Random Discrete Distributions Invariant Under Size-biased Permutation

Jim Pitman *
Department of Statistics
U.C. Berkeley CA 94720

July 25, 1995

Abstract

Invariance of a random discrete distribution under size-biased permutation is equivalent to a conjunction of symmetry conditions on its finite-dimensional distributions. This is applied to characterize residual allocation models with independent factors that are invariant under size-biased permutation. Apart from some exceptional cases and minor modifications, such models form a two-parameter family of generalized Dirichlet distributions.

RESIDUAL ALLOCATION MODEL; EXCHANGEABLE RANDOM PARTITION; GENERALIZED DIRICHLET DISTRIBUTION.

1991 Math. Subject Classification: Primary 62E10. Secondary 60G09

*Research supported by N.S.F. Grants MCS91-07531 and DMS-9404345

1 Introduction

Let $(P_n) = (P_1, P_2, \cdots)$ be a random discrete distribution over $\{1, 2, \cdots\}$. That is to say, the P_n are non-negative random variables with $\sum_n P_n = 1$. Define $(\tilde{P}_n)$, a *size-biased random permutation* (SBP) of (P_n), by the following procedure. Given (P_n), let $X(1), X(2), \cdots$ be i.i.d random variables with distribution (P_n). Let $\nu(n) = X(N_n)$ where $1 = N_1 < N_2 < \cdots$ are the successive times that new values appear in the sample sequence $X(1), X(2), \cdots$. Let $\tilde{P}_n = P_{\nu(n)}$, with the convention $\tilde{P}_n = 0$ if there are fewer than n distinct values in the sample sequence. It is easily seen that $(\tilde{P}_n)$ is a random discrete distribution over $\{1, 2, \cdots\}$. Call (P_n) *invariant under size-biased permutation* (ISBP) if the sequence $(\tilde{P}_n)$ has same finite dimensional distributions as (P_n). The main motivation for study of random discrete distributions that are ISBP is described in the next paragraph. See Patil and Taillie [15], Donnelly and Joyce [4], Donnelly [3], Ewens [6], Zabell [22] for further background.

Let π be a random discrete probability measure on the real line, or a standard Borel space. Given π, let $X_1, X_2, \cdots$ be i.i.d random variables with distribution π. Let P_n be the π-measure of the nth distinct value observed in the random sample (X_i) from π, with the convention $P_n = 0$ if there are fewer than n distinct values in the sample sequence. That is to say, P_n is the almost sure limiting frequency in the sequence $(X_1, X_2, \cdots)$ of the nth distinct value observed in the sequence of exchangeable random variables $(X_1, X_2, \cdots)$. Think of the atoms of π as representing the frequencies with which various species are present in an infinite population. Then $X_1, X_2, \cdots$ represents the sequence of species obtained by random sampling. And P_n is the proportion in the whole population of the nth species observed in the random sample. It is easily verified that the most general possible distribution for (P_n) is one that is ISBP.

As a consequence of the previous paragraph and the representation theory of random partitions due to Kingman[12, 13], the same class of random discrete distributions that are ISBP can be described a little differently as follows. Consider a partition of the positive integers into equivalence classes $A_1, A_2, \cdots$ for an exchangeable random equivalence relation, where the classes A_n are ordered by their least elements. Then the long run relative frequencies P_n of A_n exist almost surely, and, assuming $\sum_n P_n = 1$ almost surely, (P_n) is ISBP. The distribution of the random partition can then be recovered from the product moments of (P_n) displayed later in formula (5).

The problem considered in this paper is how to characterize those random discrete probability distributions (P_n) that are ISBP. A basic result in this vein is the following:

Theorem 1 (McCloskey [14]). *Suppose that*

$$P_n = \overline{W}_1 \overline{W}_2 \cdots \overline{W}_{n-1} W_n, \;\; n \geq 1 \tag{1}$$

where $W_1, W_2, \cdots$ are i.i.d with values in $[0,1]$, and $\overline{W}_i = 1 - W_i$. Then (P_n) is ISBP iff the common distribution of the W_i is beta$(1,\theta)$ for some $0 \leq \theta < \infty$.

Here, for $a > 0,\ b > 0$, the beta(a,b) distribution on $[0,1]$ has density

$$B(a,b)^{-1} x^{a-1} \overline{x}^{b-1}, \qquad 0 < x < 1,$$

where $\overline{x} = 1 - x$, and the beta$(a,0)$ distribution is a unit mass at 1. McCloskey derived the "if" part of his result by showing that if (P_n) is the size-biased permutation of the random probability distribution (π_n) on $\{1, 2, \cdots\}$ defined by

$$\pi_n = X_n / \Sigma, \tag{2}$$

where $X_1 > X_2 > \cdots$ are the points of a Poisson point process on$(0,\infty)$ with intensity measure

$$\Lambda(dx) = \theta x^{-1} e^{-\lambda x}, \;\; x > 0$$

for some $\lambda > 0,\ \theta > 0$, and $\Sigma = \sum_n X_n$, then (1) holds for independent W_i with beta$(1,\theta)$ distribution. The distribution of the ranked sequence (π_n) is then the *Poisson-Dirichlet distribution* with parameter θ, introduced by Kingman [11]. See also Watterson [21], Ignatov [10], Hoppe [9].

Perman, Pitman and Yor[16] found the joint law of (P_n) derived as above from a Poisson process on $(0,\infty)$ with a more general intensity measure Λ such that $\Sigma < \infty$ a.s.. McCloskey formulated the "only if" part of his result assuming (P_n) was the SBP of (π_n) derived as in (2) from the points of such a Poisson process on $(0,\infty)$. But McCloskey's argument establishes the more general result stated in Theorem 1, as it is based equating moments in the following identity, which holds for any (P_n) that is ISBP: If P' is a random variable which given (P_n) is such that

$$P' = \begin{cases} P_1 \text{ with probability } P_1 \\ P_2 \text{ with probability } 1 - P_1, \end{cases} \tag{3}$$

then P' has the same distribution as P_1. To see this, let P_n represent the π-measure of the nth distinct value observed in a random sample (X_i) from some random discrete distribution π. Then $P_1 = \pi(X_1)$, $P_2 = \pi(X_2)$ if $X_2 \neq X_1$, and P' as above is obtained by setting $P' = \pi(X_2)$. A multidimensional form of this property appears in Theorem 4 below, which gives a symmetry condition on the joint distribution of $(P_1, \cdots, P_n)$ for $n = 2, 3, \cdots$ that is both necessary and sufficient for (P_n) to be ISBP.

