter IBP, defined as

$$\begin{aligned} V_{n,k}^{3IBP,\alpha,\theta}(\gamma) &= \frac{1}{k!}\Big(\frac{\gamma}{(\theta+1)_{n-1\uparrow}}\Big)^k \exp\Big(-\sum_{i=1}^n \gamma \frac{(\alpha+\theta)_{i-1\uparrow}}{(1+\theta)_{i-1\uparrow}}\Big) \\ &= \frac{1}{k!}\Big(\frac{\gamma}{(\theta+1)_{n-1\uparrow}}\Big)^k \exp\Big(-\gamma\Big(\frac{\Gamma(\theta+1)\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)} - \frac{\theta}{\alpha}\Big)\Big), \end{aligned}$$

for all $\gamma \geq 0$. Define $\mathcal{P}_{V^{3IBP,\alpha,\theta}} = \{P_{V^{3IBP,\alpha,\theta}(\gamma)} \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}} : \gamma \geq 0\}$.

Proposition 2.3. Let $0 < \alpha < 1$ and $\theta > -\alpha$.

a) The elements of the set $\mathcal{P}_{V^{3IBP,\alpha,\theta}}$ belong to the maximal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ and they are induced by those paths $w \in \mathbb{N}_0^\infty$ s.t. $\frac{w_m}{m^\alpha} \to c$, where $c = \frac{\gamma\Gamma(\theta+1)}{\alpha\Gamma(\alpha+\theta)}$;

b) The elements of $\mathcal{P}_{V^{3IBP,\alpha,\theta}}$ also belong to the minimal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, i.e., they are extreme points of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$;

c) The elements of $\mathcal{P}_{V^{3IBP,\alpha,\theta}}$ are the only extreme points, i.e., $\mathcal{P}_{V^{3IBP,\alpha,\theta}}$ coincides with the maximal and the minimal boundary.

Proof.

a) In Appendix A.4, we check that

$$\lim_{\substack{m\to\infty \\ \frac{\omega_m}{m^\alpha}\to c}} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} = V_{n,k}^{3IBP,\alpha,\theta}\Big(\frac{c\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}\Big). \tag{21}$$

b) From Theorem 4 of Berti et al. (2015), it follows that

$$P_{V^{3IBP,\alpha,\theta}(\frac{c\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)})}\Big(\frac{K_n}{n^\alpha} \to c\Big) = 1.$$

c) In Appendix A.4, we show that the elements of $\mathcal{P}_{V^{3IBP,\alpha,\theta}}$ are the only ones belonging to the maximal boundary, i.e., there are not any other regular paths except those of part (a).

□

In Proposition 2.3, the case $\gamma = 0$ corresponds to the degenerate feature allocation with no features with probability one, corresponding to $V_{n,0} = 1$ and $V_{n,k} = 0$ for all $n \in \mathbb{N}$ and $k \geq 1$. This solution is induced by the path $\omega_m = 0$ for all $m \in \mathbb{N}$, which has probability one under this degenerate law.

2.3.2 Case $\alpha = 0$

For θ fixed and positive, the V of the 2-parameter IBP are of the form

$$\begin{aligned} V_{n,k}^{2IBP,\theta}(\gamma) &= \frac{1}{k!}\left(\frac{\gamma}{(\theta+1)_{n-1\uparrow}}\right)^k \exp\left(-\sum_{i=1}^{n}\gamma\frac{(\theta)_{i-1\uparrow}}{(1+\theta)_{i-1\uparrow}}\right) \\ &= \frac{1}{k!}\left(\frac{\gamma}{(\theta+1)_{n-1\uparrow}}\right)^k \exp\left(-\gamma\sum_{i=1}^{n}\frac{\theta}{\theta+i-1}\right), \end{aligned}$$

with the convention that, when $\gamma = 0$, we recover the degenerate feature allocation with no features. Define $\mathcal{P}_{V^{2IBP,\theta}} = \{P_{V^{2IBP,\theta}(\gamma)} \in \mathcal{P}_{\mathcal{V}_{0,\theta}} : \gamma \geq 0\}$.

Proposition 2.4. Let $\alpha = 0$ and $\theta > 0$.

a) The elements of the set $\mathcal{P}_{V^{2IBP,\theta}}$ belong the maximal boundary of $\mathcal{P}_{\mathcal{V}_{0,\theta}}$ and they are induced by paths $w \in \mathbb{N}_0^\infty$ s.t. $\frac{w_m}{\log(m)} \to \gamma$;

b) The elements of $\mathcal{P}_{V^{2IBP,\theta}}$ also belong to the minimal boundary of $\mathcal{P}_{\mathcal{V}_{0,\theta}}$, i.e., they are extreme points of $\mathcal{P}_{\mathcal{V}_{0,\theta}}$;

c) The elements of $\mathcal{P}_{V^{2IBP,\theta}}$ are the only extreme points, i.e., $\mathcal{P}_{V^{3IBP,\theta}}$ coincides with the maximal and the minimal boundary.

