CONTINUITY AND WEAK CONVERGENCE OF RANKED AND SIZE-BIASED PERMUTATIONS ON THE INFINITE SIMPLEX

Peter DONNELLY
School of Mathematical Sciences, Queen Mary College, University of London, Mile End Road, London E1 4NS, UK

Paul JOYCE
Department of Statistics, University of Washington, Seattle, Washington 98195, USA

Received 28 April 1988

Ranked and size-biased permutations are particular functions on the set of probability measures on the simplex. They represent two recently studied schemes for relabelling groups in certain stochastic models, and are of particular interest in describing the limiting behaviour of such models. We prove that the ranked permutations of a sequence of measures converge if and only if the size-biased permutations converge, and give conditions under which weak convergence of measures guarantees weak convergence of both permutations. Applications include a proof of the fact that the GEM distribution is the size-biased permutation of the Poisson-Dirichlet and a new proof of the fact that when labelled in a particular way, normalized cycle lengths in a random permutation converge to the GEM distribution. These techniques also allow some problems concerned with the random splitting of an interval to be related to known results in other fields.

1. Introduction

Many applications of probability involve the description of collections of individuals or objects which fall naturally into distinct groups. We have in mind settings which range from mathematical biology, where the individuals might be animals and the groups species or families, or genetics, where individuals are grouped on the basis of genetic type, to random permutations, where the objects on which a particular permutation acts may be grouped into cycles. In some contexts there will be a natural way of labelling the groups, in others, the only salient feature is that the groups may be distinguished. It is often of interest to study the limiting behaviour of such models as the number of individuals increases. This brings the labelling problem sharply into focus—under some labelling schemes the model may exhibit sensible limiting behaviour while for others it may be degenerate; for some, limits may be tractable, while for others they may be intractable.

Our purpose here is to consider two particular relabelling schemes which have received attention in the recent literature. The first of these, the ranked permutation,
is deterministic; it simply labels groups in order of decreasing size. The second, the so-called size-biased permutation, is random. Informally, it samples from the collection of individuals; each sampled individual belongs to a particular group, and (ignoring repetitions) the groups are assigned labels in the order in which they arise through this sampling process. The terminology has the same connotation as it does in sampling theory, namely that larger groups are more likely to be labelled first. Fundamentally our concern is with the interrelation between these two permutations and their relationship with any labelling which is intrinsic to the problem, asking specifically whether these permutations commute with each other or with limiting operations. Alternatively, they may be viewed as functions on probability measures, and in this terminology (with the topology of weak convergence) our interest is in the continuity of these functions.

The remainder of this section is devoted to the establishment of an appropriate framework in which to set the discussion, and the formal definition of the functions involved. In Section 2 we state and prove the main results. These are, effectively, that size-biased permutations converge if and only if ranked permutations converge, and that under certain conditions each of these functions is continuous. The final section is devoted to a number of applications of these results. We remark at the outset that in parts of the literature some of these results seem already to have been assumed. Part of our aim here is to provide a rigorous basis for such applications. In this spirit we give a proof of the fact that the size-biased permutation of the Poisson-Dirichlet distribution is the GEM distribution and show how our results may be used to provide insight into, and easy proofs of, several existing results, notably in the fields of random splitting and random permutations.

Denote by \( \Delta \) the infinite simplex,

\[
\Delta = \left\{ x = (x_1, x_2, \ldots) : x_i \geq 0 \text{ for all } i, \text{ and } \sum_{i=1}^{\infty} x_i = 1 \right\}.
\]

It will be convenient to embed \( \Delta \) in the larger set

\[
\bar{\Delta} = \left\{ x = (x_1, x_2, \ldots) : x_i \geq 0 \text{ for all } i, \text{ and } \sum_{i=1}^{\infty} x_i < 1 \right\},
\]

which we topologise as a closed (and hence compact) subset of the product space \([0, 1]^\mathbb{N}\). We endow \( \bar{\Delta} \) with the Borel \( \sigma \)-algebra, \( \mathcal{B} \), and throughout the sequel all measures will be defined on \((\bar{\Delta}, \mathcal{B})\). For many purposes, and for ours here, any finite collection of "exchangeable" objects may be well described (that is without losing essential information) as a point \( x = (x_1, x_2, \ldots) \) in \( \Delta \). The interpretation of such a description is that (with a particular labelling of groups) a proportion \( x_1 \) of the objects are in the first group, \( x_2 \) in the second, and so on. This kind of description is also appropriate for a hypothetically infinite population. If the collection (or equivalently the assignment of groups, and group sizes) is random then it may be described by a measure \( \mu \) on \( \bar{\Delta} \). In the finite case such a measure will of course satisfy \( \mu(\Delta) = 1 \). As the completion of \( \Delta \), the set \( \bar{\Delta} \) is the natural setting in which
to study limiting behaviour. We note that \( \Delta \) may also be viewed as \( \mathcal{P}(\mathbb{N}) \), the set of probability measures on \( \mathbb{N} = \{1, 2, 3, \ldots\} \), and given the topology of weak convergence. Although we omit the details, this approach is exactly equivalent to the one we adopt. The anologue of \( \bar{\Delta} \) is \( \mathcal{P}(\bar{\mathbb{N}}) \) where \( \bar{\mathbb{N}} \) denotes the one point compactification of \( \mathbb{N} \).

