Random Discrete Distributions Invariant Under Size-biased Permutation

Jim Pitman*

Technical Report No. 344
June 1992

*Research supported by N.S.F. Grant MCS91-07531

Department of Statistics
University of California
Berkeley, California 94720

Random Discrete Distributions Invariant Under Size-biased Permutation

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

June 23, 1992

1 Introduction

Let π be a random discrete probability measure. Given π, let $X_1, X_2, \ldots$ 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 $(X_1, X_2, \ldots)$ of the nth distinct value observed in the sequence of exchangeable random variables $(X_1, X_2, \ldots)$. Think of the atoms of π as representing the frequencies with which various species are present in an infinite population. Then $X_1, X_2, \ldots$ 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. Given that the random discrete distribution π has atoms of sizes say

$$\pi_1 \geq \pi_2 \geq \ldots > 0, \text{ with } \sum \pi_i = 1,$$

the (P_n) are a *size-biased random permutation (SBP)* of these atoms: $P_1 = \pi_i$ with probability π_i; given $P_1 = \pi_k$, and $P_1 < 1$, $P_2 = \pi_j$ for $j \neq k$ with probability $\pi_j/(1 - P_1)$, and so on:

*Research supported by N.S.F. Grant MCS91-07531

given $P_i = \pi_{j_i}$ for $1 \leq i \leq n$, with $\sum_{i=1}^n P_i < 1$, $P_{n+1} = \pi_j$ with probability $\pi_j/(1 - \sum_{i=1}^n P_i)$ for $j \notin \{j_1, \ldots j_n\}$.

The most general possible distribution for the sequence (P_n) is one that is *invariant under size-biased permutation* (ISBP). See Patil and Taillie [12], Donnelly and Joyce [4], Donnelly [3], Ewens [6], Zabell [16] for background, and motivation for the study of random discrete distributions that are ISBP. One reason for interest in such distributions is that due to the representation theory of random partitions of Kingman[9, 10], these are the only possible joint distributions for a proper distribution (P_n) derived from an exchangeable random partition of the positive integers as the long run relative frquencies of classes ordered by their least elements.

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 [11]). *Suppose that*

$$P_n = \bar{W}_1 \bar{W}_2 \ldots \bar{W}_{n-1} W_n, \ n \geq 1 \tag{1}$$

where $W_1, W_2, \ldots$ are i.i.d with values in $[0,1]$, and $\bar{W}_i = 1 - W_i$. 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

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

where $\bar{x} = 1 - x$, 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 $\{1, 2, \ldots\}$ defined by

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

where $X_1 > X_2 > \ldots$ 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. Perman, Pitman and Yor[13] generalize this argument to find the joint law of (P_n) derived this way from a Poisson process on $(0, \infty)$ with arbitrary 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 a Poisson process on $(0, \infty)$. But McCloskey's argument establishes the more general result formulated above, as it is based only the following property shared by every (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. A multidimensional form of this property appears in Theorem 4 below, which gives a symmetry on the joint distribution of $(P_1, \dots, P_n)$ 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[12], and solved in this paper: for what independent non-identically sequences (W_n) does the formula

$$P_n = \bar{W}, \dots, \bar{W}_{n-1} W_n \tag{4}$$

define a random discrete distribution (P_n) that is ISBP? The model (4) 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 $\{1, 2, \dots\}$ 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 $(P_1, P_2, \dots)$ 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 $(P_1, \dots, P_m)$ is the *Dirichlet* $(a_1, \dots, a_m)$ *distribution*, that is to say the joint distribution of

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

where $Y_1, \dots, Y_m$ are independent gamma random variables with common scale parameter and shape parameters $a_1, \dots, a_m$, and $\sum = \sum_{i=1}^{m} Y_i$. Patil and Taillie[12] noted that if $(\pi_1, \ldots, \pi_m)$ has Dirichlet $(\beta, \ldots, \beta)$ distribution, then the SBP of $(\pi_1, \ldots, \pi_m)$ follows a RAM that is generalized Dirichlet with parameters $a_n \equiv 1 + \beta$, $b_n = m\beta - n\beta$, $n = 1, \ldots, m$.

