The Annals of Applied Probability
2018, Vol. 28, No. 3, 1423–1448
https://doi.org/10.1214/17-AAP1333

A CHARACTERIZATION OF PRODUCT-FORM EXCHANGEABLE FEATURE PROBABILITY FUNCTIONS

BY MARCO BATTISTON[*,1], STEFANO FAVARO[†,2,3], DANIEL M. ROY[‡,4] AND YEE WHYE TEH[*,1]

University of Oxford[*], *University of Torino*[†] *and University of Toronto*[‡]

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 $(m_1, \dots, m_k)$ for these features. Each element of this class is parametrized by a countable matrix V and two sequences U and W of nonnegative 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. We prove that the only members of this class satisfying the consistency condition are mixtures of three-parameter Indian buffet Processes over the mass parameter γ, mixtures of N-dimensional Beta–Bernoulli models over the dimension N, or degenerate limits thereof. Hence, we provide a characterization of these two models as the only consistent exchangeable feature allocations having the required product form, up to randomization of the parameters.

1. Introduction. Random feature allocations are popular models within machine learning. These models posit a set of n individuals, each possessing a random (possibly empty) set of features. Specifically, let $(\mathcal{X}, \mathcal{B})$ be a 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 n individuals $X_1, \dots, X_n$, a *feature allocation* of $[n] := \{1, \dots, n\}$, denoted $\{F_{n,1}, \dots, F_{n,K_n}\}$, is a multiset of nonempty subsets of $[n]$, where K_n is the total number of observed features, that is, the cardinality of $\bigcup_{1 \le i \le n} X_i$, and $F_{n,j}$ is the subset of $[n]$ collecting the indices of those individuals having the jth feature. Intuitively, given n individuals $X_1, \dots, X_n$, the associated feature allocation forgets the particular feature values, but retains only the information about the sharing of features among individuals.

Received July 2016; revised June 2017.

[1]Supported in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071.

[2]Supported by the European Research Council through StG N-BNP 306406.

[3]Also affiliated to Collegio Carlo Alberto.

[4]Supported in part by an NSERC Discovery Grant, Connaught Award and U.S. Air Force Office of Scientific Research Grant FA9550-15-1-0074.

MSC2010 subject classifications. Primary 60K99; secondary 60C99.

Key words and phrases. Indian buffet process, exchangeable feature allocations, Gibbs-type partitions.

A random feature allocation is *exchangeable* when its distribution is invariant under permutations of the indexes of the individuals, that is, the feature allocation induced by the random sets $\{X_1, \dots, X_n\}$ is equal in distribution to that induced by $\{X_{\sigma(1)}, \dots, X_{\sigma(n)}\}$, for all permutation σ of $[n]$. Moreover, as pointed out in [2], it is usually convenient to assign an order to the K_n features of a feature allocation of n individuals. A way of achieving this purpose consists in drawing K_n values from a continuous distribution and ordering the K_n features accordingly. The resulting feature allocation is said to be a *randomly ordered feature allocation* and is denoted by $\{\bar{F}_{n,1}, \dots, \bar{F}_{n,K_n}\}$. Its distribution is obtained from that of $\{F_{n,1}, \dots, F_{n,K_n}\}$ by multiplying a combinatorial coefficient as outlined in formula (4) of [2].

Let $M_n = (M_{n,1}, \dots, M_{n,K_n})$ denote the random cardinalities of the randomly ordered feature allocation $\{\bar{F}_{n,1}, \dots, \bar{F}_{n,K_n}\}$ of $[n]$, that is, $M_{n,i} = |\bar{F}_{n,i}|$ for all $i \in [K_n]$. In [2], the authors study the class of randomly ordered exchangeable feature allocations admitting as a sufficient statistics the vector M_n, that is, those satisfying

$$P(\{\bar{F}_{n,1}, \dots, \bar{F}_{n,K_n}\} = \{f_1, \dots, f_k\}) = \pi_n(m_1, \dots, m_k) \tag{1.1}$$

for a symmetric function $\pi_n : \bigcup_{k=0}^{\infty}[n]^k \to [0, 1]$, called an *exchangeable feature probability function* (EFPF), defined over (possibly empty) finite vectors with entries in $[n]$.

When studying random feature allocations generated by a sequence of individuals, it is natural to demand projectivity of their distributions: that is, the distribution of the feature allocation for n individuals should coincide with that for $(n-1)$ individuals, when the last individual is integrated out. Random feature allocations satisfying this condition are said to be *consistent*. When considering randomly ordered exchangeable feature allocations with EFPFs, the consistency condition specializes to the condition

$$\begin{aligned} &\pi_n(m_1, \dots, m_k) \\ &\quad = \sum_{j=0}^{\infty} \binom{k+j}{j} \sum_{\underline{z} \in \{0,1\}^k} \pi_{n+1}(m_1 + z_1, \dots, m_k + z_k, \underbrace{1, \dots, 1}_{j}) \end{aligned} \tag{1.2}$$

for all $n \geq 1$ and where the index $\underline{z}$ ranges over all binary vectors with k elements and z_k denotes the kth entry of this vector. This condition is obtained by combining the last condition on page 4 of [2] with formula (4) of the same paper.

The most remarkable example of a consistent exchangeable feature allocation with an EFPF is the *Indian buffet Process* (IBP), initially introduced in [6], in its one parameter version, and then extended to its two, [4], and three parameters versions, [10]. The EFPF of a 3-parameter (γ, α, θ) IBP has the following form:

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

where $(x)_{m\uparrow}$ denotes the rising factorial, that is, $(x)_{m\uparrow} = \prod_{i=0}^{m-1}(x+i)$ with the proviso $(x)_{0\uparrow} = 1$, and the parameters must satisfy the conditions $\gamma \geq 0$, $0 \leq \alpha \leq 1$, and $\infty > \theta \geq -\alpha$. 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 referred to [7].