Motivation for this development is provided by the following problem posed by Patil and Taillie[15], and solved in this paper: for what independent, perhaps non-identically distributed sequences (W_n) does the formula (1) define a random discrete distribution (P_n) that is ISBP? The model (1) for a random discrete distribution (P_n), with independent W_n, is called a *residual allocation model* (RAM). This model has been considered in a number of contexts. Freedman[8], Fabius[7], and Connor and Mosimann[2] studied the model in the setting of Bayesian statistics. A prior distribution of this form over probabilities on $\{1, 2, \cdots\}$ has the feature that given data from a sequence of observations, which given (P_n) are i.i.d according to (P_n), the posterior distribution of (P_n) is of the same form. Such priors are called *tail free*, or *completely neutral.* If for each n the distribution of W_n is beta(a_n, b_n) for some a_n, b_n, the joint distribution of $(P_1, P_2, \cdots)$ is known as a *generalized Dirichlet* distribution. In particular, in case $b_k = \sum_{i=k+1}^m a_i$ for some $m \geq 2$, the joint distribution of $(P_1, \cdots, P_m)$ is the *Dirichlet* $(a_1, \cdots, a_m)$ *distribution*, that is to say the joint distribution of

$$(Y_1/\Sigma, \cdots, Y_m/\Sigma)$$

where $Y_1, \cdots, Y_m$ are independent gamma random variables with common scale parameter and shape parameters $a_1, \cdots, a_m$, and $\sum = \sum_{i=1}^m Y_i$. Patil and Taillie[15] noted that if $(\pi_1, \cdots, \pi_m)$ has Dirichlet $(\beta, \cdots, \beta)$ distribution, then the SBP of $(\pi_1, \cdots, \pi_m)$ follows a RAM that is generalized Dirichlet with parameters $a_n \equiv 1 + \beta$, $b_n = m\beta - n\beta$, $n = 1, \cdots, m$.

2 Results

According to the following theorem and its corollary, apart from some simple examples and modifications, the only RAM's that are ISBP are the McCloskey and Patil-Taillie schemes discussed above, and a scheme considered in quite different contexts by Engen[5] and Perman, Pitman and Yor[16]. These schemes form a two parameter family of generalized Dirichlet distributions, as indicated in cases (i) and (ii)a) of the theorem. The scheme in (ii)b) is obtained from (ii)a) by letting $\beta \to \infty$. And the scheme in (ii)c) is the most general distribution of the SBP of a random probability distribution on two points. Theorem 1 is an immediate consequence of Theorem 2.

Theorem 2 *Let $(P_1, P_2, \cdots)$ be such that $P_n \geq 0$, $\sum_n P_n = 1$, $P_1 < 1$, and $P_n = \overline{W}_1 \cdots \overline{W}_{n-1} W_n$ for independent W_i. Then (P_n) is ISBP iff one of the following four conditions* (i), (ii)a), (ii)b) or (ii)c) *holds:*

(i) *$P_n > 0$ a.s. for all n, in which case the distribution of W_n is* beta$(1-\alpha, \theta+n\alpha)$, *for every $n = 1, 2, \cdots$, for some $0 \leq \alpha < 1$, $\theta > -\alpha$.*

or (ii) *$\{n : P_n > 0\} = \{1, \cdots, m\}$ a.s. for some integer constant m, in which case either*

a) *for some $\beta > 0$, the distribution of W_n is* beta$(1 + \beta, m\beta - n\beta)$ *for $n = 1, \cdots, m$.*

or b) *$W_n = 1/(m - n + 1)$ a.s., that is to say $P_n = 1/m$ a.s., for $n = 1, \cdots, m$,*

or c) *$m = 2$ and the distribution F on $(0, 1)$ defined by*

$$F(dw) = \overline{w} P(W_1 \in dw)/E(\overline{W}_1)$$

is symmetric about $1/2$.

For $0 < \alpha < 1$, $\theta > 0$, Engen[5] showed that for (P_n) as in case (i) of Theorem 2, a single size-biased pick from (P_n) has the same distribution as P_1. The full invariance of (P_n) under size-biased permutation in this case follows from the work of Perman, Pitman and Yor [16]. For $0 < \alpha < 1$, $\theta = 0$, they showed that (P_n) as in case (i) of Theorem 2 appears as the SBP of (X_n/Σ) derived from a Poisson process of points X_n with intensity measure

$$\Lambda(dx) = Kx^{-\alpha-1}dx, \; x > 0$$

for a constant K, so that the distribution of $\Sigma = \sum_n X_n$ is stable with index α. For $0 < \alpha < 1$ and $\theta > -\alpha$, they showed that if the underlying probability measure is changed by a density factor proportional to $\Sigma^{-\theta}$, relative to the Poisson model, then (P_n) remains ISBP, while the W_n are independent with beta$(1-\alpha, \theta+n\alpha)$ distributions. (In the special case $\theta = k\alpha$ for some positive integer k, that (P_n) stays ISBP follows from the case $\theta = 0$ by simply shifting along the sequence: if $(W_1, W_2, \cdots)$ induces (P_n) that is ISBP, then so does $(W_{k+1}, W_{k+2}, \cdots)$ given $\overline{W}_1\overline{W}_2\cdots\overline{W}_k > 0$, for each $k \geq 1$. This can be seen directly, or obtained as a consequence of Remark 8.)