PROOF.

a) In Appendix A.5, we check that

$$\lim_{\substack{m\to\infty \\ \frac{\omega_m}{\log(m)}\to\gamma}} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} = V_{n,k}^{2IBP,\theta}(\gamma). \tag{22}$$

b) This also follows from Theorem 4 of Berti et al. (2015), which establish that

$$P_{V^{2IBP,\theta}(\gamma)}\left(\frac{K_n}{\log(n)} \to \gamma\right) = 1.$$

c) In Appendix A.5, we check that there are not any other regular paths but those of part (a).

□

2.3.3 Case $\alpha < 0$

From formula (3), we see that the Beta–Bernoulli is of form (4), with V of the form

$$V_{n,k}^{BB,\alpha,\theta}(N) = \frac{\binom{N}{k}\left(\frac{-\alpha\Gamma(\theta+\alpha)}{\Gamma(\theta+\alpha+n)}\right)^k}{\left(\frac{\Gamma(\theta+\alpha)\Gamma(\theta+n)}{\Gamma(\theta+\alpha+n)\Gamma(\theta)}\right)^N}. \tag{23}$$

for all $N \in \mathbb{N}$. As before, when $N = 0$, we consider the feature allocation with no feature with probability one. Define $\mathcal{P}_{V^{BB,\alpha,\theta}} = \{P_{V^{BB,\alpha,\theta}(N)} \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}} : N \in \mathbb{N}_0\}$.

Proposition 2.5. Let $\alpha < 0$ and $\theta > -\alpha$.

a) The elements of the set $\mathcal{P}_{V^{BB,\alpha,\theta}}$ belong the maximal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ and they are induced by paths $w \in \mathbb{N}_0^\infty$ s.t. $w_m \to N$;

b) The elements of $\mathcal{P}_{V^{BB,\alpha,\theta}}$ also belong to the minimal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, i.e., they are extreme points of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$;

c) The elements of $\mathcal{P}_{V^{BB,\alpha,\theta}}$ are the only extreme points, i.e., $\mathcal{P}_{V^{BB,\alpha,\theta}}$ coincides with the maximal and the minimal boundary.

PROOF.

a) In Appendix A.6, we check that

$$\lim_{\substack{m\to\infty \\ \omega_m \to N}} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} = V_{n,k}^{BB,\alpha,\theta}(N). \tag{24}$$

b) This follows since under a Beta–Bernoulli model with N features, $K_m \to N$ a.s. Indeed, the probability of each feature, q_j, is a.s. strictly positive, being Beta distributed. The probability of this feature having all zeros in a m-individuals allocation is $(1-q_k)^m$, which tends to zero as $m \to \infty$.

c) In Appendix A.6, we check that there are not any other regular paths but those of part (a).

□

3 Discussion

In this work, we have considered the class of consistent exchangeable feature allocations with EPFP of the form (4). While this is a tractable family, the only elements of this class are mixtures over γ of the 2 and 3-parameter IBP or mixtures over N of the Beta–Bernoulli model. From both an applied and theoretical perspective, it would be of interest to have larger but still tractable classes of exchangeable feature allocations. Finding new tractable priors for feature models is still an active area of research. A possible direction of research would be to study a more general class than (4), with form $V_{n,k}\prod_{l=1}^k W_{n,m_l}$, for a triangular array $W = (W_{n,k} : n \in \mathbb{N}, 0 \le k \le n)$. However, a characterization of W in this case would seem to be much more complicated than Proposition 2.1.

A Appendix: Some Facts and Proofs

A.1 EFPF of the Beta–Bernoulli model

The *Beta–Bernoulli* is described by considering a finite space of features, which can numbered using the integers in $[N]$, where N is the cardinality of the feature set. To each feature we associate a random number q_j with distribution $\text{Beta}(\eta_1, \eta_2)$. Each individual X_i possesses feature j with probability q_j. Let $Z_{i,j}$ be a binary random variable denoting the presence or absence of feature j in individual i. Then, the Beta–Bernoulli model can be written as

$$\begin{aligned} Z_{i,j}|q_j &\sim \text{Bernoulli}(p_j) && i = 1, \dots, n;\ j = 1, \dots, N; \\ q_j|\eta_1, \eta_2 &\sim \text{Beta}(\eta_1, \eta_2) && j = 1, \dots, N. \end{aligned}$$

The conditional probability that $Z = (Z_{i,j})_{i \le n, j \le N}$ is equal to $z = (z_{i,j})_{i \le n, j \le N}$ given $q = (q_1, \dots, q_N)$ is

$$p(Z = z|q) = \prod_{j=1}^{N} \prod_{i=1}^{n} \text{Bernoulli}(z_{i,j}|q_j).$$

Integrating q out, we obtain the probability mass function of Z

$$p(Z = z) = \left(\frac{\Gamma(\eta_1 + \eta_2)}{\Gamma(\eta_1)\Gamma(\eta_2)} \right)^N \prod_{i=1}^{N} \frac{\Gamma(m_i + \eta_1)\Gamma(n - m_i + \eta_2)}{\Gamma(n + \eta_1 + \eta_2)},$$

where $m_i = \sum_{j=1}^{n} z_{i,j}$. If $(m_1, \dots, m_N)$ has k non-zero entries, this probability becomes

$$\left(\frac{\Gamma(\eta_1 + \eta_2)}{\Gamma(\eta_1)\Gamma(\eta_2)\Gamma(n + \eta_1 + \eta_2)} \right)^N \left(\Gamma(\eta_1)\Gamma(n + \eta_2)\right)^{N-k} \prod_{i=1}^{k} \Gamma(m_i + \eta_1)\Gamma(n - m_i + \eta_2).$$