Two further subsets of \( \bar{\Delta} \) which are destined to play a central role are

\[
\nabla = \{ x \in \bar{\Delta}: x_1 \geq x_2 \geq \cdots \}
\]

and its closure

\[
\bar{\nabla} = \{ x \in \bar{\Delta}: x_1 \geq x_2 \geq \cdots \}.
\]

We now define the ranking function, \( \rho \). For \( x = (x_1, x_2, \ldots) \in \bar{\Delta} \) define the permutation \( R_x \) by

\[
R_x(1) = \min \{ i: x_i \geq x_j, j = 1, 2, \ldots \},
\]

\[
R_x(k) = \min \{ i: x_i \geq x_j, j \in \mathbb{N} \setminus \{ R_x(1), R_x(2), \ldots, R_x(k-1) \} \},
\]

for \( k = 2, 3, \ldots \). Now define \( \rho : \bar{\Delta} \to \Delta \) by

\[
\rho((x_1, x_2, \ldots)) = (x_{R_x(1)}, x_{R_x(2)}, \ldots).
\]

Note that the function \( \rho \) rearranges the components of \( x \) into non-increasing order. It is straightforward to check that \( \rho \) is a (Borel) measurable function. (One approach is to note that \( \rho \) is the pointwise limit of the functions \( \rho_1, \rho_2, \ldots \), where \( \rho_N \) rearranges the first \( N \) components of \( x \) into nonincreasing order but leaves other components unchanged, and to observe that each of the functions \( \rho_N \) is continuous and hence measurable.) Thus \( \rho \) induces a mapping on \( \mathcal{P}(\bar{\Delta}) \), the set of probability measures on \( \bar{\Delta} \), in a natural way: for \( \mu \in \bar{\Delta} \) and Borel sets \( A \),

\[
\rho(\mu)(A) = \mu(\rho^{-1}(A)) = \mu \{ x: \rho(x) \in A \}.
\]

(The standard notation for \( \rho(\mu) \) is \( \mu \rho^{-1} \). Our choice of notation allows convenient comparison with size-biased permutations and highlights the viewpoint that this is a function on \( \mathcal{P}(\bar{\Delta}) \).) We will refer to \( \rho(\mu) \) as the ranked permutation of \( \mu \). Note that we use the same notation for the functions defined at (2) and (3). The exact meaning should be clear from the context.

Let \( S_N \) denote the set of permutations on \( \mathbb{N} \). For \( x = (x_1, x_2, \ldots) \in \bar{\Delta} \), we define the size-biased permutation of \( x \), \( \sigma_x \in \mathcal{P}(\bar{\Delta}) \), by first constructing a random element \( \pi_x \), of \( S_N \cup \{ A \} \), where \( A \notin S_N \) is arbitrary. Put

\[
P(\pi_x = A) = 1 - P(\pi_x \in S_N) = 1 - \sum_{i=1}^{\infty} x_i.
\]

Now if \( x \) has an infinite number of non-zero components, conditional on \( \pi_x \in S_N \),
its distribution is uniquely determined by the requirements

\[ P(\pi_x(1) = i) = \left( \sum_{i=1}^{\infty} x_i \right)^{-1} x_i, \quad i = 1, 2, \ldots, \]

\[ P(\pi_x(k) = i_k \mid \pi_x(1) = i_1, \pi_x(2) = i_2, \ldots, \pi_x(k-1) = i_{k-1}) = \frac{x_{i_k}}{(1 - x_{i_1} - x_{i_2} - \cdots - x_{i_{k-1}})} \]

where \( i_1, i_2, \ldots, i_k \in \mathbb{N} \) are distinct. If \( x \) has only a finite number, \( n > 0 \), of non-zero components, \( x_{i_1}, \ldots, x_{i_k} \) say, put

\[ m_1 = \min\{i \in \mathbb{N} : x_i = 0\}, \]

and for \( k = 2, 3, \ldots, \)

\[ m_k = \min\{i \in \mathbb{N} : i > m_{k-1}, x_i = 0\}, \]

in which case \( \pi_x \) has distribution

\[ P(\pi_x(1) = j_1, \pi_x(2) = j_2, \ldots, \pi_x(n) = j_n, \pi_x(n+1) = m_1, \pi_x(n+2) = m_2, \ldots \mid \pi_x \in S_N) \]

\[ = x_{j_1} \frac{x_{j_2}}{1 - x_{j_1}} \frac{x_{j_3}}{1 - x_{j_1} - x_{j_2}} \cdots \frac{x_{j_n}}{1 - x_{j_1} - \cdots - x_{j_{n-1}}} \left( \sum_{i=1}^{\infty} x_i \right)^{-1}. \]

Finally, for Borel sets \( A \), define the measure \( \sigma_x \) by

\[ \sigma_x(A) = P(\pi_x = A) 1_A((0, 0, \ldots)) + P(\pi_x \in S_N \text{ and } (x_{\pi(1)}, x_{\pi(2)}, \ldots) \in A). \]

(Where \( 1_A \) denotes the indicator of the set \( A \).) Informally, if \( x \in \Delta \), \( \sigma_x \) is the distribution of a random rearrangement of the components of \( x \). In this rearrangement, the \( i \)th component will appear first with probability \( x_i \); conditional on this, the \( j \)th component will appear second with probability \( x_j/(1 - x_i) \), and so on. (Care is needed if there are only a finite number of non-zero components; our convention is that the zero components are "tacked on the end" in their original order.) If \( x \in \Delta \setminus A \), the collection of all such rearrangements has probability less than one; the deficient mass is concentrated on the sequence \((0, 0, \ldots)\). In applications, one is usually concerned only with the size-biased permutation of points in \( \Delta \). In our current framework, some extension to \( \Delta \) is convenient; this particular choice has the advantage that all of our results carry over appropriately.