2 Results

According to the following theorem and its corollary, apart from some rather trivial examples and modifications, the only RAM's which 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[13]. 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 $(P_1, P_2, \ldots)$ be such that $P_n > 0$, $\sum_n P_n = 1$, $P_1 < 1$, and $P_n = \bar{W}_1 \ldots \bar{W}_{n-1} W_n$ 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 $n = 1, 2, \ldots$, for some $0 \leq \alpha < 1$, $\theta > -\alpha$.*

or (ii) *$\{n : P_n > 0\} = \{1, \ldots, m\}$ 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 $n = 1, \ldots, m$.*

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

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

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

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 [13]. For $0 < \alpha < 1$, $\theta = 0$, they showed that (P_n) as in case (i) of Theorem 2 appears as the SBP of $(X_n/\sum)$ 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. In case $0 < \alpha < 1$, for arbitrary $\theta > -\alpha$, they showed that this sequence (P_n) remains ISBP, and that the W_n become independent with beta$(1-\alpha, \theta + n\alpha)$ distributions, if the underlying probability measure is changed by a density factor proportional to $\Sigma^{-\theta}$. (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 $W_1, W_2, \ldots$ induce (P_n) that is ISBP, then so do $W_{k+1}, W_{k+2}, \ldots$, for any $k \geq 1$, given $\bar{W}_1\bar{W}_2\ldots\bar{W}_k > 0$).

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 rather trivial possibility that $P(P_1 = 1) > 0$:

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

The above results show the ISBP condition imposes severe restrictions in the joint law of (P_n). These restrictions seem at first hard to understand, as the definition of ISBP 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 $(P_1, P_2, \ldots)$ 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 $(P_1, P_2, \ldots)$ be a sequence of random variables satisfying the almost sure constraints $P_i \geq 0$ for $i \geq 2$, and $\sum_{i=1}^{n} P_i \leq 1$, $n = 1, 2, \ldots$.*

Let G_k denote the measure on $\mathbb{R}^k$ whose density with respect to the joint probability distribution of $(P_1, \ldots, P_k)$ at $(p_1, \ldots, p_k)$ is $\prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} p_j)$:

$$G_k(dp_1, \ldots, dp_k) = P(P_1 \in dp_1, \ldots, P_k \in dp_k) \prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} p_j) \tag{5}$$

The following statements are equivalent:

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

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

$$(n_1, \ldots, n_k) \to E\left[\prod_{i=1}^{k} P_i^{n_i - 1} \prod_{i=1}^{k-1}(1 - \sum_{j=1}^{i} P_j)\right], \tag{6}$$

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

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 $P_1, \ldots, P_k$. (For the condition $\sum_i P_i = 1$ a.s. imposes no constraint on the law of $P_1, \ldots, P_k$ 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 interpretations of the measure G_k in (5). And it shows that the function in (6) 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 such the P_n 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 [15, 14] for analysis of the two-parameter family of random partitions corresponding to Theorem 2.

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

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

where $\pi = (\pi(1), \pi(2), \ldots)$ is a random discrete probability distribution distributed like $(P_1, P_2, \ldots)$, 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

$$Z_k = 1\{X_1, \ldots, X_k \text{ are all distinct}\}. \tag{7}$$

Then for each k the random vector

$$(P_1, P_2, \ldots, P_k)Z_k = (\pi(X_1), \pi(X_2), \ldots, \pi(X_k))Z_k \tag{8}$$

clearly has an exchangeable joint distribution. But since

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

the distribution of the exchangeable random vector (8) and measure G_k defined by (5) 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 [10] representation of exchangeable random partitions of $\mathbb{N} := \{1, 2, \ldots\}$. See Pitman [14] for generalizations of the argument to the setting of partially exchangeable random partitions of $\mathbb{N}$.