Finally, taking into account all $\binom{N}{k}$ possible uniform orderings of the $N - k$ features not possessed by any individual, which give rise to the same uniformly ordered feature allocation, we obtain

$$\binom{N}{k} \left(\frac{\Gamma(\eta_1 + \eta_2)}{\Gamma(\eta_1)\Gamma(\eta_2)\Gamma(n + \eta_1 + \eta_2)} \right)^N \left(\Gamma(\eta_1)\Gamma(n + \eta_2)\right)^{N-k} \cdot \prod_{i=1}^{k} \Gamma(m_i + \eta_1)\Gamma(n - m_i + \eta_2),$$

which can be be rewritten as in formula (3), by using rising factorials and by changing the parametrization to $\alpha = -\eta_1$ and $\theta = \eta_2 + \eta_1$, with $\alpha < 0$ and $\theta > -\alpha$.

A.2 Some Facts

Proposition A.1. $\mathcal{V}_{\alpha,\theta}$ is a convex set.

proof. We want to show that $\mathcal{V}_{\alpha,\theta}$ is a convex set, i.e., for all probability measures μ on $\mathcal{B}_{\mathcal{V}}$, $V^\mu \in \mathcal{V}_{\alpha,\theta}$. We have

$$V^\mu_{n,k} = \int_{\mathcal{V}_{\alpha,\theta}} V_{n,k} \mu(\mathrm{d}V)$$

$$
\begin{aligned}
&= \int_{\mathcal{V}_{\alpha,\theta}} \sum_{j=0}^{\infty} \binom{k+j}{j} ((\alpha+\theta)_{n\uparrow})^j (\theta+n)^k V_{n+1,k+j} \mu(\mathrm{d}V) \\
&= \sum_{j=0}^{\infty} \binom{k+j}{j} ((\alpha+\theta)_{n\uparrow})^j (\theta+n)^k \int_{\mathcal{V}_{\alpha,\theta}} V_{n+1,k+j} \mu(\mathrm{d}V) \\
&= \sum_{j=0}^{\infty} \binom{k+j}{j} ((\alpha+\theta)_{n\uparrow})^j (\theta+n)^k V^{\mu}_{n+1,k+j},
\end{aligned}
$$

for all (n,k), where the first and last equality follow from the definition of barycenter, and the second from the monotone convergence theorem. In a similar manner,

$$
\begin{aligned}
\sum_{j=0}^{\infty} V^{\mu}_{1,j} &= \sum_{j=0}^{\infty} \int_{\mathcal{V}_{\alpha,\theta}} V_{1,j} \mu(\mathrm{d}V) \\
&= \int_{\mathcal{V}_{\alpha,\theta}} \sum_{j=0}^{\infty} V_{1,j} \mu(\mathrm{d}V) \\
&= \int_{\mathcal{V}_{\alpha,\theta}} 1 \mu(\mathrm{d}V) = 1.
\end{aligned}
$$

□

Proposition A.2. $T(V) = P_V$ is an isomorphism between convex sets.

proof. According to (Dynkin, 1978, pg 706), the map $T(V) = P_V$ is a convex isomorphism if T is invertible and T and T^{-1} are measurable and preserve the convex structure. T is 1-1 from Proposition 2.2 and it is onto by construction. We prove T is measurable and preserves the convex structure, proving that the same is true for T^{-1} can be done in similar way.

$\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$ is endowed with the smallest σ-algebra $\mathcal{B}_{\mathcal{P}}$ that makes the maps $P_V \mapsto P_V(A)$ measurable for all $A \in \mathcal{C}(\mathbb{N}_0^{\infty})$. A generator of this σ-algebra is composed by sets $\{P_V \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}} : P_V(K_n = k) \leq x\}$ for $(n,k) \in \mathbb{N} \times \mathbb{N}_0$ and $x \in [0,1]$. The inverse image under T of this set is $\{V \in \mathcal{V}_{\alpha,\theta} : V_{n,k} d_{n,k} \leq x\}$ (see Proposition 2.2 for the definition of $d_{n,k}$), which lies in $\mathcal{B}_{\mathcal{V}}$. Hence, T is measurable.