The family of measures \( \{\sigma_x : x \in \Delta\} \), induces a function \( \sigma : \mathcal{P}(\Delta) \to \mathcal{P}(\Delta) \) in the obvious way: for Borel sets \( A \), put

\[ \sigma(\mu)(A) = \int_{\Delta} \sigma_x(A) \mu(dx). \]

We call \( \sigma(\mu) \) the size-biased permutation of \( \mu \). The interpretation is that we first choose a point \( x \in \Delta \) according to \( \mu \) and then (randomly, as above) size-bias; \( \sigma(\mu) \) is the distribution of the resulting point. We omit the details, but note that it is not too difficult to prove that the function \( \sigma \) is measurable, from which it follows that \( \sigma \) is also measurable.
We conclude the section by making two observations which are almost immediate consequences of these definitions. Firstly, for any $\mu \in \mathcal{P}(\Delta)$,

$$\sigma(\rho(\mu)) = \sigma(\mu)$$

(4)

and secondly, if $\mu(\Delta) = 1$,

$$\rho(\sigma(\mu)) = \rho(\mu).$$

(5)

Throughout the sequel we shall use the symbol $\Rightarrow$ to denote weak convergence, and denote by $\tilde{C}(\Delta)$ the set of bounded continuous functions from $\Delta$ into $\mathbb{R}$.

2. Main results

In this section we state and prove the continuity results. Informally these state that ranked permutations of measures converge if and only if their size-biased permutations converge (Theorems 1 and 2) and that if a sequence of measures converges to a measure concentrated on $\Delta$, then so do their ranked, and hence size-biased, permutations.

**Theorem 1.** Suppose $\{\mu_n\}$ is a sequence of measures on $\Delta$ with $\mu_n(\bar{\Delta}) = 1$ for each $n$, and further suppose that $\mu_n \Rightarrow \mu$ for some $\mu$. Then $\mu(\bar{\Delta}) = 1$ and $\sigma(\mu_n) \Rightarrow \sigma(\mu)$.

**Proof.** Since $\bar{\Delta}$ is closed, the Portmanteau Theorem ensures that $\mu(\bar{\Delta}) = 1$.

Now let $f: \Delta \rightarrow \mathbb{R}$ have the form

$$f(x) = x_{i_1}^{\nu_1} x_{i_2}^{\nu_2} \cdots x_{i_k}^{\nu_k},$$

$\nu_i \in \mathbb{N}, i = 1, 2, \ldots, k$. As a consequence of the Stone-Weierstrass Theorem (and the trivial observation that $\int_{\Delta} 1 \, d\sigma(\mu_n) \rightarrow \int_{\Delta} 1 \, d\sigma(\mu)$) it is sufficient to prove that, as $n \rightarrow \infty$,

$$\int_{\Delta} f \, d\sigma(\mu_n) \rightarrow \int_{\Delta} f \, d\sigma(\mu)$$

for all such $f$.

By definition,

$$\int_{\Delta} f \, d\sigma(\mu_n) = \int_{\Delta} \int_{\Delta} f \, d\sigma_x \mu_n(dx).$$

For $x \in \bar{\Delta}$ put

$$f^*(x) = \int_{\Delta} f \, d\sigma_x.$$  

(6)
Note that $f^*$ is bounded. We will prove that as a function from $\mathcal{V}$ into $\mathbb{R}$, $f^*$ is also continuous. Then, since $\mu_n(\mathcal{V}) = \mu(\mathcal{V}) = 1$, and $\mu_n \Rightarrow \mu$,

$$
\int_{\mathcal{V}} f \, d\mu_n = \int_{\mathcal{V}} f^* \, d\mu_n \rightarrow \int_{\mathcal{V}} f^* \, d\mu = \int_{\mathcal{V}} f \, d\sigma(\mu),
$$
as $n \to \infty$, as required.

Put

$$
I = \{ \pi = (\pi_1, \ldots, \pi_k): \pi_i \in \mathbb{N}, \pi_i \neq \pi_j \text{ for } i \neq j \} (\forall)
$$

and

$$
T_N = \{ \pi = (\pi_1, \ldots, \pi_k): \pi \in T \text{ and } \pi_i \leq N \text{ for } i = 1, 2, \ldots, k \}.
$$

Then, by definition,

$$
f^*(x) = \int_{\mathcal{V}} f \, d\sigma_x
$$

say. Let $f^*_N(x) = \sum_{\pi \in T_N} h_\pi(x)$. The functions $f^*_N, N = 1, 2, \ldots$, depend in a continuous way on only a finite number of coordinates, and so are continuous. We now show that, as $N \to \infty$,

$$
f^*_N(x) \Rightarrow f^*(x),
$$

uniformly, for $x \in \mathcal{V}$. It will then follow that $f^*$ is continuous on $\mathcal{V}$. Now

$$
|f^*(x) - f^*_N(x)| = \left| \sum_{\pi \in T} h_\pi(x) - \sum_{\pi \in T_N} h_\pi(x) \right| = \sum_{\pi \in T \setminus T_N} h_\pi(x).
$$

Observe that if $x \in \mathcal{V}$, $x_j \leq j^{-1}$ (where $x_j$ denotes the $j$th coordinate of $x$). Further, if $x \in \mathcal{V}$ and $(\pi_1, \ldots, \pi_k) \in T \setminus T_N$, then for some $i = 1, \ldots, k$, $\pi_i > N$, so $x_{\pi_i} < N^{-1}$, and $x_{\pi_1} x_{\pi_2} \cdots x_{\pi_k} \leq N^{-1}$. Thus, for $x \in \mathcal{V}$,

$$
\sum_{\pi \in T \setminus T_N} h_\pi(x) \leq \frac{1}{N} \sum_{\pi \in T \setminus T_N} P(\pi_1, \ldots, \pi_k) = \frac{1}{N} \to 0,
$$

uniformly in $x$, as $N \to \infty$, and the result follows. \qed

Our next result states that convergence of size-biased permutations guarantees that ranked permutations converge, thus providing a converse to Theorem 1.

