*Bernoulli* **21**(4), 2015, 2484–2512
DOI: 10.3150/14-BEJ652

# Size-biased permutation of a finite sequence with independent and identically distributed terms

JIM PITMAN[1] and NGOC M. TRAN[2]

[1]*Department of Statistics, UC Berkeley, CA 94720, USA. E-mail: pitman@stat.berkeley.edu*
[2]*Department of Mathematics, UT Austin, TX 78712, USA. E-mail: ntran@math.utexas.edu*

This paper focuses on the size-biased permutation of $n$ independent and identically distributed (i.i.d.) positive random variables. This is a finite dimensional analogue of the size-biased permutation of ranked jumps of a subordinator studied in Perman–Pitman–Yor (PPY) [*Probab. Theory Related Fields* **92** (1992) 21–39], as well as a special form of *induced order statistics* [*Bull. Inst. Internat. Statist.* **45** (1973) 295–300; *Ann. Statist.* **2** (1974) 1034–1039]. This intersection grants us different tools for deriving distributional properties. Their comparisons lead to new results, as well as simpler proofs of existing ones. Our main contribution, Theorem 25 in Section 6, describes the asymptotic distribution of the last few terms in a finite i.i.d. size-biased permutation via a Poisson coupling with its few smallest order statistics.

*Keywords:* induced order statistics; Kingman paint box; Poisson–Dirichlet; size-biased permutation; subordinator

## 1. Introduction

Let $x = (x(1), x(2), \dots)$ be a positive sequence with finite sum $t = \sum_{i=1}^{\infty} x(i)$. Its size-biased permutation (s.b.p.) is the same sequence presented in a random order $(x(\sigma_1), x(\sigma_2), \dots)$, where $\mathbb{P}(\sigma_1 = i) = \frac{x(i)}{t}$, and for $k$ distinct indices $i_1, \dots, i_k$,

$$\mathbb{P}(\sigma_k = i_k | \sigma_1 = i_1, \dots, \sigma_{k-1} = i_{k-1}) = \frac{x(i_k)}{t - (x(i_1) + \cdots + x(i_{k-1}))}. \tag{1}$$

An index $i$ with bigger ‘size’ $x(i)$ tends to appear earlier in the permutation, hence the name size-biased. Size-biased permutation of a random sequence is defined by conditioning on the sequence values.

One of the earliest occurrences of size-biased permutation is in social choice theory. For fixed sequence length $n$, the goal is to infer the $x(i)$ given multiple observations from the random permutation defined by (1). Here the $x(i)$ are the relative scores or desirabilities of the candidates, and (1) models the distribution of their rankings in an election. Now known as the Plackett–Luce model, it has wide applications [7,29,33].

Around the same time, biologists in population genetics were interested in inferring the distribution of alleles in a population through sampling. In these applications, $x(i)$ is the abundance and $x(i)/t$ is the relative abundance of the $i$th species [12]. Size-biased permutation models the outcome of successive sampling, where one samples without replacement from the population and records the abundance of newly discovered species in the order that they appear. To account for the occurrence of new types of alleles through mutation and migration, they considered random abundance sequences and did not assume an upper limit to the number of possible types. Patil and Taillie [25] coined the term *size-biased random permutation* to describe i.i.d. sampling from a random discrete distribution. The earliest work along this vein is perhaps that of McCloskey [23], who obtained results on the size-biased permutation of ranked jumps in a certain Poisson point process (p.p.p.). The distribution of this ranked sequence is now known as the Poisson–Dirichlet distribution $\mathrm{PD}(0, \theta)$. The distribution of its size-biased permutation is the $\mathrm{GEM}(0, \theta)$ distribution. See Section 4.3 for their definitions. This work was later generalized by Perman, Pitman and Yor [26], who studied size-biased permutation of ranked jumps of a subordinator; see Section 4.

Size-biased permutation of finite sequences appear naturally through the study of partition structure. Kingman [20,21] initiated this theory to explain the Ewens sampling formula, which gives the joint distribution of $n$ independent size-biased picks from a $\mathrm{PD}(0, \theta)$ distribution. The theory of partition structure, in particular that of exchangeable partitions, is closely related to the GEM and Poisson–Dirichlet distributions. We briefly mention related results in Section 4.4. For details, see [28], Sections 2, 3.