T preserves the convex structure if, for every measure μ on $\mathcal{B}_{\mathcal{V}}$, we have $T(V^{\mu}) = P^{\mu'}$, where μ' the push-forward measure of μ on $\mathcal{B}_{\mathcal{P}}$ (i.e., $\mu' = \mu \circ T^{-1}$), and $P^{\mu'}$ is the barycenter of μ', defined as

$$
P^{\mu'}(A) = \int_{\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}} P(A) \mu'(\mathrm{d}P), \tag{25}
$$

for all $A \in \mathcal{C}(\mathbb{N}_0^{\infty})$. Using the change of variable formula, it is easy to check that T preserves the convex structure. Indeed, considering cylinder sets of the form $K_n^{-1}\{k\}$ for $n \in \mathbb{N}$ and $k \in \mathbb{N}_0$, we have

$$
\begin{aligned}
P^{\mu'}(K_n = k) &= \int_{\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}} P(K_n = k) \mu'(\mathrm{d}P) = \int_{\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}} P(K_n = k) \mu \circ T^{-1}(\mathrm{d}P) \\
&= \int_{\mathcal{V}_{\alpha,\theta}} d_{n,k} V_{n,k} \mu(\mathrm{d}V) = d_{n,k} V^{\mu}_{n,k}.
\end{aligned}
$$

Hence, $T(V^\mu) = P^{\mu'}$.

$\square$

PROPOSITION A.3. $(\mathcal{F}_n, \Pi_n)_{n\in\mathbb{N}}$ forms a specification in $(\mathbb{N}_0^\infty, \mathcal{C}(\mathbb{N}_0^\infty))$.

PROOF. According to Dynkin (1978), section 5.1, given a directed set L and a measurable space $(\Lambda, \mathcal{F})$, a *specification* on this is space $(\mathcal{F}_\Lambda, \Pi_\Lambda)_{\Lambda\in L}$ is a family of sub-σ-algebras and Markov operators satisfying

(i) $\mathcal{F}_{\Lambda'} \subseteq \mathcal{F}_\Lambda$, if $\Lambda' \succeq \Lambda$;

(ii) $\Pi_{\Lambda'}\Pi_\Lambda = \Pi_{\Lambda'}$, if $\Lambda' \succeq \Lambda$;

(iii) $\Pi_\Lambda f \in \mathcal{F}_\Lambda$, for all $f \in \mathcal{F}$;

(iv) $\Pi_\Lambda f = f$, for all $f \in \mathcal{F}_\Lambda$.

In our context, with $L = \mathbb{N}$ and the sub-σ-algebras and Markov operators defined in section 3.1, formula (14), (i), (ii), and (iv) follow immediately. To check (ii), it is enough to check for indicators of measurable sets. In particular, for $f = \mathbb{1}_A$, with $A \in \mathcal{C}(\mathbb{N}_0^\infty)$, we must check

$$\int_{\mathbb{N}_0^\infty} Q_n(\omega', A) Q_{n+1}(\omega, \mathrm{d}\omega') = Q_{n+1}(\omega, A).$$

Indeed, it is enough to check this condition for a thin cylinder A of the form $K_1^{-1}\{k_1\} \cap K_2^{-1}\{k_2\} \cap \ldots K_m^{-1}\{k_m\}$ for $m > n+1$ and $k_i \in \mathbb{N}_0$ for all $i \leq m$,

$$\begin{aligned}
&\int_{\mathbb{N}_0^\infty} Q_n(\omega', A) Q_{n+1}(\omega, \mathrm{d}\omega') \\
&\quad = \int_{\mathbb{N}_0^\infty} P(K_1 = k_1, \ldots, K_m = k_m | \mathcal{F}_n)(\omega') P((K_l)_{l\in\mathbb{N}} \in \mathrm{d}\omega' | \mathcal{F}_{n+1})(\omega) \\
&\quad = \int_{\mathbb{N}_0^\infty} P(K_1 = k_1, \ldots, K_{n-1} = k_{n-1} | K_n = k_n) \\
&\qquad\qquad \cdot \mathbb{1}(\omega'_n = k_n) \ldots \mathbb{1}(\omega'_m = k_m) P((K_l)_{l\in\mathbb{N}} \in \mathrm{d}\omega' | \mathcal{F}_{n+1})(\omega) \\
&\quad = P(K_1 = k_1, \ldots, K_{n-1} = k_{n-1} | K_n = k_n) \\
&\qquad\qquad \cdot \int_{\mathbb{N}_0^\infty} \mathbb{1}(\omega'_n = k_n) \ldots \mathbb{1}(\omega'_m = k_m) P((K_l)_{l\in\mathbb{N}} \in \mathrm{d}\omega' | \mathcal{F}_{n+1})(\omega) \\
&\quad = P(K_1 = k_1, \ldots, K_{n-1} = k_{n-1} | K_n = k_n) P(K_n = k_n | \mathcal{F}_{n+1})(\omega) \\
&\qquad\qquad \cdot \mathbb{1}(\omega_{n+1} = k_{n+1}) \ldots \mathbb{1}(\omega_m = k_m) \\
&\quad = P(K_1 = k_1, \ldots, K_m = k_m | \mathcal{F}_{n+1})(\omega) = Q_{n+1}(\omega, A).
\end{aligned}$$