**Theorem 2.** Suppose $\{\mu_n\}$ is a sequence of measures on $\mathcal{D}$ with $\sigma(\mu_n) \Rightarrow \mu$ as $n \to \infty$, for some $\mu$. Then $\rho(\mu_n) \Rightarrow \rho(\mu)$ and $\mu$ is the size-biased permutation of some measure on $\mathcal{D}$. 

Proof. First note that if $\rho(\mu_n) \Rightarrow \rho(\mu)$ then as a consequence of (4) and Theorem 1, $\sigma(\mu_n) = \sigma(\rho(\mu)) \Rightarrow \sigma(\rho(\mu))$. But by hypothesis $\sigma(\mu_n) \Rightarrow \mu$ so $\mu = \sigma(\rho(\mu))$ which proves the second assertion.

Now, let $f \in \tilde{C}(\Delta)$. By definition, and (4) again,

$$\int_{\Delta} f \, d\sigma(\mu_n) = \int_{\Delta} \left( \int_{\Delta} f \, d\sigma_\lambda \right) \mu_n(dx)$$

$$= \int_{\Delta} \left( \int_{\Delta} f \, d\sigma_\lambda \right) \rho(\mu_n)(dx) = \int_{\varepsilon} f^* \, d(\rho(\mu_n))$$

with $f^*$ defined by (6). Similarly $\int_{\Delta} f \, d\sigma(\mu) = \int_{\varepsilon} f^* \, d(\rho(\mu))$. Thus, by hypothesis, whenever $f \in \tilde{C}(\Delta)$,

$$\int_{\varepsilon} f^* \, d(\rho(\mu_n)) \rightarrow \int_{\varepsilon} f^* \, d(\rho(\mu))$$

as $n \rightarrow \infty$. If $f$ has the particular form

$$f(x) = x_1^{a_1}x_2^{a_2} \cdots x_k^{a_k}(1 - x_1)(1 - x_1 - x_2) \cdots (1 - x_1 - \cdots - x_{k-1}),$$

for $a_1, \ldots, a_k \in \mathbb{N} \cup \{0\}$, then

$$f^*(x) = \sum_{\pi \in \Gamma} x_{\pi_1}^{a_1+1}x_{\pi_2}^{a_2+1} \cdots x_{\pi_k}^{a_k+1},$$

where $\Gamma$ is defined at (7). Kingman (1977) uses a Stone-Weierstrass argument to prove that, together with the constant function, a subset of the functions of the form (9) generates an algebra which is dense (in the sup norm) in the set $\tilde{C}(\tilde{\mathcal{V}})$, the set of bounded continuous functions from $\tilde{\mathcal{V}}$ into $\mathbb{R}$, and hence that as measures on $\tilde{\mathcal{V}}$, $\rho(\mu_n) \Rightarrow \rho(\mu)$. The desired result follows from the Continuous Mapping Theorem. □

We now prove that if a sequence of measures converges weakly to a measure $\mu$ which is concentrated on $\Delta$, then their ranked and size-biased permutations converge to the ranked, respectively size-biased, permutation of the limit measure.

Theorem 3. Suppose $\{\mu_n\}$ is a sequence of measures on $\tilde{\Delta}$ with $\mu_n \Rightarrow \mu$, and further suppose $\mu(\Delta) = 1$. Then $\rho(\mu_n) \Rightarrow \rho(\mu)$ and $\sigma(\mu_n) \Rightarrow \sigma(\mu)$.

Proof. The following proposition shows that (as a function from $\Delta$ to $\tilde{\Delta}$) $\rho$ is continuous on $\Delta$. The Continuous Mapping Theorem then guarantees that $\rho(\mu_n) \Rightarrow \rho(\mu)$. Theorem 1 and the observation, (4), that $\sigma(\rho(\mu)) = \sigma(\mu)$ completes the proof. □

Proposition 4. Suppose $\{x^{(n)}\}$ is a sequence in $\tilde{\Delta}$ and $x^{(n)} \rightarrow x \in \Delta$. Then $\rho(x^{(n)}) \rightarrow \rho(x)$.

Proof. Denote $\rho(x^{(n)})$ and $\rho(x)$ by $(x_1^{(n)}, x_2^{(n)}, \ldots)$ and $(x_1, x_2, \ldots)$ respectively, and denote by $R_n$ and $R$ respectively the ranked permutations $R(x^{(n)})$ and $R_x$ defined at (1). We must show that for fixed but arbitrary $k$, $x_{i(j)}^{(n)} \rightarrow x_{i(j)}$, $i = 1, 2, \ldots, k$. 