The IBP is derived as the limit of a Beta–Bernoulli model in [6]. 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

$$\binom{N}{k}\left(\frac{-\alpha}{(\theta+\alpha)_{n\uparrow}}\right)^k\left(\frac{(\theta+\alpha)_{n\uparrow}}{(\theta)_{n\uparrow}}\right)^N \prod_{i=1}^{k}(1-\alpha)_{m_i-1\uparrow}(\theta+\alpha)_{n-m_i\uparrow}, \tag{1.3}$$

where $\infty < \alpha < 0$ and $\infty > \theta \geq -\alpha$. As a limiting case, for $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q$, we obtain the homogeneous Bernoulli model of parameter q. We refer to this model as the Beta–Bernoulli model with $\alpha = -\infty$. 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 [9] for a complete review of exchangeable random partitions. The most important distribution for random partitions is the Ewens–Pitman formula, which is a generalization of the famous Ewens formula. Starting from this distribution, [5] considers a larger class of random partitions, having an exchangeable partition probability function (see [8] 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 nonnegative weights. Theorem 12 of [5] characterizes all elements of this class of distributions for random partitions satisfying a consistency condition similar to (1.2). The resulting class of distributions is termed *Gibbs-type partitions*.

Motivated by the work [5] 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

$$\pi_n(m_1, \ldots, m_k) = V_{n,k} \prod_{l=1}^{k} W_{m_l} U_{n-m_l} \tag{1.4}$$

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 nonnegative 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 (1.4), up to randomization of their γ and N parameters, respectively. Consistency and exchangeability imply that the two sequences of weights, W and U, can be uniquely rewritten with the same form as in the IBP or Beta–Bernoulli model, for two parameters α and θ satisfying either $-\infty < \alpha \leq 1$ and $-\alpha \leq \theta < \infty$ or $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q \in (0, 1)$. In addition, V must satisfy a recursion with coefficients depending on α and θ and the set of solutions of this recursion forms a convex set. For each fixed α and θ, we describe the extreme points of this convex set. Their form remarkably depends on the value of α. For $0 < \alpha \leq 1$, the set of extreme points coincides with the family of V of a 3-parameter IBP. For $\alpha = 0$, this set of extreme points 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. In summary, we prove the following theorem.

THEOREM 1.1. *The distribution of a consistent exchangeable feature allocation can be represented by an EFPF of form* (1.4) *iff one of the following three cases holds*:

1. *W and U can be uniquely written as $W_m = (1-\alpha)_{m-1\uparrow}$ and $U_m = (\theta+\alpha)_{m\uparrow}$, for constants α, θ satisfying $-\infty < \alpha < 1$ and $-\alpha < \theta < \infty$, and the elements of V satisfy the recursion*
$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \big((\theta+\alpha)_{n\uparrow}\big)^j (\theta+n)^k V_{n+1,k+j}.$$
2. *W and U can be uniquely written $W_m = q^{m-1}$ and $U_m = (1-q)^m$, for some $q \in (0, 1)$, and V satisfies the recursion*
$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} (1-q)^{nj} V_{n+1,k+j},$$
corresponding to the limiting case $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q$.
3. *One of the following two degenerate cases holds.*
   (a) *There is no feature sharing, that is, $M_{n,i} = 1$ almost surely, for all $i \leq K_n$. In this case, $W_m = (1-\alpha)_{m-1\uparrow}$ for $\alpha = 1$, and $\tilde{V}_{n,k} := V_{n,k} U_{n-1}^k$ satisfies $\tilde{V}_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \tilde{V}_{n+1,k+j}$.*
   (b) *There is complete feature sharing, that is, $M_{n,i} = n$ almost surely, for all $i \leq K_n$. In this case, $U_m = (\theta+\alpha)_{m\uparrow}$ for $\theta = -\alpha$, and $\tilde{V}_{n,k} := V_{n,k} W_n^k$ satisfies $\tilde{V}_{n,k} = \tilde{V}_{n+1,k}$.*

Moreover, for fixed (α, θ), the set of solutions of these recursions is:

1. *for $0 < \alpha \leq 1$, 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 Section 2.1, we begin by characterizing U and W, and identifying the recursion defining V. In Section 2.2, we describe how to derive the extreme solutions of these recursions. Finally, in Section 2.3, we study the three cases $0 < \alpha \le 1$, $\alpha = 0$, and $\alpha < 0$ and the two degenerate cases.

2. Proof of Theorem 1.1. The problem is to describe all distributions for exchangeable feature allocations with EFPF (1.4) subject to the consistency constraint (1.2), which becomes

$$V_{n,k} \prod_{i=1}^{k} W_{m_i} U_{n-m_i} = \sum_{j=0}^{\infty} \binom{k+j}{j} U_n^j W_1^j \sum_{\underline{z} \in \{0,1\}^k} V_{n+1,k+j} \prod_{i=1}^{k} W_{m_i+z_i} U_{n+1-m_i-z_i} \tag{2.1}$$

for all $n \in \mathbb{N}$, $k \in \mathbb{N}_0$, and $m_i \le n$, for $i \le k$. We start by noting that the representation (1.4) is not unique. Specifically, we can tilt the weights in the following ways, for $\kappa > 0$, and obtain the same EFPF:

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

By imposing $W_1 = 1$, we remove the first ambiguity, and with $U_0 = 1$, we remove the second one. These conditions also exclude the third ambiguity, but do not exclude the last geometric tilting, which we address in Proposition 2.1.

2.1. *Characterization of W and U.* The following proposition shows that W and U must have the same form as in the IBP and Beta–Bernoulli model and V is constrained to satisfy a particular recursion. In the proof of Proposition 2.1, $(x)_{n\uparrow\tau}$ denotes the generalized rising factorial, that is, $(x)_{n\uparrow\tau} = \prod_{i=0}^{n-1}(x + i\tau)$.

PROPOSITION 2.1. *Let $W_j, U_j > 0$ for all $j > 0$. Then, the weights V, W, and U, with the normalizations $W_1 = U_0 = 1$, define a consistent exchangeable feature allocation of form* (1.4) *iff $\sum_{j \ge 0} V_{1,j} = 1$ and one of the following two cases holds*:

1. *For some constants α, θ satisfying $-\infty < \alpha < 1$ and $-\alpha < \theta < \infty$, W and U can be uniquely expressed as $W_j = (1-\alpha)_{j-1\uparrow}$ for all $j \ge 1$ and $U_j = (\theta+\alpha)_{j\uparrow}$ for all $j \ge 0$, and V satisfies, for all $(n,k) \in \mathbb{N} \times \mathbb{N}_0$,*

$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \big((\theta+\alpha)_{n\uparrow}\big)^j (\theta+n)^k V_{n+1,k+j}. \tag{2.2}$$

2. *For some constant $q \in (0,1)$, W and U can be uniquely expressed as $W_j = q^{j-1}$ for all $j \geq 1$ and $U_j = (1-q)^j$ for all $j \geq 0$, and V satisfies, for all $(n,k) \in \mathbb{N} \times \mathbb{N}_0$,*

$$V_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} (1-q)^{nj} V_{n+1,k+j}. \tag{2.3}$$