$\square$

A.3 Proof of Proposition 2.2

First, note that for $m > n$ and $l \geq k$ $P_V(K_m = l|K_n = k) = V_{m,l} d^{m,l}_{n,k}$, for a function $d^{m,l}_{n,k}$ independent of V. Indeed, from (11), the probability of a path $(k_{n+1}, k_{n+2}, \ldots, k_{m-1}, l)$ depends only on the last $V_{m,l}$. Summing over all possible paths from $K_n = k$ to $K_m = l$, we see that $P_V(K_m = l|K_n = k)$ must be of the form $V_{m,l} d^{m,l}_{n,k}$. In addition, by considering $P_V(K_m = l) = \sum_{i=0}^{l} P_V(K_m = l|K_1 = i) \cdot P_V(K_1 = i)$, $P_V(K_m = l)$ must be of the form $V_{m,l} d^{m,l}$. Also, from

$$P_V\left(K_m = l\right) = \sum_{j=0}^{l} P_V\left(K_m = l|K_{m-1} = j\right) \cdot P_V\left(K_{m-1} = j\right),$$

it follows that, for $l > 2$, the function $d^{m,l}$ must satisfy

$$V_{m,l} d^{m,l} = \sum_{j=0}^{l} \frac{V_{m,l}}{V_{m-1,j}} \tbinom{l}{l-j} ((\alpha+\theta)_{m-1\uparrow})^{l-j} (\theta+m-1)^j V_{m-1,j} d^{m-1,j},$$

which gives the following the recursion

$$d^{m,l} = \sum_{j=0}^{l} \tbinom{l}{l-j} ((\alpha+\theta)_{m-1\uparrow})^{l-j} (\theta+m-1)^j d^{m-1,j}.$$

Substituting $d^{m-1,j}$, we find

$$d^{m,l} = \sum_{j=0}^{l} \tbinom{l}{l-j} ((\alpha+\theta)_{m-1\uparrow})^{l-j} (\theta+m-1)^j \sum_{i=0}^{j} \binom{j}{j-i} ((\alpha+\theta)_{m-2\uparrow})^{j-i} (\theta+m-2)^i d^{m-2,i}.$$

Grouping together all coefficient multiplying $d^{m-2,k}$ on the right hand side ($0 \leq k \leq l$), we find

$$\begin{aligned}
d^{m,l}_{n-2,k} &= \tbinom{l}{l-k} \left((\alpha+\theta)_{m-1\uparrow}\right)^{l-k} (\theta+m-1)^k (\theta+m-2)^k \\
&+ \tbinom{l}{l-k-1} \left((\alpha+\theta)_{m-1\uparrow}\right)^{l-k-1} (\theta+m-1)^{k+1} \left((\alpha+\theta)_{m-2\uparrow}\right) (\theta+m-2)^k \\
&+ \tbinom{l}{l-k-2} \left((\alpha+\theta)_{m-1\uparrow}\right)^{l-k-2} (\theta+m-1)^{k+2} \binom{k+2}{2} \left((\alpha+\theta)_{m-2\uparrow}\right)^2 (\theta+m-2)^k \\
&\quad \vdots \\
&+ \left((\alpha+\theta)_{m-1\uparrow}\right) (\theta+m-1)^{l-1} \binom{l-1}{l-1-k} \left((\alpha+\theta)_{m-2\uparrow}\right)^{l-1-k} (\theta+m-2)^k \\
&+ (\theta+m-1)^l \binom{l}{l-k} \left((\alpha+\theta)_{m-2\uparrow}\right)^{l-k} (\theta+m-2)^k \\
&= ((\theta+m-1)(\theta+m-2))^k \left((\alpha+\theta)_{m-1\uparrow} + (\theta+m-1)(\alpha+\theta)_{m-2\uparrow}\right)^{l-k} \binom{l}{l-k}.
\end{aligned}$$

So, the recursion for $d^{m,l}$ becomes

$$d^{m,l} = \sum_{j=0}^{l} \binom{l}{l-j} \left((\theta+m-1)(\theta+m-2)\right)^j \left((\alpha+\theta)_{m-1\uparrow} + (\theta+m-1)(\alpha+\theta)_{m-2\uparrow}\right)^{l-j} d^{m-2,j}.$$

In the same manner, we find

$$\begin{aligned} d^{m,l}_{m-3,k} = &\left((\theta+m-1)(\theta+m-2)(\theta+m-3)\right)^k \binom{l}{l-k} \\ &\cdot \left((\alpha+\theta)_{m-1\uparrow} + (\theta+m-1)(\alpha+\theta)_{m-2\uparrow} + (\theta+m-1)(\theta+m-2)(\alpha+\theta)_{m-3\uparrow}\right)^{l-k}. \end{aligned}$$

Finally, we obtain,

$$\begin{aligned} d^{m,l}_{n,k} &= \binom{l}{l-k} ((\theta+m-1)_{m-n\downarrow})^k \Big(\sum_{j=1}^{m-n} (\alpha+\theta)_{m-j\uparrow} (\theta+m-1)_{j-1\downarrow}\Big)^{l-k} \\ &= \binom{l}{k} ((\theta+n)_{m-n\uparrow})^k \Big(\sum_{j=1}^{m-n} (\alpha+\theta)_{m-j\uparrow} (\theta+m-1)_{j-1\downarrow}\Big)^{l-k}. \end{aligned}$$