We first prove that $x_{(1)}(n) \to x_{(1)}$. Note that if $x_{(1)}(n) \to 0$ along a subsequence $\{n_j\}_{j=1}^\infty$, it follows that $x_{i}(n) \to 0$ for each $i$, which contradicts the fact that $x(n) \to x \in \Delta$. Thus there is a $\delta > 0$ with $x_{(1)}(n) > \delta$ for infinitely many $n_j$. As a consequence the sequence $\{R_{n}(1)\}_{n=1}^\infty$ is bounded above, for otherwise, for fixed $m$ and infinitely many $n > m$,

$$\sum_{i=1}^{m} x_{i}(n) < 1 - x_{(1)}(n)$$

which will be less than 1 $\delta$ infinitely often. This would imply that

$$\sum_{i=1}^{\infty} x_{i} = \lim_{m \to \infty} \sum_{i=1}^{m} x_{i} = \lim_{m \to \infty} \sum_{n=1}^{m} x_{i}(n) = 1 - \delta$$

which contradicts $x \in \Delta$. Thus there exist $N$ and $n_0$ with

$$R_{n}(1) \leq N \quad \text{for all } n \geq n_0.$$  

Hence for $n \gg n_0$,

$$x_{(1)}^{(n)} = \max\{x_{1}^{(n)}, \ldots, x_{N}^{(n)}\} \to \max\{x_{1}, \ldots, x_{N}\} = x' \quad \text{say.} \quad (10)$$

It remains to prove that $x' = x_{(1)}$. Again suppose otherwise, so that for some $k > N$, $x_{k} > x'$. Since $x_{k}(n) \to x_{k}$, we must have $x_{k}^{(n)} > x' + (x_{k} - x')/2$ for all large $n$. But by definition, $x_{i}(n) \equiv x_{k}(n)$, so

$$\lim_{n \to \infty} \inf x_{i}(n) \geq \lim_{n \to \infty} \inf x_{k}(n) > x', \quad (10)$$

which contradicts (10). Thus $x' = x_{(1)}$, and from (10), $x_{(1)}^{(n)} \to x_{(1)}$ as $n \to \infty$.

Now suppose $x_{(i)}(n) \to x_{(i)}$, $i = 1, 2, \ldots, j - 1$. We will show $x_{(j)}(n) \to x_{(j)}$. There are three cases to consider.

Case 1: $x_{(j-1)} = 0$. If $x_{(j-1)} = 0$, then by definition $x_{(j)} = 0$. By assumption $x_{(i)}(n) \to x_{(j-1)} = 0$. Also $0 \leq x_{(j)}^{(n)} \leq x_{(j-1)}^{(n)}$, so as $n \to \infty$, $x_{(j)}^{(n)} \to 0 = x_{(j)}$, as required.

Case 2: $x_{(j-1)} > 0$ and $\{R_{n}(j)\}_{n=1}^{\infty}$ bounded above. Thus there exists a $\delta > 0$ and an $n_0$ with

$$x_{(1)}^{(n)} \geq x_{(2)}^{(n)} \geq \cdots \geq x_{(j-1)}^{(n)} > \delta \quad \text{for all } n > n_0.$$  

An argument similar to the one above shows that $x \in \Delta$ implies that each of the sequences $\{R_{n}(i)\}_{n=1}^{\infty}$, $i = 1, 2, \ldots, j - 1$, is bounded above. This and our assumption guarantees the existence of an $N$ with $R_{n}(i) \leq N$ for $i = 1, 2, \ldots, j$ and all $n$. Thus, arguing as above

$$x_{(i)}^{(n)} \to x_{(i)}^{'}, \quad i = 1, \ldots, j,$$

as $n \to \infty$, where $x_{1}^{'}, x_{2}^{'}, \ldots, x_{j}^{'}$ are the $j$ largest members (allowing repetitions in the obvious way) of the sequence $(x_{1}, \ldots, x_{N})$. The existence of a $k > N$ with $x_{k} > x_{j}^{'}$ will give a contradiction as before, so $x_{j}^{'} = x_{(j)}$ and in particular $x_{(j)}^{(n)} \to x_{(j)}$ as required.
Case 3: \( x_{(j-1)} > 0 \) and \( \{R_n(j)\}_{n=1}^{\infty} \) not bounded above. An argument similar to the one above shows that \( x \in \Delta \) and \( \{R_n(j)\}_{n=1}^{\infty} \) not bounded above implies that

\[
x_i^{(n)}(j) \to 0
\]

as \( n \to \infty \). By definition,

\[
x_i^{(n)}(j) \equiv \min\{x_{R(1)}^{(n)}, \ldots, x_{R(j)}^{(n)}\} \to \min\{x_R(1), \ldots, x_R(j)\}
\]

as \( n \to \infty \), since by assumption \( x^{(n)} \to x \). But by definition of \( R \),

\[
x_{(j)} = x_{R(j)} = \min\{x_{R(1)}, \ldots, x_{R(j)}\}.
\]

It follows from (11), (12), and (13) that \( x_{(j)} = 0 \), so \( x_i^{(n)}(j) \to x_{(j)} \) as required.

We remark that in fact the ranking function is discontinuous for all \( x \in \bar{\Delta} \setminus \Delta \). For such an \( x = (x_1, x_2, \ldots) \) put

\[
x_0 = 1 - \sum_{i=1}^{\infty} x_i > 0
\]

and define a sequence \( \{x_i^{(n)}\}_{n=1}^{\infty} \) by

\[
x_i^{(n)} = \begin{cases} x_i, & i = 1, 2, \ldots, n-1, \\ x_0, & i = n, \\ x_{i-1}, & i = n+1, n+2, \ldots. \end{cases}
\]

It is clear that \( x^{(n)} \to x \) and straightforward to verify that \( \rho(x^{(n)}) \neq \rho(x) \).

Proposition 4, the continuity of the ranking function on \( \Delta \), may be of some interest in its own right. For example it leads to a direct proof of one of the results of Vershik and Shmidt (1977), the Corollary to Theorem 2.4.