PROOF. The consistency condition (2.1) for $k=1$ gives

$$V_{n,1} W_{m_1} U_{n-m_1} = \sum_{j=0}^{\infty} (j+1) V_{n+1,j+1} U_n^j (W_{m_1+1} U_{n-m_1} + W_{m_1} U_{n+1-m_1}). \tag{2.4}$$

The consistent exchangeable feature allocation with no features with probability one can be represented as in (1.4), with $V_{n,0} = 1$ for all $n \in \mathbb{N}$ and $V_{n,k} = 0$ for $k \geq 1$. Except this case, $\sum_{j=0}^{\infty} (j+1) V_{n+1,j+1} U_n^j$ cannot be equal to zero. Therefore, because of $W_j, U_j > 0$ for all $j > 0$, condition (2.4) implies that, for all $n \in \mathbb{N}$ and for all $m_1 \leq n$,

$$\frac{W_{m_1+1}}{W_{m_1}} + \frac{U_{n+1-m_1}}{U_{n-m_1}} = \frac{V_{n,1}}{\sum_{j=0}^{\infty} (j+1) V_{n+1,j+1} U_n^j}. \tag{2.5}$$

Since the right-hand side of (2.5) 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

$$\frac{W_{n+1}}{W_n} - \frac{W_n}{W_{n-1}} = \frac{U_2}{U_1} - U_1 = \tau. \tag{2.6}$$

This last condition implies, for all $n > 1$,

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

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

$$\frac{U_n}{U_{n-1}} - \frac{U_{n-1}}{U_{n-2}} = \frac{W_3}{W_2} - W_2 = \tau. \tag{2.7}$$

As before, this formula implies $U_n = \prod_{i=1}^{n} (U_1 + i\tau)$.

If $\tau \neq 0$, we can introduce the parametrization $\alpha = 1 - \frac{W_2}{\tau}$ and $\theta = \frac{U_1+W_2}{\tau} - 1$. Then $W_n = (W_2)_{n-1\uparrow\tau} = \tau^{n-1}(1-\alpha)_{n-1\uparrow}$ and $U_n = (U_1)_{n\uparrow\tau} = \tau^n(\theta+\alpha)_{n\uparrow}$. We can fix geometric tilting by the normalization $\tau = 1$, in which case α, θ satisfy $-\infty < \alpha < 1$ and $-\alpha < \theta < \infty$.

If $\tau = 0$, then $W_n = (W_2)^{n-1}$ and $U_n = (U_1)^n$. Let $W_2 = q > 0$. Then $W_n = q^{n-1}$. In this case, we fix geometric tilting by imposing $W_2 = 1 - U_1$, in which case it must also hold that $q < 1$ and $U_n = (1-q)^n$.

The recursions (2.2) and (2.3) follow from rewriting (2.1) as

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

noticing that

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

and substituting the forms for W and U just obtained. Finally, $\sum_{j=0}^{\infty} V_{1,j} = 1$ follows 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 (1.4) and V, W, and U as in the statement of the proposition satisfies the consistency condition (2.1). $\square$

In Proposition 2.1, we assumed $W, U > 0$. In Propositions A.1 and A.2, we show that only two feature allocations having $W_j = 0$ or $U_j = 0$ for some j are the feature allocation with $W_j = \mathbb{1}(j=1)$ and the feature allocation with $U_j = \mathbb{1}(j=0)$.

The first solution, $W_j = \mathbb{1}(j=1)$, corresponds to a feature allocation where no features are shared between individuals, that is, with probability one, $M_{n,j} = 1$ for all $j \leq K_n$. This sequence of weights can be represented as $W_m = (1-\alpha)_{m-1\uparrow}$ for $\alpha = 1$. Moreover, the only nonnegative EFPF is $\pi_n(m_1, \dots, m_k) = V_{n,k} U_{n-1}^k$, with $m_j = 1$ for all $j \leq k$, which is a function of n and k only. Therefore, we can define $\tilde{V}_{n,k} = V_{n,k} U_{n-1}^k$, which must satisfy the recursion $\tilde{V}_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \tilde{V}_{n+1,k+j}$ according to (2.1).

The second solution, $U_j = \mathbb{1}(j=0)$, is the feature allocation in which every individual possesses the exact same features, that is, with probability one, $M_{n,j} = n$ for all $j \leq K_n$ (and K_n is independent of n). This sequence U can be represented as $U_m = (\theta+\alpha)_{m\uparrow}$ with $\alpha = -\theta$ and any value of θ. The only nonnegative EFPF is $\pi_n(m_1, \dots, m_k) = V_{n,k} W_n^k$, with $m_j = n$ for all $j \leq k$, which is a function of n and k only. Again, we can introduce a new parametrization, $\tilde{V}_{n,k} = V_{n,k} W_n^k$, which must satisfy $\tilde{V}_{n+1,k} = \tilde{V}_{n,k}$ according to (2.1).

We will exclude these two degenerate cases in our analysis until Sections 2.3.4 and 2.3.5.

2.2. *General tools to derive the extreme V.* Fix constants α, θ satisfying $-\infty < \alpha < 1$ and $-\alpha < \theta < \infty$. (We will consider the limiting case $\alpha \to -\infty$ at the end of the section.) Let $\mathcal{V}_{\alpha,\theta}$ be the set of those elements $V \in \mathbb{R}_+^{\mathbb{N}\times\mathbb{N}_0}$ satisfying (2.2). 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,

$$V_{n,k}^{\mu} = \int_{\mathcal{V}_{\alpha,\theta}} V_{n,k}\mu(\mathrm{d}V). \tag{2.8}$$

It is easy to check that $\mathcal{V}_{\alpha,\theta}$ is a convex set, that is, 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, [3] describes a general theory, which can be applied to show the space is a simplex, and, then, determine its extreme points. Similar results have been studied and rediscovered on several occasions (see references in [5]). In order to apply the results of [3] to our problem, we will follow the same strategy used by [5]: rather than studying $\mathcal{V}_{\alpha,\theta}$ directly, we consider a space that is isomorphic to $\mathcal{V}_{\alpha,\theta}$ and easier to study, and then we find its extreme points by applying the results in [3].