Sections 3 and 4 of this paper provide a unified proof of Theorem 2, without using the Poisson representation for the "if" part. The following immediate corollary of Theorem 2 takes care of the possibility that $P(P_1 = 1) > 0$. See Zabell [23] for an application where this possibility is of some interest.

Corollary 3 *Let (P_n) be a random discrete distribution on $\{1, 2, \cdots\}$, represented as $P_n = \overline{P}_1\overline{W}_2\cdots\overline{W}_{n-1}W_n$, $n \geq 2$, for independent P_1, $W_2, W_3, \cdots$. Assuming that $P(P_1 < 1) > 0$, let W_1 be independent of $W_2, W_3, \cdots$ with the distribution of P_1 given $P_1 < 1$. Then (P_n) is ISBP iff either $P_1 = 1$ a.s. or $W_1, W_2, \cdots$ is of one of the forms described in Theorem 2.*

Put another way, the distribution of (P_n) is a probabilistic mixture of the degenerate distribution that makes $P_1 = 1$, and one of the distributions of (P_n) described by Theorem 2. The above results show that the ISBP condition imposes severe restrictions on the joint law of (P_n). These restrictions seem at first hard to understand, as the ISBP condition appears to be essentially infinite dimensional. The central result of this paper is that despite these appearances, a simple conjunction of conditions on the *finite*-dimensional joint distributions of a sequence $(P_1, P_2, \cdots)$ is equivalent to ISBP. This is stated in the following theorem, which is established in Section 3 , and applied to prove Theorem 2 in Section 4.

Theorem 4 *Let $(P_1, P_2, \cdots)$ be a sequence of random variables satisfying the almost sure constraints $P_i \geq 0$ for $i \geq 1$, and $\sum_{i=1}^n P_i \leq 1$, $n = 1, 2, \cdots$. Let G_k denote the measure on $\mathbb{R}^k$ whose density with respect to the joint probability distribution of $(P_1, \cdots, P_k)$ at $(p_1, \cdots, p_k)$ is $\prod_{i=1}^{k-1}(1 - \sum_{j=1}^i p_j)$:*

$$G_k(dp_1, \cdots, dp_k) = P(P_1 \in dp_1, \cdots, P_k \in dp_k) \prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} p_j) \tag{4}$$

The following statements are equivalent:

(i) $\sum_i P_i = 1$ *a.s. and* $(P_1, P_2, \cdots)$ *is ISBP.*

(ii) $P_1 > 0$ *a.s. and for each* $k = 2, 3, \cdots$, *the measure* G_k *is symmetric with respect to permutations of the coordinates in* $\mathbb{R}^k$.

(iii) $P_1 > 0$ *a.s. and for each* $k = 2, 3, \cdots$ *the function of* k*-tuples of positive integers*

$$(n_1, \cdots, n_k) \to E\left[\prod_{i=1}^{k} P_i^{n_i - 1} \prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} P_j)\right], \tag{5}$$

is a symmetric function of $(n_1, \cdots, n_k)$.

Note the surprising feature of Theorem 4 that the condition $\sum_i P_i = 1$ a.s. in (i) is *not* assumed in (ii) and (iii), but is nonetheless implied by these symmetry conditions. By contrast, for arbitrary random variables $P_i \geq 0$ the condition $\sum_i P_i = 1$ a.s. alone is obviously not just a simple conjunction of conditions on the joint distributions of $P_1, \cdots, P_k$. (For the condition $\sum_i P_i = 1$ a.s. imposes no constraint on the law of $P_1, \cdots, P_k$ besides $\sum_1^k P_i \leq 1$, and the conjunction of these conditions is $\sum_i P_i \leq 1$ a.s.).

The proof of Theorem 4 provides a probabilistic interpretation of the measure G_k in (4). And it shows that the function in (5) for $(n_1, \ldots, n_k)$ with $\sum_i n_i = n$ defines the distribution of a random partition of the first n positive integers derived from an exchangeable random partition of all positive integers, constructed in such a way that the P_k are the long run relative frequencies of classes ordered by their least elements. In the case of McCloskey's Theorem 1, the corresponding random partition of n is that defined by Ewens' sampling formula. See Ewens [6]. See Pitman [18, 17] for analysis of the two-parameter family of random partitions corresponding to Theorem 2, and Pitman-Yor [20] regarding the two-parameter analog of the Poisson-Dirichlet distribution.

3 Symmetry in size-biased sampling

This section presents a proof of Theorem 4, then draws some corollaries.
Proof of Theorem 4
(i) $\Rightarrow$ (ii): Because (P_k) is ISBP, it can be assumed that (P_k) is represented as

$$P_k = \pi(X_{N(k)}), \ k = 1, 2, \cdots$$

where $\pi = (\pi(1), \pi(2), \cdots)$ is a random discrete probability distribution distributed like $(P_1, P_2, \cdots)$, given π the (X_i) are i.i.d. according to π, and the $N(k)$ are the times that successive distinct X-values appear, with the convention $P_k = 0$ in case fewer than k distinct X-values ever appear. Define indicator random variables

$$Z_k = 1\{X_1, \cdots, X_k \text{ are all distinct}\}. \tag{6}$$

Then for each k the random vector

$$(P_1, P_2, \cdots, P_k)Z_k = (\pi(X_1), \pi(X_2), \cdots, \pi(X_k))Z_k \tag{7}$$

clearly has an exchangeable joint distribution. But since

$$P(Z_k = 1 | P_1, \cdots, P_k) = \prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} P_j),$$

the distribution of the exchangeable random vector (7) and measure G_k defined by (4) are identical when restricted to $\mathbb{R}^k - \{0\}$, where 0 is the origin in $\mathbb{R}^k$. Thus G_k is symmetric.
(ii) $\Leftrightarrow$ (iii): This is immediate from the definition of G_k, and the fact that polynomials are dense in the space of continuous function on $[0,1]^k$.
(iii) $\Rightarrow$ (i): The argument is based on the approach of Aldous [1] to Kingman's [13] representation of exchangeable random partitions of $\mathbb{N} := \{1, 2, \cdots\}$. See Pitman [17] for generalizations of the argument to the setting of partially exchangeable random partitions of $\mathbb{N}$.