This paper focuses on *finite i.i.d. size-biased permutation*, that is, the size-biased permutation of $n$ independent and identically distributed (i.i.d.) random variables from some distribution $F$ on $(0, \infty)$. Our setting is a finite dimensional analogue of the size-biased permutation of ranked jumps of a subordinator studied in [26], as well as a special form of *induced order statistics* [3,8]. This intersection grants us different tools for deriving distributional properties. Their comparison lead to new results, as well as simpler proofs of existing ones. By considering size-biased permutation of i.i.d. triangular arrays, we derive convergence in distribution of the remaining $u$ fraction in a successive sampling scheme. This provides alternative proofs to similar statements in the successive sampling literature. Our main contribution, Theorem 25 in Section 6, describes the asymptotic distribution of the last few terms in a finite i.i.d. size-biased permutation via a Poisson coupling with its few smallest order statistics.

## 1.1. Organization

We derive joint and marginal distribution of finite i.i.d. size-biased permutation in Section 2 through a Markov chain, and re-derive them in Section 3 using induced order statistics. Section 4 connects our setting and its infinite version of [26]. As the sequence length tends to infinity, we derive asymptotics of the last $u$ fraction of finite i.i.d. size-biased permutation in Section 5, and that of the first few terms in Section 6.

## 1.2. Notation

We shall write $\mathrm{gamma}(a, \lambda)$ for a Gamma distribution whose density at $x$ is $\lambda^a x^{a-1} e^{-\lambda x} / \Gamma(a)$ for $x > 0$, and $\mathrm{beta}(a, b)$ for the Beta distribution whose density at $x$ is $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

for $x \in (0,1)$. For an ordered sequence, not necessarily order statistics, $(Y_n(k), k = 1, \dots, n)$, let $Y_n^{\mathrm{rev}}(k) = Y_n(n-k+1)$ be the same sequence presented in reverse. For order statistics, we write $Y^{\uparrow}$ for the increasing sequence, and $Y^{\downarrow}$ for the decreasing sequence. Throughout this paper, unless otherwise indicated, we use $X_n = (X_n(1), \dots, X_n(n))$ for the underlying i.i.d. sequence with $n$ terms, and $(X_n[1], \dots, X_n[n])$ for its size-biased permutation. To avoid having to list out the terms, it is also sometimes convenient to write $X_n^* = (X_n^*(1), \dots, X_n^*(n))$ for the size-biased permutation of $X_n$.

## 2. Markov property and stick-breaking

Assume that $F$ has density $\nu_1$. Let $T_{n-k} = X_n[k+1] + \cdots + X_n[n]$ denote the sum of the last $n-k$ terms in an i.i.d. size-biased permutation of length $n$. We first derive joint distribution of the first $k$ terms $X_n[1], \dots, X_n[k]$.

**Proposition 1 (Barouch–Kaufman [1]).** *For $1 \le k \le n$, let $\nu_k$ be the density of $S_k$, the sum of $k$ i.i.d. random variables with distribution $F$. Then*

$$\begin{aligned} &\mathbb{P}\big(X_n[1] \in \mathrm{d}x_1, \dots, X_n[k] \in \mathrm{d}x_k\big) \\ &\quad = \frac{n!}{(n-k)!}\left(\prod_{j=1}^{k} x_j \nu_1(x_j)\,\mathrm{d}x_j\right) \int_0^\infty \nu_{n-k}(s) \prod_{j=1}^{k} (x_j + \cdots + x_k + s)^{-1}\,\mathrm{d}s \qquad (2) \\ &\quad = \frac{n!}{(n-k)!}\left(\prod_{j=1}^{k} x_j \nu_1(x_j)\,\mathrm{d}x_j\right) \mathbb{E}\left(\prod_{j=1}^{k} \frac{1}{x_j + \cdots + x_k + S_{n-k}}\right). \qquad (3) \end{aligned}$$