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 nth coordinate projection on the product space, the Markov law associated to V has an initial distribution given by

$$P_V(K_1 = j) = V_{1,j}, \tag{2.9}$$

and transition probabilities

$$P_V(K_{n+1} = j + k | K_n = k) = \binom{k+j}{j} \big((\alpha+\theta)_{n\uparrow}\big)^j (\theta+n)^k \frac{V_{n+1,k+j}}{V_{n,k}}, \tag{2.10}$$

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 (1.4) 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, that is, K_n is the cardinality of $\bigcup_{1\leq i\leq n} X_i$.

As we will see from Proposition 2.2, for every $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma(K_n, K_{n+1}, \dots)$ is a sufficient σ-algebra for $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$. Hence, for each $n \in \mathbb{N}$, there exists a common 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 $A \in \mathcal{C}(\mathbb{N}_0^\infty)$,

$$Q_n(\omega, A) = P_V\big((K_m)_{m\in\mathbb{N}} \in A | \mathcal{F}_n\big)(\omega) \tag{2.11}$$

for P_V-almost all $\omega \in \mathbb{N}_0^\infty$. When $A \in \sigma(K_1, \dots, K_n)$, we will take advantage of the Markov property of $(K_m)_{m\in\mathbb{N}}$ and write

$$Q_n(\omega, A) = P\big((K_m)_{m\in\mathbb{N}} \in A|\mathcal{F}_n\big)(\omega) = P\big((K_m)_{m\in\mathbb{N}} \in A|K_n\big)(\omega), \tag{2.12}$$

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.

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

$$\Pi_n f(\omega) = \int f(\omega')Q_n(\omega, \mathrm{d}\omega') \tag{2.13}$$

for all f bounded and $\mathcal{C}(\mathbb{N}_0^\infty)$-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 [3], 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 [3]) that, for all P_V that are extreme,

$$P_V\left(\left\{\omega \in \mathbb{N}_0^\infty : \forall f \in \mathcal{C}(\mathbb{N}_0^\infty), \lim_{n\to\infty} \Pi_n f(\omega) = \int f\,\mathrm{d}P_V\right\}\right) = 1. \tag{2.14}$$

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 points 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 (2.14), that is, they assign probability one to the set of regular paths inducing them.

In our context, to identify the maximal boundary, it is enough to check (2.14) 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 $(K_n)_{n\in\mathbb{N}}$. We denote by $(x)_{n\downarrow}$ the falling factorial, that is, $(x)_{n\downarrow} = \prod_{i=0}^{n-1}(x - i)$ with the proviso $(x)_{0\downarrow} = 1$.

PROPOSITION 2.2. *For $-\infty < \alpha < 1$ and $-\alpha < \theta < \infty$, the cotransition probabilities are*

$$P(K_n = k|K_m = l) = \frac{d_{n,k}^{m,l}}{d^{m,l}} d^{n,k} \tag{2.15}$$

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

$$P_V(K_n = k) = V_{n,k} d^{n,k}, \tag{2.16}$$

where

$$d^{m,l} = \left(\sum_{j=1}^{m} (\theta+\alpha)_{m-j\uparrow}(\theta+1+m-j)_{j-1\uparrow}\right)^{l} \tag{2.17}$$

and

$$d_{n,k}^{m,l} = \binom{l}{k}\left((\theta+n)_{m-n\uparrow}\right)^{k}\left(\sum_{j=1}^{m-n} (\alpha+\theta)_{m-j\uparrow}(\theta+m-1)_{j-1\downarrow}\right)^{l-k}. \tag{2.18}$$

PROOF. See Appendix A.3. □

Note that the cotransition probabilities are independent of V.

Along the same lines we can deal with the limiting case $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q \in (0,1)$. In particular, the transition probability (2.10) is replaced by

$$P_V(K_{n+1} = j+k \mid K_n = k) = \binom{k+j}{j}(1-q)^{nj}\frac{V_{n+1,k+j}}{V_{n,k}}, \tag{2.19}$$

and Proposition 2.2 still holds with

$$d^{m,l} = \left(\sum_{j=1}^{m} (1-q)^{m-j}\right)^{l} \tag{2.20}$$

and

$$d_{n,k}^{m,l} = \binom{l}{k}\left(\sum_{j=1}^{m-n} (1-q)^{m-j}\right)^{l-k}. \tag{2.21}$$

We omit the proof of these formulas because it follows the same steps of the proof of Proposition 2.2 with (2.19) in place of (2.10).

2.3. *Characterization of* $V_{n,k}$. In this section, we study the three cases $0 < \alpha < 1$, $\alpha = 0$, and $\alpha < 0$, and then study the degenerate cases $\alpha = 1$ and $\alpha = -\theta$ 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

$$\lim_{m\to\infty} P(K_n = k \mid K_m = \omega_m) = \lim_{m\to\infty} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} d^{n,k} = \bar{V}_{n,k} d^{n,k} \tag{2.22}$$

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 $-\alpha < \theta < \infty$, let $V^{3\mathrm{IBP},\alpha,\theta}(\gamma)$ be the V of the 3-parameter IBP, defined as

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

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

PROPOSITION 2.3. *Let* $0 < \alpha < 1$ *and* $-\alpha < \theta < \infty$.

(a) *The elements of the set* $\mathcal{P}_{V^{3\mathrm{IBP},\alpha,\theta}}$ *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* $\mathcal{P}_{V^{3\mathrm{IBP},\alpha,\theta}}$ *also belong to the minimal boundary of* $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, *that is, they are extreme points of* $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$.

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

PROOF. (a) In Appendix A.4, we check that

$$\lim_{\substack{m\to\infty \\ \frac{\omega_m}{m^\alpha}\to c}} \frac{d^{m,\omega_m}_{n,k}}{d^{m,\omega_m}} = V^{3\mathrm{IBP},\alpha,\theta}_{n,k}\biggl(\frac{c\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}\biggr). \tag{2.23}$$

(b) From Theorem 4 of [1], it follows that

$$P_{V^{3\mathrm{IBP},\alpha,\theta}(\frac{c\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)})}\biggl(\frac{K_m}{m^\alpha} \to c\biggr) = 1.$$

(c) In Appendix A.4, we show that the elements of $\mathcal{P}_{V^{3\mathrm{IBP},\alpha,\theta}}$ are the only ones belonging to the maximal boundary, that is, there are no other regular paths except those of part (a). $\square$

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, nonnegative, and finite, the V of the 2-parameter IBP are of the form