Define a sequence Π_n of partitions of $\mathbb{N}_n := \{1, \cdots, n\}$ as follows: $\Pi_1 = \{1\}$; and for each $n \in \mathbb{N}$, conditionally given $\Pi_n = \{\{A_i\}_1^k\}$, where $\{A_i\}$ is a partition of $\mathbb{N}_n$ into non-empty subsets of sizes n_i that satisfy the *order constraint*: $1 \in A_1$, the least element not in A_1 is in A_2, and so on,

Π_{n+1} is an extension of Π_n in which element $n+1$ attaches to class A_i with probability P_i, $1 \leq i \leq k$, and forms a new class with probability $R_i := 1 - P_1 - \cdots - P_i$. By construction, the partitions Π_n are consistent as n varies, so they induce a random partition Π of $\mathbb{N}$. Also by construction, for $\{A_i\}_1^k$ that satisfy the order constraint:

$$P(\Pi_n = \{A_i\}_1^k) = E(\prod_{i=1}^{k} P_i^{n_i - 1} \prod_{i=1}^{k-1} R_i). \tag{8}$$

This probability depends on $A_1, \cdots, A_k$ only through their sizes $n_1, \cdots, n_k$, and hypothesis (iii) amounts to symmetry of the right hand side of (8) as a function of $(n_1, \cdots, n_k)$, for each $k \geq 2$. It follows that Π is exchangeable in Aldous' sense. Aldous [1] uses further randomization to construct a random probability distribution π on $[0,1]$, and a random sequence $(X_1, X_2, \cdots)$, which given π is i.i.d. according to π, and which generates Π as the collection of equivalence classes for the equivalence relation

$$i \sim j \Leftrightarrow X_i = X_j, \ i, j \in \mathbb{N}.$$

From the original construction of Π and the law of large numbers, P_k is the long run relative frequency of numbers in the kth class of Π to appear. But Aldous' construction identifies P_k as the π-measure of the kth distinct value to appear in the sequence $(X_1, X_2, \cdots)$. If $P(\sum_k P_k = 1) < 1$ then π must have a non-trivial continuous component with positive probability, and given this event there is positive probability that X_j is distinct from X_1 for every $j > 1$, in which case $P_1 = 0$. But this possibility is precluded by the assumption that $P_1 > 0$ a.s.. Thus $\sum_k P_k = 1$ a.s., and (P_k) is a size-biased presentation of the atoms of π. Therefore (P_k) is ISBP. $\Box$

An immediate consequence of Theorem 4 is

Corollary 5 *Suppose $(P_1, P_2, \cdots)$ is a sequence of random variables such that for each n,*

$$P(P_1 \in dp_1, \cdots, P_n \in dp_n) = f_n(p_1, \cdots, p_n) dp_1, \cdots, dp_n,$$

for a joint density f_n such that $f_n(p_1, \cdots, p_n) = 0$ unless $p_i \geq 0$ and $\sum_{i=1}^n p_i \leq 1$, and

$$f_n(p_1, \cdots, p_n) \prod_{j=1}^{n-1} (1 - \sum_{i=1}^{j} p_i) \tag{9}$$

is a symmetric function of $(p_1, \cdots, p_n)$*. Then* $(P_1, P_2, \cdots)$ *defines a random discrete probability distribution which is ISBP.*

By change of variables the above condition on f_n becomes a simpler condition on the joint density g_n of $W_1, \cdots, W_n$ such that $P_n = \overline{W}_1 \cdots \overline{W}_{n-1} W_n$, namely:

$$g_n(w_1, \cdots, w_n) = 0 \text{ unless } 0 \le w_i \le 1,$$

and

$$g_n\left(p_1, \frac{p_2}{1-p_1}, \cdots, \frac{p_n}{1-p_1-\cdots-p_{n-1}}\right) \tag{10}$$

is a symmetric function of $p_1, \cdots, p_n$. Characterization of such joint densities g_n of product form is provided in the next section. More generally, the result of Theorem 4 can be reformulated as follows.

Definition 6 Say the joint distribution of a pair of random variables (W_1, W_2) is *acceptable* if $0 < W_i \le 1$ a.s., and for $P_1 = W_1,\ P_2 = \overline{W}_1 W_2$, the joint law of (P_1, P_2) is such that the distribution G_2 in (4) is symmetric.

In view of (5), a joint distribution for (W_1, W_2) is acceptable iff $0 < W_i \le 1$ a.s., and

$$m(r, s) := E[W_1^r \overline{W}_1^{s+1} W_2^s] \tag{11}$$

is such that $m(r, s) = m(s, r)$ for all pairs of non-negative integers r and s.

Corollary 7 *Let* $(W_1, W_2, \cdots)$ *be a sequence of random variables with* $0 < W_1 \le 1$ *a.s. and let* $P_n = \overline{W}_1 \cdots \overline{W}_{n-1} W_n$, $n = 1, 2, \cdots$*. The following are equivalent:*

(i) (P_n) *is a random probability distribution that is ISBP.*

(ii) (P_n) *satisfies each of these three conditions:*

- a) *the law of* (W_1, W_2) *is acceptable;*
- b) *for each* $k = 1, 2, \cdots$*, on the event* $P_1 + \cdots + P_k < 1$ *there is a version of the conditional law of* (W_{k+1}, W_{k+2}) *given* $(P_1, \cdots, P_k)$ *that is acceptable;*
- c) *the conditional law in* b) *can be chosen to depend symmetrically on* $(P_1, \cdots, P_k)$*.*