Define a sequence Π_n of partitions of $\mathbb{N}_n := \{1, \ldots, n\}$ 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 $R_i := 1 - P_1 - \ldots - P_i$. By constriction, 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:

$$P(\Pi_n = \{A_i\}_1^k) = E(\prod_{i=1}^{k} P_i^{n_i - 1} \prod_{i=1}^{k-1} R_i). \tag{9}$$

This probability depends on $A_1, \ldots, A_k$ only through their sizes $n_1, \ldots, n_k$, and hypothesis (iii) amounts to symmetry of the right hand side of (9) as a function of $(n_1, \ldots, n_k)$, 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 $(X_1, X_2, \ldots)$, 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 $(X_1, X_2, \ldots)$. The assumption $P_1 > 0$ implies π is discrete a.s., hence that $\sum_i P_i = 1$ a.s., and that (P_k) is a size-biased presentation of the atoms of π. Thus (P_k) is ISBP. $\square$

An immediate consequence of Theorem 4 is

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

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

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

$$f_n(p_1, \ldots, p_n) \prod_{j=1}^{n-1} (1 - \sum_{i=1}^{j} p_i) \tag{10}$$

is a symmetric function of $(p_1, \ldots, p_n)$. Then $(P_1, P_2, \ldots)$ 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 $W_1, \ldots, W_n$ such that $P_n = \bar{W}_1 \ldots \bar{W}_{n-1} W_n$, namely:

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

and

$$g_n(p_1, \frac{p_2}{1-p_1}, \ldots, \frac{p_n}{1-p_1-\ldots-p_{n-1}}) \tag{11}$$

is a symmetric function of $p_1, \ldots, p_n$. 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 $0 < W_i \leq 1$ a.s., and for $P_1 = W_1$, $P_2 = \bar{W}_1 W_2$, the joint law of (P_1, P_2) is such that the distribution G_2 in (5) is symmetric.

In view of (6), a joint distribution for (W_1, W_2) is acceptable iff $0 < W_i \leq 1$ a.s., and

$$m(r,s) := E[W_1^r \bar{W}_1^{s+1} W_2^s] \tag{12}$$

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

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

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

(ii) *The law of the pair (W_1, W_2) is acceptable, and for each $n = 1, 2, \ldots$, on the event $P_1 + \cdots + P_n < 1$ there is a version of the conditional law of the pair (W_{n+1}, W_{n+2}) given $(P_1, \ldots, P_n)$ that is acceptable, and that depends exchangeably on $(P_1, \ldots, P_n)$.*

Proof. Condition (ii) is a substitute for the condition that G_k in (5) is symmetric for all k. Equivalence of these two conditions is easily established by induction, using moments. Then the present corollary follows at once from Theorem 4.

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

4 Residual Allocation Models

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

$$P_n = \bar{W}_1 \dots \bar{W}_{n-1} W_n$$

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

Lemma 8 *Let $(W_1, W_2 \dots)$ be a sequence of independent random variables with $0 < W_i \leq 1$ a.s., and let $P_n = \bar{W}_1 \dots \bar{W}_{n-1} W_n$, $n = 1, 2, \dots$. 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 (12), all possible pairs of distributions for random variables W and Z with $0 < W \leq 1$, $0 < Z \leq 1$, such that

$$m(r, s) := E(W^r \bar{W}^{s+1}) E(Z^s) \tag{13}$$

is symmetric function of non-negative in integers r and s. From (11) 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:

$$f(p) g(\frac{q}{\bar{p}}) = f(q) g(\frac{p}{\bar{q}}) \tag{14}$$

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 (14) becomes

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

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

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

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 9 *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 9 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 8 and 9. 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 (13) that identifies an acceptable independent pair (W, Z), and Lemma 8.

Lemma 10 *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 $s = 1, 2, \ldots$*

$$E(W^r\bar{W}^{s+1})E(Z^s) = E(W^s\bar{W}^{r+1})E(Z^r) \tag{15}$$

(ii) *either*

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

$$F(dw) = \bar{w}P(W \in dw)/E(\bar{W}), \tag{16}$$

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* (15) *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 (16), so for any bounded function g,

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

Then each condition above becomes a corresponding condition on Y, as in the statement of Lemma 11 below. Thus Lemma 10 is a consequence of Lemma 11. □

Lemma 11 *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 $s = 1, 2, \ldots$,*

$$E(Y^r \bar{Y}^s)E(Z^s) = E(Y^s \bar{Y}^r)E(Z^r), \tag{17}$$

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

(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 (17) holds for all positive real r and s.*