We close this section by remarking that no more general results concerning convergence are true in this context. It is possible to have sequences \( \{\mu_n\} \) which do not converge to anything, for which \( \{\rho(\mu_n)\} \) and \( \{\sigma(\mu_n)\} \) both converge. As an example, take

\[
x_i^{(n)} = \begin{cases} \left(\frac{1}{3}, \frac{2}{3}, 0, 0, \ldots\right), & n \text{ even} \\ \left(\frac{1}{3}, \frac{1}{3}, 0, 0, \ldots\right), & n \text{ odd} \end{cases}
\]

and take \( \mu_n \) to be a unit mass at \( x^{(n)} \). Note that insisting that limit points of the sequence \( \{\mu_n\} \) place mass one on \( \Delta \) is of no avail. Furthermore, if \( \mu_n \Rightarrow \mu, \mu(\Delta) < 1 \), it may be that \( \rho(\mu_n) \) converges to something other than \( \rho(\mu) \) or that \( \rho(\mu_n) \) does not converge at all. An example of the first case is given by considering unit masses at the elements of the sequence \( \{x^{(n)}\} \) defined at (14). For the second, let \( \mu_n \) be a unit mass at \( x = (x_1, x_2, \ldots) \in \bar{\Delta} \setminus \Delta \) for \( n \) odd and a unit mass at \( x^{(n)} \) defined by (14) for \( n \) even.
3. Applications

Ranked permutations arise naturally in applications as one way of defining sensible limits of probabilistic models. If \( X^{(n)} \) is some random variable of interest with distribution \( \mu_n \), then \( \mu_n \) may have a degenerate limit (for example a point mass at zero) while many quantities of interest exhibit sensible limiting behaviour. It may be appropriate to look for a suitable relabelling of the components which does have a nondegenerate limit, and the relabelling induced by ranking is a natural choice. The point is that if \( \rho(\mu_n) \to \mu \) say, and \( f \) is a function which is invariant under relabelling of components and bounded and continuous as a function from \( \mathbb{V} \) into \( \mathbb{R} \), for which \( E(f(X^{(n)})) = \int_\mathbb{V} f \, d\mu_n \) is of interest, then

\[
\int_\mathbb{V} f \, d\mu_n = \int_\mathbb{V} f \, d\rho(\mu_n) \to \int_\mathbb{V} f \, d\mu.
\]

This explains the existence of sensible limiting behaviour for some functionals of \( \mu_n \).

This is exactly the rationale which lead to Kingman's (1975) definition of the Poisson-Dirichlet distribution. If \( (X_1, \ldots, X_n) \) has a symmetric Dirichlet distribution with parameters \( n \) and \( \alpha \), and \( \lambda_n \) is the distribution of the random point \( (X_1, X_2, \ldots, X_n, 0, 0, \ldots) \in \Delta \) then as \( n \to \infty \) with \( n\alpha \to \theta \in (0, \infty) \), although \( \lambda_n \) has a degenerate limit at \( (0, 0, \ldots) \), there is a measure \( \lambda \), called the Poisson-Dirichlet distribution with parameter \( \theta \), such that

\[
\rho(\lambda_n) \to \lambda
\]

as \( n \to \infty \) and \( n\alpha \to \theta \). In fact the Poisson-Dirichlet distribution is defined by (15). This distribution is of considerable importance in applied probability, arising for example in the fields of population genetics, mathematical ecology, the theory of self regulating filing systems, and (in the case \( \theta = 1 \)) in studying the asymptotics of random permutations. Although several equivalent representations exist for the distribution (see for example Kingman 1978), it is notoriously intractable.

One of the morals of the results of Section 2 is that whenever ranked permutations are used in this way to specify limiting behaviour, size-biased permutation will also converge to something interesting (specifically the size-biased permutations of the limit of the ranked permutations). For quantities of interest (i.e. functionals of the limiting measure) which are invariant under relabelling, use of either the limit of the ranked permutations or its size biased permutation will give the same answer. This gives alternative representations of the limiting model and the point is that one is at liberty to choose whichever of these is the more convenient for particular purposes.

The Poisson-Dirichlet distribution is a case in point. While for most purposes it is intractable, its size-biased permutation, known as the GEM distribution, has a very simple structure which readily lends itself to calculation. Let \( Z_1, Z_2, \ldots \) be independent and identically distributed random variables with probability density...
The GEM distribution with parameter \( \theta, 0 < \theta < \infty \), which we denote by \( \gamma \), is defined for Borel sets \( A \subseteq \Delta \) by
\[
\gamma(A) = P\left( (Z_1, (1 - Z_1)Z_2, (1 - Z_1)(1 - Z_2)Z_3, \ldots) \in A \right).
\]
Note that \( \gamma(\Delta) = 1 \). The GEM distribution is the size-biased permutation of the Poisson-Dirichlet, and the Poisson-Dirichlet is the ranked permutation of GEM. This result, due originally to Patil and Taillie (1977), is used continually throughout the literature. To our knowledge however, no published proof exists. We give one here as an application of Theorems 1 and 2.

**Theorem 5.** Let \( \lambda \) denote the Poisson-Dirichlet distribution with parameter \( \theta \), and let \( \gamma \) denote the GEM distribution with parameter \( \theta \). Then \( \sigma(\lambda) = \gamma \) and \( \rho(\gamma) = \lambda \).

**Proof.** Note that the second observation follows immediately from the first and (5), since then \( \rho(\gamma) = \rho(\sigma(\lambda)) = \rho(\lambda) = \lambda \).