Proof. Denote by $p(n_1, \cdots, n_k)$ the function of $(n_1, \cdots, n_k)$ defined by (5). Use the notation

$$R_i = 1 - \sum_{j=1}^{i} P_j = \prod_{j=1}^{i} \overline{W}_j$$

to see that for $k = 1, 2 \cdots$

$$p(n_1, \cdots, n_{k+1}) = E\left[\left(\prod_{i=1}^{k} P_i^{n_i - 1}\right)\left(\prod_{i=1}^{k-1} R_i\right) R_k^{n_{k+1}} W_{k+1}^{n_{k+1}-1}\right] \tag{12}$$

$$p(n_1, \cdots, n_{k+1}, n_{k+2}) = E\left[\left(\prod_{i=1}^{k} P_i^{n_i - 1}\right)\left(\prod_{i=1}^{k-1} R_i\right) R_k^{n_{k+1}+n_{k+2}} W_{k+1}^{n_{k+1}-1} \overline{W}_{k+1}^{n_{k+2}} W_{k+2}^{n_{k+2}-1}\right] \tag{13}$$

According to part (iii) of Theorem 4, (P_n) is ISBP iff both (W_1, W_2) is acceptable, and $p(n_1, \cdots, n_{k+1}, n_{k+2})$ is symmetric for every $k = 1, 2, \cdots$. The implication (i) $\Rightarrow$ (ii) follows from this. For example (ii)b) comes from the symmetry of $p(n_1, \cdots, n_{k+1}, n_{k+2})$ as a function of (n_{k+1}, n_{k+2}) for fixed $(n_1, \cdots, n_k)$, and the fact that polynomials are dense among continuous functions of $(P_1, \cdots, P_k)$. And (ii)c) follows from the symmetry of $p(n_1, \cdots, n_{k+1}, n_{k+2})$ as a function of $(n_1, \cdots, n_k)$ for fixed (n_{k+1}, n_{k+2}). The key step of this argument can be understood in lighter notation as follows. If X and Y are random variables such that

$$E[\, f(X, Y) g(X, Y)\,] = E[f(Y, X) g(X, Y)\,]$$

for some fixed product measurable g and all non-negative product measurable f, and Z is a random variable with

$$E[\, f(X, Y) g(X, Y) h(Z)\,] = E[f(Y, X) g(X, Y) h(Z)\,]$$

for some non-negative measurable h, and f, g as before, then there is a version of $E[h(Z)|X, Y]$ that equals a symmetric function of (X, Y) on the event $\{g(X, Y) > 0\}$. And if this condition holds for enough functions h there is a regular conditional distribution for Z given (X, Y) that depends symmetrically on (X, Y) on the event $\{g(X, Y) > 0\}$.

The converse implication (ii) $\Rightarrow$ (i) is proved by induction on k, starting from $k = 2$. Note first that from (ii)b) and (13) with $k - 1$ instead of k that

$$p(n_1, \cdots, n_{k+1}) \text{ is symmetric in } (n_k, n_{k+1}) \text{ for fixed } (n_1, \cdots, n_{k-1}). \tag{14}$$

Also, condition (ii) c) implies

(ii)c') for each $k = 1, 2, \cdots$ the conditional distribution of W_{k+1} given $(P_1, \cdots, P_k)$ on the event $R_k > 0$ can be chosen to be a symmetric function of $(P_1, \cdots, P_k)$

Now make the inductive hypothesis that $p(n_1, \cdots, n_k)$ is symmetric, and use the condition (ii)c') in formula (12) to see that

$$p(n_1, \cdots, n_{k+1}) \text{ is symmetric in } (n_1, \cdots, n_k) \text{ for fixed } n_{k+1}. \qquad (15)$$

Combining (14) and (15) shows that $p(n_1, \cdots, n_{k+1})$ is fully symmetric in $(n_1, \cdots, n_{k+1})$, completing the induction. □

Remark 8 The proof shows that the corollary holds just as well with (ii)c) replaced by the weaker condition (ii) c'). A variation of the argument shows that every (P_n) that is ISBP satisfies the following stronger from of (ii) b)-c):

(ii)d) *for each $k = 1, 2, \cdots$ there is a version of the conditional distribution of $(P_{k+n}/R_k, n = 1, 2 \cdots)$ given $(P_1, \cdots, P_k)$ on the event $R_k > 0$ that is ISBP and that depends symmetrically on $P_1, \cdots, P_k$.*

Remark 9 Condition (ii) of the Corollary is analogous to the following necessary and sufficient condition for a sequence $(X_1, X_2, \cdots)$ to be exchangeable: (X_1, X_2) is exchangeable, and for each $n = 2, 3, \cdots$ there is a version of the conditional law of (X_{n+1}, X_{n+2}) given $(X_1, \cdots, X_n)$ that is exchangeable, and that depends exchangeably on $(X_1, \cdots, X_n)$. The proof of Corollary 7 follows the same pattern as the proof of this more intuitively obvious result, with extra density factors in the calculations.

4 Residual Allocation Models

This section applies the general results of Section 3 to the RAM

$$P_n = \overline{W}_1 \cdots \overline{W}_{n-1} W_n$$

for independent W_i. The final result is Theorem 2 stated in the introduction. The first step is provided by

Lemma 10 *Let $(W_1, W_2 \cdots)$ be a sequence of independent random variables with $0 < W_i \leq 1$ a.s., and let $P_n = \overline{W}_1 \cdots \overline{W}_{n-1} W_n$, $n = 1, 2, \cdots$. Then the following are equivalent:*

(i) (P_n) *is a random probability distribution that is ISBP.*

(ii) *the law of* (W_n, W_{n+1}) *is acceptable for every* $n < m$, *where*

$$m = \inf\{n : P(W_n = 1) = 1\}.$$

Proof. This follows immediately from Corollary 7.