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

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

PROPOSITION 2.4. *Let $\alpha = 0$ and $0 < \theta < \infty$.*

(a) *The elements of the set $\mathcal{P}_{V^{2\mathrm{IBP},\theta}}$ belong to 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 $\mathcal{P}_{V^{2\mathrm{IBP},\theta}}$ also belong to the minimal boundary of $\mathcal{P}_{\mathcal{V}_{0,\theta}}$, that is, they are extreme points of $\mathcal{P}_{\mathcal{V}_{0,\theta}}$.*
(c) *The elements of $\mathcal{P}_{V^{2\mathrm{IBP},\theta}}$ are the only extreme points, that is, $\mathcal{P}_{V^{3\mathrm{IBP},\theta}}$ coincides with the maximal and the minimal boundary.*

PROOF. (a) In Appendix A.5, we check that

$$\lim_{\substack{m\to\infty \\ \frac{\omega_m}{\log(m)}\to\gamma}} \frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} = V_{n,k}^{2\mathrm{IBP},\theta}(\gamma). \tag{2.24}$$

(b) This also follows from Theorem 4 of [1], which establish that

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

(c) In Appendix A.5, we check that there are no other regular paths but those of part (a). $\square$

2.3.3. *Case $\alpha < 0$.* From formula (1.3), we see that, if $-\infty < \alpha < 0$ and $-\alpha < \theta < \infty$, the Beta–Bernoulli is of form (1.4), with V of the form

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

and, in the limiting case $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q \in (0, 1)$, from (A.1) we have

$$V_{n,k}^{\mathrm{BB},\alpha,\theta}(N) = \binom{N}{k}(1-q)^{n(N-k)}q^k$$

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

PROPOSITION 2.5. *Let $-\infty < \alpha < 0$ and $-\alpha < \theta < \infty$ or $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q \in (0, 1)$.*

(a) *The elements of the set $\mathcal{P}_{V^{\mathrm{BB},\alpha,\theta}}$ belong to 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 $\mathcal{P}_{V^{\mathrm{BB},\alpha,\theta}}$ also belong to the minimal boundary of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$, that is, they are extreme points of $\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}$.*

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

PROOF. (a) In Appendix A.6, we check that

$$\lim_{\substack{m\to\infty \\ \omega_m \to N}} \frac{d^{m,\omega_m}_{n,k}}{d^{m,\omega_m}} = V^{\mathrm{BB},\alpha,\theta}_{n,k}(N). \tag{2.25}$$

(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 (or being fixed to q for $\alpha \to -\infty$). 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 no other regular paths but those of part (a). □

2.3.4. *Degenerate case*: $\alpha = 1$. Proceeding along the same lines of Section 2.2, we can show that for the recursion $\tilde{V}_{n,k} = \sum_{j=0}^{\infty} \binom{k+j}{j} \tilde{V}_{n+1,k+j}$, we have $d^{m,l}_{n,k} = \binom{l}{k}(m-n)^{l-k}$ and $d^{m,l} = m^l$. As $m \to \infty$ and $\frac{\omega_m}{m} \to \gamma$, the ratio $d^{m,\omega_m}_{n,k}/d^{m,\omega_m}$ converges to $\tilde{V}_{n,k} = \frac{1}{k!}\gamma^k \exp(-n\gamma)$, which is the EFPF of the IBP with $\alpha = 1$ and all cardinalities $m_l = 1$. Moreover, it is easy to show that these are the only regular paths. Therefore, all feature allocations with no sharing are mixtures over γ of the IBP with $\alpha = 1$.

2.3.5. *Degenerate case*: $\alpha = -\theta$. In this case $\tilde{V}$ satisfies $\tilde{V}_{n+1,k} = \tilde{V}_{n,k}$ and $\sum_j \tilde{V}_{1,j} = 1$. Therefore, the first row of $\tilde{V}$ is a generic mass function over $\mathbb{N}_0$ and all other rows are equal to the first. This feature allocation can be written as a mixture over the parameter N of Beta–Bernoulli models with $\alpha = -\theta$. Indeed, in a Beta–Bernoulli model $(N, \alpha, -\alpha)$ the underlying Beta distribution is degenerate and it is a point mass at 1. Therefore, all individuals display all N features. Mixing over N with mixing measure $(\tilde{V}_{1,j})_{j \in \mathbb{N}_0}$, we recover all possible feature allocations with complete sharing.

3. Discussion.

In this work, we have considered the class of consistent exchangeable feature allocations with EFPF of the form (1.4). 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 (1.4), 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.

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 parameter Q_j with distribution $\mathrm{Beta}(\eta_1, \eta_2)$ and independent of the parameters of other features. Given Q_j, each individual X_i possesses feature j with probability Q_j, independently of other individuals and other features. 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

$$\begin{aligned} Z_{i,j} \mid Q_j &\overset{\text{ind}}{\sim} \mathrm{Bernoulli}(Q_j), \qquad i = 1, \dots, n; \ j = 1, \dots, N; \\ Q_j \mid \eta_1, \eta_2 &\overset{\text{i.i.d.}}{\sim} \mathrm{Beta}(\eta_1, \eta_2), \qquad j = 1, \dots, N. \end{aligned}$$

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

$$P(Z = z \mid Q) = \prod_{j=1}^{N} \prod_{i=1}^{n} \mathrm{Bernoulli}(z_{i,j} \mid Q_j).$$

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

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

where $\overline{m}_i = \sum_{j=1}^{n} z_{i,j}$. If $(\overline{m}_1, \dots, \overline{m}_k)$ has k nonzero entries, denoted $(m_1, \dots, m_N)$, this probability becomes

$$\begin{aligned} &\left(\frac{\Gamma(\eta_1 + \eta_2)}{\Gamma(\eta_1)\Gamma(\eta_2)\Gamma(n + \eta_1 + \eta_2)} \right)^N \big(\Gamma(\eta_1)\Gamma(n + \eta_2)\big)^{N-k} \\ &\quad \times \prod_{i=1}^{k} \Gamma(m_i + \eta_1)\Gamma(n - m_i + \eta_2). \end{aligned}$$

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 the EFPF

$$\begin{aligned} &\binom{N}{k} \left(\frac{\Gamma(\eta_1 + \eta_2)}{\Gamma(\eta_1)\Gamma(\eta_2)\Gamma(n + \eta_1 + \eta_2)} \right)^N \big(\Gamma(\eta_1)\Gamma(n + \eta_2)\big)^{N-k} \\ &\quad \times \prod_{i=1}^{k} \Gamma(m_i + \eta_1)\Gamma(n - m_i + \eta_2), \end{aligned}$$

which can be rewritten as in formula (1.3), by using rising factorials and by changing the parametrization to $\alpha = -\eta_1$ and $\theta = \eta_2 + \eta_1$, with $\alpha < 0$ and $\theta \ge -\alpha$.