Now suppose that \((X_1, \ldots, X_n)\) have a symmetric Dirichlet distribution with parameters \( n \) and \( \alpha \) and define
\[
\lambda_n(A) = P((X_1, \ldots, X_n, 0, 0, \ldots) \in A)
\]
for Borel sets \( A \). It is known (Patil and Taillie, 1977), and in any case a straightforward calculation, to show that the size-biased permutation of \( \lambda_n \) has the representation
\[
\sigma(\lambda_n)(A) = P((\tilde{X}_1^n, \ldots, \tilde{X}_n^n, 0, 0, \ldots) \in A)
\]
with
\[
\tilde{X}_1^n = \tilde{Z}_1, \\
\tilde{X}_2^n = (1 - \tilde{Z}_1)\tilde{Z}_2, \\
\tilde{X}_n^n = (1 - \tilde{Z}_1)(1 - \tilde{Z}_2)\cdots(1 - \tilde{Z}_{n-1})\tilde{Z}_n,
\]
where \( \tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_n \) are independent, and for \( j = 1, 2, \ldots, n, \tilde{Z}_j \) has probability density function
\[
f_j(z) = \frac{\Gamma((n-j+1)\alpha+1)}{\Gamma(\alpha+1)\Gamma((n-j)\alpha)} z^\alpha (1-z)^{(n-j)\alpha-1}, \quad 0 < z < 1.
\]
Now fix \( k \). It is evident from the fact that \( f_j(z) \to f(z) \) as \( n \to \infty, n\alpha \to \theta \), that under this limiting regime,
\[
(\tilde{X}_1^n, \tilde{X}_2^n, \ldots, \tilde{X}_k^n) \Rightarrow (Z_1, (1 - Z_1)Z_2, \ldots, (1 - Z_1)(1 - Z_2)\cdots(1 - Z_{k-1})Z_k).
\]
It then follows from the fact that the functions which depend on only finitely many coordinates are dense in \( C(\Delta) \) that
\[
\sigma(\lambda_n) \Rightarrow \gamma,
\]
the GEM distribution with parameter \( \theta \), as \( n \to \infty, n\alpha \to \theta \). By definition \( \rho(\lambda_n) \Rightarrow \lambda \), so \( \sigma(\lambda) = \gamma \), as required, by Theorem 1. \( \square \)
For applications of this Theorem, see for example Patil and Taillie (1977), Donnelly (1986), Hoppe (1986), or Ewens (1988). As a further application, consider the problem of recursive splitting of an interval; for background, see for example Lloyd and Williams (1988). After splitting, the unit interval is broken into pieces of length

\[ X_1 = Y_1, \]
\[ X_2 = (1 - Y_1)Y_2, \]
\[ X_3 = (1 - Y_1)(1 - Y_2)Y_3, \]

and so on, for \( Y_1, Y_2, \ldots \) independent and identically distributed with common distribution \( F \) on \([0, 1]\). (Note that Lloyd and Williams use \( X_i \) in place of our \( (1 - Y_i) \).) Among other things, interest centres on the length of the longest piece, \( X_{(1)} \), say. If the \( Y_i \) have the particular Beta distribution (16), then by definition the point \((X_1, X_2, \ldots)\) has a GEM distribution and, as a consequence of Theorem 5, its rearrangement into non-increasing order has a Poisson–Dirichlet distribution. Many results follow immediately from this and properties of the Poisson–Dirichlet. In particular the distribution of \( X_{(1)} \) is that of the first component of a Poisson–Dirichlet, about which much is known; see for example Ewens (1988) or Watterson (1976).

Now use the other assertion of Theorem 5 and construct the lengths \( X_1, X_2, \ldots \) at (17) as the size-biased permutation of a point \((X_{(1)}, X_{(2)}, \ldots)\) which has a Poisson–Dirichlet distribution. Condition on \( X_{(1)}, X_{(2)}, \ldots \), and note that with probability one these will be distinct. The probability that the first piece (i.e. the one of length \( X_1 \)) is the longest is exactly the probability that the first component of \((X_{(1)}, X_{(2)}, \ldots)\) is the one chosen first in performing the operation of size-biasing, but this is just \( X_{(1)} \). Thus

\[ P(\text{first piece is longest}) = E(X_{(1)}). \] (18)

Lloyd and Williams (1988) consider in detail the case where \( F \) is uniform. They observe that (18) is an "intriguing theoretical consequence" of one of their analytical results and ask for a heuristic argument. The uniform case corresponds to \( \theta = 1 \) in (16). The above discussion shows that (18) is in fact a consequence of the result (Theorem 5) that the random variables (17) arise as the size-biased permutation of their ranked permutation. Furthermore, the result (18) obtains for a class of distributions more general than the uniform and for any of these distributions, exactly the same argument gives

\[ P(\text{first piece is kth longest}) = E(X_{(k)}). \]

The right hand side of (18) cannot be evaluated in closed form. It has been evaluated numerically several times in different contexts. See for example Golomb (1964), Shepp and Lloyd (1966), Lloyd and Williams (1988) for the case \( \theta = 1 \), and Watterson and Guess (1977) for several values of \( \theta \). We remark that the connection (provided
by Theorem 5) between the GEM distribution and the Poisson-Dirichlet allows the translation of many of the problems and results of Lloyd and Williams (1988) into much studied problems in other contexts, notably population genetics and random permutations.

Another property of the GEM distribution which has proved crucial in applications is its invariance under size-biased permutations. This result is originally due to Engen (1975), who used direct (moment) calculations. As Patil and Taillie (1977) remark, it is also immediate from Theorem 5 and (4), since then

\[ \sigma(\gamma) = \sigma(\rho(\gamma)) = \sigma(\lambda) = \gamma. \]

Another sense in which the continuity results of Section 2 are useful is in allowing symmetry arguments in discrete models to carry over to limiting distributions. We illustrate by considering random permutations, but note that the technique is much more widely applicable.