The problem now boils down to characterizing all acceptable laws for (W, Z) say, where W and Z are independent. That is to say, from (11), all possible pairs of distributions for random variables W and Z with $0 < W \leq 1$, $0 < Z \leq 1$, such that

$$m(r, s) := E(W^r \overline{W}^{s+1}) E(Z^s) \tag{16}$$

is symmetric function of non-negative integers r and s. From (10) for $n = 2$ we obtain an elementary sufficient condition for acceptability of independent W and Z with densities say f and g on $(0, 1)$, namely:

$$f(p) g\left(\frac{q}{\overline{p}}\right) = f(q) g\left(\frac{p}{\overline{q}}\right) \tag{17}$$

for $0 < p < 1$, $0 < q < 1$. In particular, in case f and g are beta densities with parameters (a, b) and (c, d) respectively (17) becomes

$$p^{a-1}\overline{p}^{b-1}\left(\frac{q}{\overline{p}}\right)^{c-1}\left(\frac{\overline{p}-q}{\overline{p}}\right)^{d-1} = q^{a-1}\overline{q}^{b-1}\left(\frac{p}{\overline{q}}\right)^{c-1}\left(\frac{\overline{q}-p}{\overline{q}}\right)^{d-1},$$

for $p > 0, q > 0, p + q < 1$, which simplifies to

$$p^{a-1}\overline{p}^{b-c-d+1} q^{c-1} = p^{c-1} q^{a-1} \overline{q}^{b-c-d+1}.$$

Clearly, this identity holds iff $c = a$ and $d = b - a + 1$. And it is easy to see that for W and Z with beta densities these conditions are in fact necessary for (W, Z) to be acceptable. Thus we obtain

Lemma 11 *If* W *and* Z *are independent with beta* (a, b) *distribution and beta* (c, d) *distribution, respectively, for strictly positive* a, b, c, d, *then* (W, Z) *is acceptable iff* $c = a$ *and* $d = b - a + 1$.

In particular for $a = 1$ and $b > 0$, Lemma 11 shows that W and Z i.i.d. beta$(1, b)$ makes an acceptable pair. So the "if" part of McCloskey's Theorem 1 follows at once from Lemmas 10 and 11. So does the "if" part of Theorem 2 in case (i). The entirety of Theorem 2 follows from the next lemma combined with the symmetry condition for the moments (16) that identifies an acceptable independent pair (W, Z), and Lemma 10.

Lemma 12 *For a random variable W with $0 < W < 1$, the following statements* (i) *and* (ii) *are equivalent:*

(i) *there exists a r.v. Z with $0 < Z \leq 1$ such that for $r = 0$ and 1, and $s = 1, 2, \cdots$*

$$E(W^r \overline{W}^{s+1})E(Z^s) = E(W^s \overline{W}^{r+1})E(Z^r) \tag{18}$$

(ii) *either*

A) *the distribution F on $(0, 1)$, defined by*

$$F(dw) = \overline{w}P(W \in dw)/E(\overline{W}), \tag{19}$$

is symmetric about $1/2$.

or

B) *W has beta(α, β) distribution for some $\alpha < \beta + 1$*

or

C) *$W = c$ a.s. for some constant c with $0 < c < 1/2$.*

In case A), *$Z = 1$ a.s., whereas in case* B), *Z has beta$(\alpha, \beta + 1 - \alpha)$ distribution, and in case* C), *$Z = c/(1 - c)$ a.s. In any case, identity* (18) *holds for all positive real r and s.*

Proof. Given some distribution for W, let Y be a r.v. with distribution F as in (19), so for any bounded function g,

$$Eg(Y) = E(g(W)\overline{W})/E(\overline{W}).$$

Then each condition above becomes a corresponding condition on Y, as in the statement of Lemma 13 below. Thus Lemma 12 is a consequence of Lemma 13. □

Lemma 13 *For a random variable Y with $0 < Y < 1$, the following statements* (i) *and* (ii) *are equivalent:*

(i) *there exists a random variable Z with $0 < Z \leq 1$ such that for $r = 0$ and 1, and $s = 1, 2, \cdots$,*

$$E(Y^r \overline{Y}^s) E(Z^s) = E(Y^s \overline{Y}^r) E(Z^r), \tag{20}$$

where $\overline{Y} = 1 - Y$.

(ii) *either*

A) *Y has a distribution symmetric about $1/2$,*

or

B) *Y has beta(a, b) distribution for some $a < b$.*

or

C) *$Y = c$ a.s. for some constant c with $0 < c < 1/2$.*

In case A), *$Z = 1$ a.s., whereas in case* B), *Z has beta$(a, b - a)$ distribution, and in case* C), *$Z = c/(1 - c)$ a.s. In any case, identity (20) holds for all positive real r and s.*

Proof. Suppose $0 < Y < 1$ and (i) holds. For $k = 0, 1, \cdots$ let

$$\mu_k = EY^k; \ \nu_k = EZ^k.$$

Note first that (20) for $r = 0$, $s = k$ implies

$$\nu_k = E(Y^k)/E(\overline{Y}^k), \tag{21}$$

In particular

$$\nu_1 = \mu_1/(1 - \mu_1) \tag{22}$$

and in general

$$\nu_k = \mu_k / p_k(\mu_1, \cdots, \mu_k) \tag{23}$$

for some polynomial p_k. Thus the distribution of Z is determined by that of Y. In particular, Y is symmetric about $1/2$, that is $E(Y^k) = E(\overline{Y}^k)$ for all

k, iff $Z = 1$ a.s.. And $Y = c$ a.s. for some constant c iff $\nu_k = (c/\overline{c})^k$ for all k, that is $Z = c/\overline{c}$, in which case $c \leq 1/2$ by the assumption $Z \leq 1$.