In addition,

$$\begin{aligned} d^{m,l} &= \sum_{i=0}^{l} d^{m,l}_{1,i} = \sum_{i=0}^{l} \binom{l}{i} ((\theta+1)_{m-1\uparrow})^i \Big(\sum_{j=1}^{m-1} (\alpha+\theta)_{m-j\uparrow} (\theta+m-1)_{j-1\downarrow}\Big)^{l-i} \\ &= \Big((\theta+1)_{m-1\uparrow} + \sum_{j=1}^{m-1} (\alpha+\theta)_{m-j\uparrow} (\theta+1+m-j)_{j-1\uparrow}\Big)^l. \end{aligned}$$

□

A.4 Proof of Proposition 2.3

We begin with some technical results about asymptotic equivalence of functions: Write $f \approx g$ to denote that $f(m)/g(m) \to 1$ as $m \to \infty$, and note that $f_i \approx g_i$ implies $f_1 + f_2 \approx f_1 + g_2 \approx g_1 + g_2$ and $f_1 f_2 \approx f_1 g_2 \approx g_1 g_2$. In general, it does not hold that $f_1^{f_2} \approx g_1^{g_2}$. The following two results characterize special cases:

Lemma A.4. Let $g(m) \to \infty$ as $m \to \infty$, let $f \approx g$ and $h/g \to c$ for some constant $c \geq 0$. Then, for every $p, q \in \mathbb{R}$, we have

$$\left(\frac{f(m)-p}{f(m)-q}\right)^{h(m)} \approx \left(\frac{g(m)-p}{g(m)-q}\right)^{c\,g(m)} \to e^{c\,(q-p)} \tag{26}$$

proof We prove only the first equivalence, because the limiting exponential form is well known.

Taking logarithms, we have

$$\log\left[\left(\frac{f(m)-p}{f(m)-q}\right)^{h(m)}\left(\frac{g(m)-q}{g(m)-p}\right)^{c\,g(m)}\right] \tag{27}$$

$$= c\,g(m)\,\log\frac{(f(m)-p)\,(g(m)-q)}{(f(m)-q)\,(g(m)-p)} \tag{28}$$

$$+\,(h(m)-c\,g(m))\,\log\frac{f(m)-p}{f(m)-q}. \tag{29}$$

The arguments to the logarithms can we written as

$$\frac{(f(m)-p)\,(g(m)-q)}{(f(m)-q)\,(g(m)-p)} = 1+\frac{(f(m)-g(m))\,(p-q)}{(f(m)-q)(g(m)-p)} \tag{30}$$

and

$$\frac{f(m)-p}{f(m)-q} = 1+\frac{q-p}{f(m)-q}. \tag{31}$$

Using the fact that $z(m)\to 0$ implies $\log(1+z(m))\approx z(m)$, and that both terms (30) and (31) converge to one, it follows that

$$(27)\approx c\,g(m)\frac{(f(m)-g(m))\,(p-q)}{(f(m)-q)(g(m)-p)}+(h(m)-c\,g(m))\,\frac{q-p}{f(m)-q}. \tag{32}$$

It is straightforward to show that the r.h.s. converges to 0.

□

LEMMA A.5. Let $f(m)\to\infty$ as $m\to\infty$, let $g(m)/f(m)\to\infty$ as $m\to\infty$, and let $h\approx f$. For every $p>q$,

$$\left(\frac{g(m)}{f(m)}\right)^{k}\left(\frac{h(m)-p}{h(m)-q}\right)^{g(m)}\to 0 \qquad \text{as } m\to\infty.$$

PROOF. Taking logarithms

$$\begin{aligned}
&k\log\frac{g(m)}{f(m)}+g(m)\log\Big\{1+\frac{q-p}{h(m)-q}\Big\}\\
&\approx k\log\frac{g(m)}{f(m)}+(q-p)\frac{g(m)}{h(m)-q}\\
&\approx k\log\frac{g(m)}{f(m)}+(q-p)\frac{g(m)}{f(m)}\to-\infty
\end{aligned}$$

as $m\to\infty$, completing the proof.

□

We now proceed to prove each part of Proposition 2.3:

a) We must check the limit (21). From Proposition 2.2,

$$
\begin{aligned}
\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} &= \binom{\omega_m}{k} \frac{\left[(\theta+m-1)_{m-n\downarrow}\right]^k \left[\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}\right]^{\omega_m-k}}{\left[(\theta+m-1)_{m-1\downarrow}+\sum_{i=1}^{m-1}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}\right]^{\omega_m}} \\
&= \binom{\omega_m}{k}\left[\frac{(\theta+m-1)_{m-n\downarrow}}{\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}\right]^k \\
&\quad \cdot \left[\frac{\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}{(\theta+m-1)_{m-1\downarrow}+\sum_{i=1}^{m-1}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}\right]^{\omega_m} \\
&= \binom{\omega_m}{k}\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta+n)}}{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}\right]^k \\
&\quad \cdot \left[\frac{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta+1)}+\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+1)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+1)}}\right]^{\omega_m},
\end{aligned}
\tag{33}
$$

where the third equality follows from the identity

$$
\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow} = \frac{\Gamma(\alpha+\theta+m)}{\alpha\Gamma(\alpha+\theta)} - \frac{\Gamma(\theta+m)\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)},
$$

which itself arises from rewriting the sum as a difference of two infinite hypergeometric series evaluated at 1 and applying the Gauss theorem for hypergeometric series.