Let \( S_n \) be the set of permutations of the set \( \{1, 2, \ldots, n\} \), and let \( \eta_n \) be a randomly chosen member. (That is, \( \eta_n \) is equally likely to be any of the \( n! \) permutations in \( S_n \).) Among other things, interest centres on the asymptotic distribution of the (normalized) cycle lengths of \( \eta_n \). Specifically, suppose \( \eta_n \) has \( l \) cycles, and denote the cycle lengths by \( C_1, C_2, \ldots, C_l \). There is no single natural way of labelling the cycles. One approach is to label them by decreasing length, so that \( C_{c_1} \) is the length of the longest cycle, \( C_{c_2} \) the length of the second longest, \ldots, and so on (with some convention for ties). In this way we can associate with \( \eta_n \) a point \( X^{(n)} \in \Delta \) by

\[ X^{(n)} = (n^{-1}C_{(1)}, n^{-1}C_{(2)}, \ldots, n^{-1}C_{(l)}, 0, 0, \ldots). \]

An alternative labelling of cycles is to denote by \( B_1 \) the length of the cycle containing the element 1, by \( B_2 \) the length of the cycle containing the smallest integer not in the cycle containing 1, and so on. This associates another point \( Y^{(n)} \in \Delta \) with \( \eta_n \):

\[ Y^{(n)} = (n^{-1}B_1, n^{-1}B_2, \ldots, n^{-1}B_l, 0, 0, \ldots). \]

For our purposes, the following observation is crucial.

**Lemma 6.** Conditional on \( X^{(n)} = x = (x_1, x_2, \ldots) \) the distribution of \( Y^{(n)} \) is \( \sigma_x \), the size-biased permutation of \( x \).

**Proof.** Imagine constructing \( \eta_n \) by first choosing the cycle lengths \( C_{(1)}, C_{(2)}, \ldots, C_{(l)} \), from the appropriate distribution, and then assigning integers to these cycles. Conditional on \( C_{(1)}, C_{(2)}, \ldots, C_{(l)} \), let \( I_1, I_2, \ldots, I_l \in \{1, 2, \ldots, n\} \) denote the labels of the cycles to which 1, 2, \ldots, \( n \) respectively are assigned. It follows from the symmetry of the problem (or the uniformity of the distribution of \( \eta_n \)) that the collection of random variables \( I_1, \ldots, I_l \) is exchangeable. Thus

\[ P(\text{integer 1 belongs to a particular cycle of length } C_{(k)} | C_{(1)}, \ldots, C_{(l)}) = C_{(k)} / n. \]

Furthermore if the cycle containing 1 is deleted, the resulting permutation is a uniform random permutation on the remaining \( n - B_1 \) objects. Successive repetitions of these arguments give the desired result. \( \Box \)
It follows that if $\mu_n$ and $\nu_n$ denote the distributions of $X^{(n)}$ and $Y^{(n)}$ respectively, then

$$\nu_n = \sigma(\mu_n) \quad \text{and} \quad \mu_n = \rho(\nu_n).$$

(19)

Theorems 1 and 2 now guarantee that if either $\mu_n$ or $\nu_n$ converges as $n \to \infty$, then so will the other. Aldous (1985, Section 10) uses an elegant argument founded on the observation in the penultimate sentence of the proof of Lemma 6, and a characterization result of Kingman's, to prove that as $n \to \infty$,

$$\mu_n \Rightarrow \lambda,$$

(20)

the Poisson-Dirichlet distribution with parameter $\theta = 1$. As a consequence of (19) and Theorems 1 and 5, we have immediately that as $n \to \infty$

$$\nu_n \Rightarrow \gamma,$$

(21)

the GEM distribution with parameter $\gamma = 1$. These results are not new. In this case the distributions are tractable (or symmetric) enough to allow an explicit derivation of (21), see for example Vershik and Shmidt (1977). (We remark in passing that (19) and Theorem 2 also provide a proof of their Corollary to Theorem 2.4, by a less direct route than that discussed at the end of Section 2.) The novelty of the current approach is that no limiting calculations are needed: Aldous' proof of (20) uses a characterization result which forces the limiting distribution to be Poisson-Dirichlet (an easy calculation identifies the appropriate value of $\theta$) and convergence to GEM follows from symmetry and our continuity results. Note that the symmetry argument leading to Lemma 6 and hence (19) is inherently finite; a continuity result (in this case Theorem 1) is needed to force the limiting distributions to inherit this symmetry.

We remark again that our discussion of random permutations should be seen as an application of a general idea. In fact in this case it is possible to derive (19) by direct calculation. (The distributions $\nu_n$ and $\mu_n$ are given in, for example, Joyce and Tavaré (1987).) Informally, for any exchangeable combinatorial "object", the assignment of "individuals" to "groups" (integers to cycles in the case of permutations) will be exchangeable, and it will follow that any relabelling of groups on the basis of the individuals they contain will effectively be a size-biased relabelling of the ordered labelling. If the limiting behaviour is known in one case, the other is immediate. Aldous (1985) treats a number of other settings (notably random functions) to which these same arguments apply.

For related applications of the continuity results of Section 2 the interested reader is referred to Ewens (1988) and Donnelly (1989).

Acknowledgements

The first author was supported in part by NSF grant DMS 86-08857.
References


F.M. Hoppe, Size biased filtering of Poisson-Dirichlet samples with an application to partition structures in genetics, J. Appl. Prob. 23 (1986) 1008–1012.