Next, suppose that (i) holds, and that Y is neither constant, nor symmetric about $1/2$. Then by the preceding argument, $P(Z = 1) < 1$. Since we assume $0 < Z \leq 1$ a.s., this implies

$$1 > \nu_1 > \nu_2 > \cdots > 0 \tag{24}$$

Now (20) for $r = 1$ states that for $s = 2, 3 \cdots$

$$E(Y\overline{Y}^s)\nu_s = E(Y^s\overline{Y})\nu_1,$$

which rearranges to show that for $s = 2, 3, \cdots$

$$\mu_{s+1} = \frac{r_s(\mu_1, \mu_2, \cdots, \mu_s)}{\nu_s(-1)^s + \nu_1} \tag{25}$$

for some polynomial r_s, where the denominator never vanishes, because $0 < \nu_s < \nu_1$ by (24). Now (23) combined with (25) shows that

$$\mu_{s+1} = f_s(\mu_1, \cdots, \mu_s), \;\; s = 2, 3, \cdots$$

for some function f_s, hence by induction

$$\mu_k = g_k(\mu_1, \mu_2); \quad \nu_k = h_k(\mu_1, \mu_2), \quad k = 3, 4, \cdots \tag{26}$$

for some functions g_k and h_k. To summarise we have established the following:

Uniqueness Claim. *For all μ_1 and μ_2 with*

$$0 < \mu_1^2 < \mu_2 < \mu_1 < 1/2, \tag{27}$$

there is at most one distribution for Y with $0 < Y < 1$, $E(Y) = \mu_1$, $E(Y^2) = \mu_2$, and at most one distribution of Z with $0 < Z \leq 1$, such that (20) holds.

To complete the argument, note that the a priori constraints (27) determine unique a, b with $0 < a < b$ and

$$\mu_1 = \frac{a}{a+b}, \quad \mu_2 = \frac{a}{(a+b)}\frac{(a+1)}{(a+b+1)}.$$

If Y has beta(a, b) distribution, then

$$E(Y^r \overline{Y}^s) = \frac{[a]_r [b]_s}{[a+b]_{r+s}}$$

where e.g. $[a]_r = a(a+1)\cdots(a+r-1) = \Gamma(a+r)/\Gamma(a)$. But then $EY = \mu$, $EY^2 = \mu_2$, and it is obvious that (20) holds for Z with the beta$(a, b-a)$ distribution which makes

$$EZ^k = [a]_k/[b]_k.$$

Remark 14 Imposing conditions to avoid cases A) and C) above gives two characterizations of the beta family of distributions on $(0,1)$

1) A r.v. Y with non-degenerate distribution on $(0,1)$ has beta(a, b) distribution for some $0 < a < b$ iff there exists Z with $0 < Z < 1$ such that (20) holds.

2) A r.v. Z with non-degenerate distribution on $(0,1)$ has beta(a, c) distribution for some $a > 0$, $c > 0$ iff there exists Y with $0 < Y < 1$ such that (20) holds.

5 Structural Distributions

Given a random discrete distribution (P_n) let F denote the distribution on $(0,1]$ of $\tilde{P}_1$ which is a *size-biased pick* from (P_n), that is the first term of a SBP $(\tilde{P}_n)$ of (P_n). Following Engen [5], call F the *structural distribution* of (P_n). As noted by Engen and others, this one-dimensional probability distribution carries a great deal of information about the distribution of the whole sequence (P_n). For example, any positive measurable function g

$$E\left[\sum_n g(P_n)\right] = \int_0^1 \frac{g(v)}{v} F(dv) \tag{28}$$

Taking $g(v) = 1(a < v < b)$ gives an expression in terms of F for the mean number of n such that $a < P_n < b$. Thus $v^{-1}F(dv)$ is the measure known as the *frequency spectrum* in population genetics [6].

The structural distribution F also appears in formulae related to Kingman's partition structure induced by (P_n). To illustrate, given (P_n), let

$N_1, N_2, \cdots$ be i.i.d. with distribution (P_n), and let the set $\mathbb{N}_n := \{1, 2, \cdots, n\}$ be partitioned by the random equivalence relation $i \sim j$ iff $N_i = N_j$. Let K_n be the number of classes in this random partition of $\mathbb{N}_n$. That is to say K_n is the number of distinct values observed in the sequence $(N_1, \cdots, N_n)$. It is elementary and well known that

$$P(K_n = 1) = \int_0^1 v^{n-1} F(dv). \tag{29}$$

As noted by Engen [5], $E(K_n)$ is a polynomial moment of F of degree $n-1$:

$$E(K_n) = \int_0^1 \left[\frac{1-(1-v)^n}{v} \right] F(dv) \tag{30}$$

In particular, the distribution of both K_2 and K_3 can be recovered from moments the structural distribution. In applications to species sampling as in [5], K_n would represent the number of distinct species found in a sample of size n from a population in which distinct species are present with random frequencies (P_n). In population genetics, K_n might represent the number of distinct alleles found in a sample of size n from a population with allelic frequencies (P_n).

Call a distribution F on $(0, 1]$ a *structural distribution* if F is the structural distribution of some random discrete distribution (P_n). The set of all structural distributions coincides with the set of all possible distributions of P_1 for a sequence (P_n) that is ISBP.

Proposition 15 *A necessary condition for a distribution F on $(0, 1]$ to be a structural distribution is the following: for every $0 < a \leq 1$ (or, equivalently, for every $0 < a \leq 1/2$),*

$$\int_0^a (1-x) F(dx) \geq \int_{1-a}^1 (1-x) F(dx) \tag{31}$$

Proof. According to Theorem 4, if P_1 is the first atom sampled from a random discrete distribution, and (X, Y) are r.v.s with

$$P(X \in dx, Y \in dy) = \frac{P(P_1 \in dx, P_2 \in dy)\overline{x}}{E(\overline{P}_1)}, \tag{32}$$

then (X, Y) is exchangeable with $X \geq 0$, $Y \geq 0$, $X + Y \leq 1$. (Assume that $P(P_1 < 1) > 0$, so $E(\overline{P}_1) > 0$.) Now it is easy to see that given some distribution for X with $0 \leq X \leq 1$, there exists an exchangeable joint distribution for (X, Y) with $X + Y \leq 1$, and the given X-marginal, iff X is stochastically smaller than $1 - X$, i.e.

$$P(X \leq a) \geq P(1 - X \leq a) \tag{33}$$

for every $0 \leq a \leq 1$ (or, equivalently, every $0 \leq a \leq 1/2$). Now from (32)

$$\frac{P(X \leq a)}{P(\overline{X} \leq a)} = \frac{E\overline{P}_1 1(P_1 \leq a)}{E\overline{P}_1 1(\overline{P}_1 \leq a)},$$

so (33) reduces to condition (31) on the distribution F of P_1.