Using the asymptotic equivalence $\Gamma(m+\delta)\approx\Gamma(m)m^\delta$ and limit $(m+\theta)^\alpha - m^\alpha \to 0$, the Stirling formula $\binom{\omega_m}{k}\approx\frac{1}{k!}\omega_m^k$ for the binomial coefficient, and then the limit $\frac{\omega_m}{m^\alpha}\to c$, the first line of (33) can be simplified to yield a limiting form:

$$
\begin{aligned}
&\binom{\omega_m}{k}\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta+n)}}{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}\right]^k \\
&\approx \binom{\omega_m}{k}\left[\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{m^\alpha}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \\
&\approx \frac{1}{k!}\left[\omega_m m^{-\alpha}\cdot\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{1}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{m^\alpha\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \\
&\approx \frac{1}{k!}\left[c\cdot\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{1}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{m^\alpha\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \to \frac{1}{k!}\left(c\frac{\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+n)}\right)^k.
\end{aligned}
\tag{34}
$$

Similarly, the second line of (33) can be simplified by Lemma A.4, using the asymptotic equivalence $\Gamma(m+\delta)\approx\Gamma(m)m^\delta$ and the limits $(m+\theta)^\alpha - m^\alpha\to 0$ and $\frac{\omega_m}{m^\alpha}\to c$, to yield

$$
\begin{aligned}
&\left[\frac{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta+1)}+\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+1)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+1)}}\right]^{\omega_m} \\
&\approx \left[\frac{m^\alpha-\frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}}{m^\alpha-\frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}}\right]^{cm^\alpha} \to \exp\left\{c\left(\frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}-\frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}\right)\right\}.
\end{aligned}
\tag{35}
$$

Substituting back into (33), we obtain the V of the 3-parameter IBP.

c) To check that the only regular paths are those paths $\omega \in \mathbb{N}_0^{\mathbb{N}}$ such that $\frac{w_m}{m^\alpha} \to c$ for some $c \geq 0$, suppose otherwise; i.e., let $\omega \in \mathbb{N}_0^{\mathbb{N}}$ be a regular path, but assume $\frac{w_m}{m^\alpha}$ does not converge to some finite $c \geq 0$. If $(\frac{w_m}{m^\alpha})_{m\in\mathbb{N}}$ has at least two distinct subsequential limits, then, from the proof of part (a), we see that $d_{n,k}^{m,\omega_m}/d^{m,\omega_m}$ has at least two distinct subsequential limits, a contradiction, and so $\frac{w_m}{m^\alpha} \to \infty$. But then it follows from equations (33), (34), and (35); the asymptotic equivalence $\Gamma(m+\delta) \approx \Gamma(m)m^\delta$ and limit $(m+\theta)^\alpha - m^\alpha \to 0$; and finally an application of Lemma A.4 that $\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} \to 0$ as $m \to \infty$ for every $k \in \mathbb{N}_0$. As these limits must define a probability distribution, this is a contradiction, completing the proof. □

A.5 Proof of Proposition 2.4

a) We must check the limit (22). From Proposition 2.2,

$$\begin{aligned}\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} &= \frac{\binom{\omega_m}{k}\left[(\theta+m-1)_{m-n\downarrow}\right]^k\left[\sum_{i=1}^{m-n}(\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}\right]^{l-k}}{\left[(\theta+m-1)_{m-1\downarrow}+\sum_{i=1}^{m-1}(\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}\right]^{\omega_m}}\\ &= \binom{\omega_m}{k}\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta+n)}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta)}\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}\right]^k\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta)}\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta)}\sum_{i=1}^{m-1}\frac{1}{\theta+m-i}+\frac{\Gamma(\theta+m)}{\Gamma(\theta+1)}}\right]^{\omega_m},\end{aligned}$$

where the second equality follows from the identity

$$\sum_{i=1}^{m-n}(\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow} = (\theta)_{m\uparrow}\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}.$$

Using the Stirling formula for the binomial coefficient, $\binom{\omega_m}{k} \approx \frac{1}{k!}\omega_m^k$, and the identity $\sum_{i=1}^{m}\frac{1}{\theta+m-i} = \sum_{i=1}^{m-n}\frac{1}{\theta+m-i} + \sum_{i=1}^{n}\frac{1}{\theta+n-i}$, we have

$$\begin{aligned}\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} &\approx \frac{1}{k!}\left[\omega_m\frac{\frac{\Gamma(\theta)}{\Gamma(\theta+n)}}{\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}\right]^k\left[\frac{\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}{\sum_{i=1}^{m}\frac{1}{\theta+m-i}}\right]^{\omega_m}\\ &\approx \frac{1}{k!}\left[\frac{\omega_m}{\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}\frac{\Gamma(\theta)}{\Gamma(\theta+n)}\right]^k\left[\frac{\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}}{\sum_{i=1}^{m-n}\frac{1}{\theta+m-i}+\sum_{i=1}^{n}\frac{1}{\theta+n-i}}\right]^{\omega_m}.\end{aligned}$$