The homogeneous Bernoulli model. In the homogeneous Bernoulli model, every individual displays any of the N possible features with a nonrandom probability q. This model is the limiting case of the Beta–Bernoulli model when $\eta_1/(\eta_1+\eta_2)\to q$ and $\eta_1,\eta_2\to\infty$. The homogeneous Bernoulli model can be written as follows:

$$Z_{i,j}\overset{\text{i.i.d.}}{\sim}\text{Bernoulli}(q),\qquad i=1,\dots,n;\ j=1,\dots,N.$$

The probability mass function of a $n\times N$ matrix Z generated by this model is

$$P(Z=z)=\prod_{j=1}^{N}\prod_{i=1}^{n}q^{z_{j,i}}(1-q)^{1-z_{j,i}}=(1-q)^{n(N-k)}\prod_{j=1}^{k}q^{m_j}(1-q)^{n-m_j},$$

where as before m_i are the column sums of the k nonzero columns in z. Taking into account the $\binom{N}{k}$ uniform orderings of the $N-k$ zero columns, the EFPF of the homogeneous Bernoulli model is

$$\binom{N}{k}(1-q)^{n(N-k)}q^k\prod_{j=1}^{k}q^{m_j-1}(1-q)^{n-m_j}. \tag{A.1}$$

A.2. Some facts.

PROPOSITION A.1. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random finite sets generating an exchangeable feature allocation with EFPF of form* (1.4). *If $W_2>0$, then $W_j>0$ for all $j>2$.*

PROOF. The random sets $(X_i)_{i\in\mathbb{N}}$, defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$, define an exchangeable feature allocation with EFPF of form (1.4) with $W_2>0$ iff, for all $n\in\mathbb{N}$, the sets

$$A_n=\left\{\omega\in\Omega:\exists x_\omega\in\bigcup_{i=1}^{n}X_i(\omega)\text{ s.t. }\sum_{i=1}^{n}\mathbb{1}\big(x_\omega\in X_i(\omega)\big)\geq 2\right\}$$

have positive probability, that is, $\mathbb{P}(A_n)>0$. We will show that for all $\omega\in A_n$ (except possibly in a null set) we have

$$\lim_{m\to\infty}\sum_{j=1}^{m}\mathbb{1}\big(x_\omega\in X_{n+j}(\omega)\big)=\infty, \tag{A.2}$$

which in turns implies that $W_j>0$ for all $j>2$.

Suppose (A.2) does not hold, that means that $\exists B_n\subseteq A_n$ with $\mathbb{P}(B_n)>0$ s.t. for all $\omega\in B_n$

$$\lim_{m\to\infty}\sum_{j=1}^{m}\mathbb{1}\big(x_\omega\in X_{n+j}(\omega)\big)<\infty.$$

Since $\sum_{j=1}^{m} \mathbb{1}(x_\omega \in X_{n+j}(\omega))$ is not decreasing, for all $\omega \in B_n$ $\exists M_\omega > 1$ s.t.

$$\lim_{m\to\infty} \sum_{j=1}^{m} \mathbb{1}\big(x_\omega \in X_{n+j}(\omega)\big) = M_\omega < \infty.$$

Now, by dominated convergence

$$\begin{aligned}\lim_{m\to\infty} \sum_{j=1}^{m} \int_{B_n} \mathbb{1}\big(x_\omega \in X_{n+j}(\omega)\big)\,\mathrm{d}\mathbb{P} &= \int_{B_n} \lim_{m\to\infty} \sum_{j=1}^{m} \mathbb{1}\big(x_\omega \in X_{n+j}(\omega)\big)\,\mathrm{d}\mathbb{P} \\ &= \mathbb{E}\big(M_\omega \mathbb{1}(B_n)\big) < \infty.\end{aligned}$$

By exchangeability of the X_i's,

$$\begin{aligned}\lim_{m\to\infty} \sum_{j=1}^{m} \mathbb{E}\big(\mathbb{1}\big(x_\omega \in X_{n+j}(\omega)\big)\mathbb{1}(B_n)\big) &= \lim_{m\to\infty} m\mathbb{E}\big(\mathbb{1}\big(x_\omega \in X_{n+1}(\omega)\big)\mathbb{1}(B_n)\big) \\ &= \mathbb{E}\big(M_\omega \mathbb{1}(B_n)\big) < \infty.\end{aligned} \tag{A.3}$$

Therefore, it must be $\mathbb{E}(\mathbb{1}(x_\omega \in X_{n+1}(\omega))\mathbb{1}(B_n)) = 0$, but on $B_n \subseteq A_n$, from exchangeability of the X_i's, $\mathbb{P}(x_\omega \in X_{n+1}(\omega)) > 0$. As a consequence, in order for (A.3) to be true it must be $\mathbb{P}(B_n) = 0$, hence a contradiction. □

PROPOSITION A.2. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random finite sets generating an exchangeable feature allocation with EFPF of form* (1.4). *If $U_1 > 0$, then $U_j > 0$ for all $j > 1$.*

PROOF. Proceed along similar lines of the proof of Proposition A.1, consider the sets

$$A_n = \left\{\omega \in \Omega : \exists x_\omega \in \bigcup_{i=1}^{n} X_i(\omega) \text{ s.t. } \sum_{i=1}^{n} \mathbb{1}\big(x_\omega \notin X_i(\omega)\big) \geq 1\right\},$$

and check that for almost all $\omega \in A_n$,

$$\lim_{m\to\infty} \sum_{j=1}^{m} \mathbb{1}\big(x_\omega \notin X_{n+j}(\omega)\big) = \infty.$$

□

PROPOSITION A.3. *$\mathcal{V}_{\alpha,\theta}$ is a convex set.*

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

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

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

$$= \sum_{j=0}^{\infty} \binom{k+j}{j} ((\alpha+\theta)_{n\uparrow})^j (\theta+n)^k V^{\mu}_{n+1,k+j},$$

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.4. *$T(V) = P_V$ is an isomorphism between convex sets.*

PROOF. According to [3], page 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 $\{P_V \in \mathcal{P}_{\mathcal{V}_{\alpha,\theta}} : P_V(K_n = k) \le x\}$ for $(n,k) \in \mathbb{N} \times \mathbb{N}_0$ and $x \in [0,1]$. The inverse image under T of this set is $\{V \in \mathcal{V}_{\alpha,\theta} : V_{n,k} d_{n,k} \le x\}$ (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

$$P^{\mu'}(A) = \int_{\mathcal{P}_{\mathcal{V}_{\alpha,\theta}}} P(A)\mu'(\mathrm{d}P) \tag{A.4}$$

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

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

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

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

PROPOSITION A.5. $(\mathcal{F}_n, \Pi_n)_{n \in \mathbb{N}}$ *forms a specification in* $(\mathbb{N}_0^{\infty}, \mathcal{C}(\mathbb{N}_0^{\infty}))$.