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

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

Note first that (17) for $r = 0$, $s = k$ implies

$$\nu_k = E(Y^k)/E(\bar{Y}^k), \tag{18}$$

In particular

$$\nu_1 = \mu_1/(1 - \mu_1) \tag{19}$$

and in general

$$\nu_k = \mu_k/p_k(\mu_1, \ldots, \mu_k) \tag{20}$$

for some polynomial p_k. Thus the distribution of Z is determined by that of Y. In particular if Y is symmetric about $1/2$, that is $E(Y^k) = E(\bar{Y}^k)$ for all k, iff $Z = 1$ a.s.. And $Y = c$ a.s. for some constant c iff $\nu_k = (c/\bar{c})^k$ for all k, that is $Z = c/\bar{c}$, 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

$$1 > \nu_1 > \nu_2 > \ldots > 0 \tag{21}$$

Now (17) for $r = 1$ states that for $s = 2, 3 \ldots$

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

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

$$\mu_{s+1} = \frac{r_s(\mu_1, \mu_2, \ldots, \mu_s)}{\nu_s(-1)^s + \nu_1} \tag{22}$$

for some polynomial r_s, where the denominator never vanishes, because $0 < \nu_s < \nu_1$ by (21). Now (20) combined with (22) shows that

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

for some function f_s, hence by induction

$$\mu_k = g_k(\mu_1, \mu_2); \quad \nu_k = h_k(\mu_1, \mu_2), \quad k = 3, 4, \ldots \tag{23}$$

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*

$$0 < \mu_1^2 < \mu_2 < \mu_1 < 1/2, \tag{24}$$

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 (17) holds.*

To complete the argument, note that the a priori constraints (24) 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

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

where e.g. $[a]_r = a(a+1)\dots(a+r-1) = \Gamma(a+r)/\Gamma(a)$. But then $EY = \mu$, $EY^2 = \mu_2$, and it is obvious that (17) holds for Z with the beta$(a, b-a)$ distribution which makes

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

Remark. 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 (17) 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 (17) holds.

5 Concuding Remarks

It is natural to look for some sequential procedure for generating the most general distribution $(P_1, P_2, \ldots)$ 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 suficient for the existence of $(P_1, P_2, \ldots)$ 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 $(P_1, P_2, \ldots)$ that is an ISBP extension of this prescription iff $c = 1/n$ for some integer $n \geq 2$.

Thus to the question "what laws for P_1, or for (P_1, P_2), are the start of a law for $(P_1, P_2, P_3, \ldots)$ that is ISBP?", the present paper offers no satisfactory answer without the side condition of independence of the ratios $W_1, W_2, \ldots$,

A Necessary Condition on P_1.

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

$$P(X \in dx, Y \in dy) = \frac{P(P_1 \in dx, P_2 \in dy)\bar{x}}{E(\bar{P}_1)}, \tag{25}$$

then (X, Y) is exchangeable with $X \geq 0$, $Y \geq 0$, $X + Y \leq 1$. (Assume that $P(P_1 < 1) > 0$, so $E(\bar{P}_1) > 0$.) 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.

$$P(X \leq a) \geq P(1 - X \leq a) \tag{26}$$

for every $0 \leq a \leq 1$ (or, equivalently, every $0 \leq a \leq 1/2$). Now from (25)

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

so the constraint on the distribution of P_1 is

$$E\bar{P}_1 1(P_1 \leq a) \geq E(\bar{P}_1 1(\bar{P}_1 \leq a)) \tag{27}$$

for every $0 \leq a \leq 1/2$, or $0 \leq a \leq 1$.

Problem. Find a necessary and sufficient condition.

References

[1] D.J. Aldous. Exchangeability and related topics. In P.L. Hennequin, editor, *Ecole 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 and size biased permutations. *Preprint*, 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] J. F. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

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

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

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

[13] M. Perman, J. Pitman, and M. Yor. Size-biased sampling of Poisson point processes and excursions. To appear in *Probability and Related Fields*, 1992.

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

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

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