Therefore, by Lemma A.4 and the fact that $\log(m) \approx \sum_{i=1}^{m-n}\frac{1}{\theta+m-i}$ and $\omega_m/\log(m) \to \gamma$, we have

$$\begin{aligned}\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} &\approx \frac{1}{k!}\left[\gamma\frac{\Gamma(\theta)}{\Gamma(\theta+n)}\right]^k\left[\frac{\log(m)}{\log(m)+\sum_{i=1}^{n}\frac{1}{\theta+n-i}}\right]^{\gamma\log(m)}\\ &\to \frac{1}{k!}\left[\gamma\frac{\Gamma(\theta)}{\Gamma(\theta+n)}\right]^k\exp\Big(-\gamma\sum_{i=1}^{n}\frac{1}{\theta+n-i}\Big),\end{aligned}$$

as $m \to \infty$, recovering the V of the 2-parameter IBP.

c) To check that the only regular paths are those paths $\omega \in \mathbb{N}_0^{\mathbb{N}}$ such that $\frac{\omega_m}{\log(m)} \to \gamma$ for some $\gamma \geq 0$, we can repeat the same argument as in the proof of Proposition 2.3, part (c). First, we note

that if $(\frac{w_m}{\log(m)})_{m\in\mathbb{N}}$ has at least two distinct subsequential limits, then ω cannot be regular, because the proof of point (a) shows that there will be two distinct induced laws, a contradiction. If $\frac{\omega_m}{\log(m)} \to \infty$ as $m \to \infty$, then it follows again from Lemma A.4 that $\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} \to 0$ as $m \to \infty$ for all $k \in \mathbb{N}_0$, a contradiction.

□

A.6 Proof of Proposition 2.5

a) We must check the limit (24). Starting with Proposition 2.2 and following similar steps as for the case $0 < \alpha < 1$, we obtain the approximation

$$\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} \approx \binom{\omega_m}{k} \left(\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{m^\alpha}{\alpha\Gamma(\alpha+\theta)} - \frac{\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}} \right)^k \left(\frac{m^\alpha - \frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}}{m^\alpha - \frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}} \right)^{\omega_m}, \tag{36}$$

assuming $\omega_m \to N$ and $\alpha < 0$. Taking the limit as $m \to \infty$, we obtain

$$\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} \to \binom{N}{k} \left[\frac{-\alpha\Gamma(\alpha+\theta)}{\Gamma(\alpha+\theta+n)} \right]^k \left[\frac{\Gamma(\alpha+\theta+n)\,\Gamma(\theta)}{\Gamma(\alpha+\theta)\,\Gamma(\theta+n)} \right]^N.$$

c) By the a.s. monotonicity of regular paths, as $m \to \infty$, the number of features ω_m either diverges or converges to a finite (integer) limit. The divergent paths cannot be regular for $\alpha < 0$, because for these paths, (36) diverges as $m \to \infty$. Hence, the only regular paths are those of part (a).

□

Acknowledgements

S. Favaro is supported by the European Research Council through StG N-BNP 306406. D.M. Roy is funded by an NSERC Discovery Grant, Connaught New Researcher Award, AFOSR grant FA9550-15-1-0074, and Newton Alumni grant. Y.W. Teh research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071.

References

Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P. (2015). Central limit theorems for an Indian buffet model with random weights. *Annals of Applied Probability* **25**(2) 523–547.

Broderick, T., Pitman, J. and Jordan, M.I. (2013). Feature Allocations, Probability Functions, and Paintboxes. *Bayesian Analysis* **8**(4) 801–836.

Dynkin, E.B. (1978). Sufficient Statistics and Extreme Points. *Annals of Probability* **6**(5) 705–730.

Ghahramani, Z., Griffiths, T.L. and Sollich, P. (2007). Bayesian nonparametric latent feature models. In *Bayesian Statistics 8*. Oxford University Press, Oxford.

Gnedin, A. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Sciences* **138**(3) 5674–5685.

Griffiths, T.L. and Ghahramani, Z. (2006). Infinite latent feature models and the Indian buffet process. *Advances in Neural Information Processing Systems* **18**.

Griffiths, T.L. and Ghahramani, Z. (2011). The Indian Buffet Process: An Introduction and Review. *Journal of Machine Learning Research* **12** 1185–1224.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102** 145–158.

Pitman, J. (2006). *Combinatorial Stochastic Processes*. Springer-Verlag, Berlin.

Teh, Y. and Görür, D. (2010). Indian buffet processes with power-law behavior. *Advances in Neural Information Processing Systems* **22**.