PROOF. According to [3], 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 2.2, formula (2.13), (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 $K_1^{-1}\{k_1\} \cap K_2^{-1}\{k_2\} \cap \cdots \cap K_m^{-1}\{k_m\}$ for $m > n+1$ and $k_i \in \mathbb{N}_0$ for all $i \leq m$,

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

$\square$

A.3. Proof of Proposition 2.2. First, note that for $m > n$ and $l \geq k$ $P_V(K_m = l | K_n = k) = V_{m,l} d_{n,k}^{m,l}$, for a function $d_{n,k}^{m,l}$ independent of V. Indeed, from (2.10), the probability of a path $(k_{n+1}, k_{n+2}, \dots, k_{m-1}, l)$ 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 $V_{m,l} d_{n,k}^{m,l}$. 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

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

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

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

which gives the following the recursion:

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

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

$$\begin{aligned} d^{m,l} = & \sum_{j=0}^{l} \binom{l}{l-j} \big((\alpha+\theta)_{m-1\uparrow}\big)^{l-j} (\theta + m - 1)^j \\ & \times \sum_{i=0}^{j} \binom{j}{j-i} \big((\alpha+\theta)_{m-2\uparrow}\big)^{j-i} (\theta + m - 2)^i d^{m-2,i}. \end{aligned}$$

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

$$\begin{aligned} d_{m-2,k}^{m,l} = & \binom{l}{l-k} \big((\alpha+\theta)_{m-1\uparrow}\big)^{l-k} (\theta + m - 1)^k (\theta + m - 2)^k \\ & + \binom{l}{l-k-1} \big((\alpha+\theta)_{m-1\uparrow}\big)^{l-k-1} (\theta + m - 1)^{k+1} \\ & \times \big((\alpha+\theta)_{m-2\uparrow}\big) (\theta + m - 2)^k \\ & + \binom{l}{l-k-2} \big((\alpha+\theta)_{m-1\uparrow}\big)^{l-k-2} (\theta + m - 1)^{k+2} \\ & \times \binom{k+2}{2} \big((\alpha+\theta)_{m-2\uparrow}\big)^2 (\theta + m - 2)^k \end{aligned}$$

$$\begin{aligned}
&\vdots\\
&+\big((\alpha+\theta)_{m-1\uparrow}\big)(\theta+m-1)^{l-1}\\
&\times\binom{l-1}{l-1-k}\big((\alpha+\theta)_{m-2\uparrow}\big)^{l-1-k}(\theta+m-2)^k\\
&+(\theta+m-1)^l\binom{l}{l-k}\big((\alpha+\theta)_{m-2\uparrow}\big)^{l-k}(\theta+m-2)^k\\
={}&\big((\theta+m-1)(\theta+m-2)\big)^k\\
&\times\big((\alpha+\theta)_{m-1\uparrow}+(\theta+m-1)(\alpha+\theta)_{m-2\uparrow}\big)^{l-k}\binom{l}{l-k}.
\end{aligned}$$

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

$$\begin{aligned}
d^{m,l}={}&\sum_{j=0}^{l}\binom{l}{l-j}\big((\theta+m-1)(\theta+m-2)\big)^j\\
&\times\big((\alpha+\theta)_{m-1\uparrow}+(\theta+m-1)(\alpha+\theta)_{m-2\uparrow}\big)^{l-j}d^{m-2,j}.
\end{aligned}$$

In the same manner, we find

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

Finally, we obtain

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

In addition,

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

$\square$

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 $m \to \infty$.

LEMMA A.6. *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*

$$\left(\frac{f(m)-p}{f(m)-q}\right)^{h(m)} \approx \left(\frac{g(m)-p}{g(m)-q}\right)^{cg(m)} \to e^{c(q-p)}. \tag{A.5}$$

PROOF. We prove only the first equivalence, because the limiting exponential form is well known. Taking logarithms, we have

$$\log\left[\left(\frac{f(m)-p}{f(m)-q}\right)^{h(m)}\left(\frac{g(m)-q}{g(m)-p}\right)^{cg(m)}\right] \tag{A.6}$$

$$= cg(m)\log\frac{(f(m)-p)(g(m)-q)}{(f(m)-q)(g(m)-p)} \tag{A.7}$$

$$+ \big(h(m)-cg(m)\big)\log\frac{f(m)-p}{f(m)-q}. \tag{A.8}$$

The arguments to the logarithms can we written as

$$\frac{(f(m)-p)(g(m)-q)}{(f(m)-q)(g(m)-p)} = 1 + \frac{(f(m)-g(m))(p-q)}{(f(m)-q)(g(m)-p)} \tag{A.9}$$

and

$$\frac{f(m)-p}{f(m)-q} = 1 + \frac{q-p}{f(m)-q}. \tag{A.10}$$

Using the fact that $z(m) \to 0$ implies $\log(1+z(m)) \approx z(m)$, and that both terms (A.9) and (A.10) converge to one, it follows that

$$\text{(A.6)} \approx cg(m)\frac{(f(m)-g(m))(p-q)}{(f(m)-q)(g(m)-p)} + \big(h(m)-cg(m)\big)\frac{q-p}{f(m)-q}. \tag{A.11}$$

It is straightforward to show that the right-hand side converges to 0. $\square$

LEMMA A.7. *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

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

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

We now proceed to prove each part of Proposition 2.3:

(a) We must check the limit (2.23). From Proposition 2.2,

$$
\begin{aligned}
\frac{d^{m,\omega_m}_{n,k}}{d^{m,\omega_m}} &= \binom{\omega_m}{k}\frac{[(\theta+m-1)_{m-n\downarrow}]^k[\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}]^{\omega_m-k}}{[(\theta+m-1)_{m-1\downarrow}+\sum_{i=1}^{m-1}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}]^{\omega_m}} \\
&= \binom{\omega_m}{k}\left[\frac{(\theta+m-1)_{m-n\downarrow}}{\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}\right]^k \\
&\quad\times\left[\frac{\sum_{i=1}^{m-n}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}{(\theta+m-1)_{m-1\downarrow}+\sum_{i=1}^{m-1}(\alpha+\theta)_{m-i\uparrow}(\theta+m-1)_{i-1\downarrow}}\right]^{\omega_m} \qquad \text{(A.12)} \\
&= \binom{\omega_m}{k}\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta+n)}}{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}\right]^k \\
&\quad\times\left[\frac{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta+1)}+\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+1)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+1)}}\right]^{\omega_m},
\end{aligned}
$$

where the third equality follows from the identity

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

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 (A.12) can be simplified to yield a limiting form:

$$
\begin{aligned}
&\binom{\omega_m}{k}\left[\frac{\frac{\Gamma(\theta+m)}{\Gamma(\theta+n)}}{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}\right]^k \\
&\quad\approx\binom{\omega_m}{k}\left[\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{m^\alpha}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \\
&\quad\approx\frac{1}{k!}\left[\omega_m m^{-\alpha}\cdot\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{1}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{m^\alpha\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \\
&\quad\approx\frac{1}{k!}\left[c\cdot\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{1}{\alpha\Gamma(\alpha+\theta)}-\frac{\Gamma(\alpha+\theta+n)}{m^\alpha\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}}\right]^k \\
&\quad\to\frac{1}{k!}\left(c\frac{\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+n)}\right)^k.
\end{aligned}
\tag{A.13}
$$

Similarly, the second line of (A.12) can be simplified by Lemma A.6, 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

$$
\begin{aligned}
&\left[\frac{\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+n)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+n)}}{\frac{\Gamma(\theta+m)}{\Gamma(\theta+1)}+\frac{\Gamma(\alpha+\theta+m)}{\alpha\cdot\Gamma(\alpha+\theta)}-\frac{\Gamma(\theta+m)\cdot\Gamma(\alpha+\theta+1)}{\alpha\cdot\Gamma(\alpha+\theta)\cdot\Gamma(\theta+1)}}\right]^{\omega_m} \\
&\quad\approx\left[\frac{m^\alpha-\frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}}{m^\alpha-\frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}}\right]^{cm^\alpha} \\
&\quad\to\exp\left\{c\left(\frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}-\frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}\right)\right\}.
\end{aligned}
\tag{A.14}
$$

Substituting back into (A.12), 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; that is, 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 (A.12), (A.13) and (A.14); 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.7 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. $\square$

A.5. Proof of Proposition 2.4. (a) We must check the limit (2.24). From Proposition 2.2,

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

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

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

Therefore, by Lemma A.6 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

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

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.7 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. $\square$

A.6. Proof of Proposition 2.5. (a) We must check the limit (2.25). For $\alpha \in (-\infty, 0)$, starting with Proposition 2.2 and following similar steps as for the case $0 < \alpha < 1$, we obtain the approximation

$$\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} \approx \binom{\omega_m}{k} \left(\frac{\frac{1}{\Gamma(\theta+n)}}{\frac{m^\alpha}{\alpha\Gamma(\alpha+\theta)} - \frac{\Gamma(\alpha+\theta+n)}{\alpha\Gamma(\alpha+\theta)\Gamma(\theta+n)}} \right)^k \left(\frac{m^\alpha - \frac{\Gamma(\alpha+\theta+n)}{\Gamma(\theta+n)}}{m^\alpha - \frac{\theta\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}} \right)^{\omega_m}, \tag{A.15}$$

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

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

For $\alpha \to -\infty$ and $-\frac{\alpha}{\theta} \to q \in (0, 1)$,

$$\begin{aligned}
\frac{d_{n,k}^{m,\omega_m}}{d^{m,\omega_m}} &= \frac{\binom{\omega_m}{k}(\sum_{j=1}^{m-n}(1-q)^{m-j})^{\omega_m-k}}{(\sum_{j=1}^{m}(1-q)^{m-j})^{\omega_m}} \\
&= \frac{\binom{\omega_m}{k}(\frac{(1-q)^n-(1-q)^m}{q})^{\omega_m-k}}{(\frac{1-(1-q)^m}{q})^{\omega_m}} \\
&\to \frac{\binom{N}{k}(\frac{(1-q)^n}{q})^{N-k}}{(\frac{1}{q})^N} = \binom{N}{k}(1-q)^{n(N-k)}q^k.
\end{aligned} \tag{A.16}$$

(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, (A.15) and (A.16) diverge as $m \to \infty$. Hence, the only regular paths are those of part (a). □

Acknowledgments. The authors thank two anonymous referees for their comments, corrections and suggestions.

REFERENCES

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

[2] BRODERICK, T., PITMAN, J. and JORDAN, M. I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836. MR3150470

[3] DYNKIN, E. B. (1978). Sufficient statistics and extreme points. *Ann. Probab.* **6** 705–730. MR0518321

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

[5] GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* (*N.Y.*) **138** 5674–5685. MR2160320

[6] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2006). Infinite latent feature models and the Indian buffet process. *Adv. Neural Inf. Process. Syst.* **18** 475–482.

[7] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2011). The Indian buffet process: An introduction and review. *J. Mach. Learn. Res.* **12** 1185–1224. MR2804598

[8] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249

[9] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. Springer, Berlin.

[10] TEH, Y. and GÖRÜR, D. (2010). Indian buffet processes with power-law behavior. *Adv. Neural Inf. Process. Syst.* **22** 1838–1846.

M. BATTISTON
Y. W. TEH
DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
24-29 ST GILES
OXFORD OX1 3LB
UNITED KINGDOM
E-MAIL: marco.battiston@stats.ox.ac.uk
y.w.teh@stats.ox.ac.uk

S. FAVARO
DEPARTMENT OF ECONOMICS AND STATISTICS
UNIVERSITY OF TORINO
CORSO UNIONE SOVIETICA 218/BIS
10134 TORINO
ITALY
E-MAIL: stefano.favaro@unito.it

D. M. ROY
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF TORONTO
100 ST. GEORGE STREET
TORONTO, ONTARIO M5S 3G3
CANADA
E-MAIL: droy@utstat.toronto.edu