Problem. Find a necessary and sufficient condition for F to be a structural distribution.

According to Pitman-Yor [19], a sufficient condition is that F admits a density $f(u) = F(du)/du$ such that $(1 - u)f(u)$ is a decreasing function of u for $0 < u < 1$. But this condition is far from necessary, just as condition (31) is far from sufficient. Some easy remarks are that the set of structural distributions is convex, not closed in the weak topology, but closed under the multiplicative analog of convolution. In particular, if f and g are two structural densities then so is h defined by

$$h(u) := \int_0^1 y^{-1} f(u/y) g(y) dy \quad (0 < u < 1) \tag{34}$$

To verify this closure property, define the *product* of two random discrete distributions (P_n) and (Q_n) defined on the same probability space to be the random discrete distribution (R_n) obtained by ranking the collection of products $\{P_m Q_n, m = 1, 2, \cdots, n = 1, 2, \cdots\}$. If $\tilde{P}_1$ is a size-biased pick from (P_n) and $\tilde{Q}_1$ is a size-biased pick from (Q_n), and $(\tilde{P}_1, P_1, P_2, \cdots)$ and $(\tilde{Q}_1, Q_1, Q_2, \cdots)$ are independent, it is easily seen that $\tilde{P}_1 \tilde{Q}_1$ is a size-biased pick from the ranked product of (P_n) and (Q_n).

6 Concluding Remarks.

It is natural to look for some sequential procedure for generating the most general distribution $(P_1, P_2, \cdots)$ that is ISBP, but it is not at all evident how to describe such a procedure.

The constraint on (P_1, P_2) that G_2 is symmetric is necessary but not sufficient for the existence of $(P_1, P_2, \cdots)$ that is ISBP with a prescribed law for (P_1, P_2). The point is that the later constraints, e.g. that G_3 is symmetric, impose further restrictions on the allowed joint laws for (P_1, P_2). A trivial example which illustrates this point is the degenerate prescription $P_1 = P_2 = c$, that is $W_1 = c$, $W_2 = c/(1-c)$ for a constant c. This (W_1, W_2) is an acceptable pair for any $c \leq 1/2$. However, it is obvious without calculation that there exists a $(P_1, P_2, \cdots)$ that is an ISBP extension of this prescription iff $c = 1/n$ for some integer $n \geq 2$. The question "what laws of $(P_1, \cdots, P_n)$ are the start of a law for $(P_1, P_2, P_3, \cdots)$ that is ISBP?" was raised in the previous section for $n = 1$, and can be posed also for $n > 1$. But the present paper offers no satisfactory answer without the side condition of independence of the ratios $W_1, W_2, \cdots$.

References

[1] D.J. Aldous. Exchangeability and related topics. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XII, Springer Lecture Notes in Mathematics, Vol. 1117.* Springer-Verlag, 1985.

[2] R.J. Connor and J.E. Mosimann. Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Am. Statist. Assoc.*, 64:194–206, 1969.

[3] P. Donnelly. The heaps process, libraries and size biased permutations. *J. Appl. Prob.*, 28:322–335, 1991.

[4] P. Donnelly and P. Joyce. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Processes and their Applications*, 31:89 – 103, 1989.

[5] S. Engen. *Stochastic Abundance Models with Emphasis on Biological Communities and Species Diversity.* Chapman and Hall Ltd., 1978.

[6] W.J. Ewens. Population genetics theory - the past and the future. In S. Lessard, editor, *Mathematical and Statistical Problems in Evolution.* University of Montreal Press, Montreal, 1988.

[7] J. Fabius. Asymptotic behavior of Bayes estimates. *Math. Statist.*, 35:846–856, 1964.

[8] D. Freedman. On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.*, 34:1386–1403, 1963.

[9] F. M. Hoppe. Size-biased filtering of Poisson-Dirichlet samples with an application to partition structures in genetics. *Journal of Applied Probability*, 23:1008 – 1012, 1986.

[10] T. Ignatov. On a constant arising in the theory of symmetric groups and on Poisson-Dirichlet measures. *Theory Probab. Appl.*, 27:136–147, 1982.

[11] J. F. Kingman. Random discrete distributions. *J. Roy. Statist. Soc. B*, 37:1–22, 1975.

[12] J. F. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

[13] J. F. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982.

[14] J. W. McCloskey. A model for the distribution of individuals by species in an environment. Ph. D. thesis, Michigan State University, 1965.

[15] G. P. Patil and C. Taillie. Diversity as a concept and its implications for random communities. *Bull. Int. Stat. Inst.*, XLVII:497 – 515, 1977.

[16] M. Perman, J. Pitman, and M. Yor. Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92:21–39, 1992.

[17] J. Pitman. Partially exchangeable random partitions. Technical Report 343, Dept. Statistics, U.C. Berkeley, 1992.

[18] J. Pitman. The two-parameter generalization of Ewens' random partition structure. Technical Report 345, Dept. Statistics, U.C. Berkeley, 1992.

[19] J. Pitman and M. Yor. Random discrete distributions derived from self-similar random sets. Preprint, 1994.

[20] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Preprint, 1994.

[21] G. A. Watterson. The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.*, 13:639–651, 1976.

[22] S.L. Zabell. Predicting the unpredictable. *Synthese*, 90:205–232, 1992.

[23] S.L. Zabell. The continuum of inductive methods revisited. Preprint, 1994.