**Proof.** Let $\sigma$ denote the random permutation on $n$ letters defined by size-biased permutation as in (1). Then there are $\frac{n!}{(n-k)!}$ distinct possible values for $(\sigma_1, \dots, \sigma_k)$. By exchangeability of the underlying i.i.d. random variables $X_n(1), \dots, X_n(n)$, it is sufficient to consider $\sigma_1 = 1, \dots, \sigma_k = k$. Note that

$$\mathbb{P}\left(\big(X_n(1), \dots, X_n(k)\big) \in \mathrm{d}x_1 \cdots \mathrm{d}x_k, \sum_{j=k+1}^{n} X_n(j) \in \mathrm{d}s\right) = \nu_{n-k}(s)\,\mathrm{d}s \prod_{j=1}^{k} \nu_1(x_j)\,\mathrm{d}x_j.$$

Thus, restricted to $\sigma_1 = 1, \dots, \sigma_k = k$, the probability of observing $(X_n[1], \dots, X_n[k]) \in \mathrm{d}x_1 \cdots \mathrm{d}x_k$ and $T_{n-k} \in \mathrm{d}s$ is precisely

$$\frac{x_1}{x_1 + \cdots + x_k + s} \frac{x_2}{x_2 + \cdots + x_k + s} \cdots \frac{x_k}{x_k + s} \nu_{n-k}(s) \left(\prod_{j=1}^{k} \nu_1(x_j)\,\mathrm{d}x_j\right) \mathrm{d}s.$$

By summing over $\frac{n!}{(n-k)!}$ possible values for $(\sigma_1, \dots, \sigma_k)$, and integrating out the sum $T_{n-k}$, we arrive at (2). Equation (3) follows by rewriting. $\square$

Note that $X_n[k] = T_{n-k+1} - T_{n-k}$ for $k = 1, \dots, n-1$. Thus we can rewrite (2) in terms of the joint law of $(T_n, T_{n-1}, \dots, T_{n-k})$:

$$\begin{aligned}&\mathbb{P}(T_n \in \mathrm{d}t_0, \dots, T_{n-k} \in \mathrm{d}t_k) \\ &\quad = \frac{n!}{(n-k)!}\left(\prod_{i=0}^{k-1} \frac{t_i - t_{i+1}}{t_i} \nu_1(t_i - t_{i+1})\right)\nu_{n-k}(t_k)\,\mathrm{d}t_0 \cdots \mathrm{d}t_k.\end{aligned} \tag{4}$$

Rearranging (4) yields the following result, which appeared as an exercise in [6], Section 2.3.

**Corollary 2 (Chaumont–Yor [6]).** *The sequence $(T_n, T_{n-1}, \dots, T_1)$ is an inhomogeneous Markov chain with transition probability*

$$\mathbb{P}(T_{n-k} \in \mathrm{d}s | T_{n-k+1} = t) = (n-k+1)\frac{t-s}{t}\nu_1(t-s)\frac{\nu_{n-k}(s)}{\nu_{n-k+1}(t)}\,\mathrm{d}s, \tag{5}$$

*for $k = 1, \dots, n-1$. Together with $T_n \stackrel{d}{=} S_n$, equation (5) specifies the joint law in (4), and vice versa.*

## 2.1. The stick-breaking representation

An equivalent way to state (5) is that for $k \geq 1$, conditioned on $T_{n-k+1} = t$, $X_n[k]$ is distributed as the first size-biased pick out of $n-k+1$ i.i.d. random variables conditioned to have sum $S_{n-k+1} = t$. This provides a recursive way to generate a finite i.i.d. size-biased permutation: first generate $T_n$ (which is distributed as $S_n$). Conditioned on the value of $T_n$, generate $T_{n-1}$ via (5), let $X_n[1]$ be the difference. Now conditioned on the value of $T_{n-1}$, generate $T_{n-2}$ via (5), let $X_n[2]$ be the difference, and so on. Let us explore this recursion from a different angle by considering the ratio $W_{n,k} := \frac{X_n[k]}{T_{n-k+1}}$ and its complement, $\overline{W}_{n,k} = 1 - W_{n,k} = \frac{T_{n-k}}{T_{n-k+1}}$. For $k \geq 2$, note that

$$\frac{X_n[k]}{T_n} = \frac{X_n[k]}{T_{n-k+1}}\frac{T_{n-k+1}}{T_{n-k+2}} \cdots \frac{T_{n-1}}{T_n} = W_{n,k}\prod_{i=1}^{k-1}\overline{W}_{n,i}. \tag{6}$$

The variables $\overline{W}_{n,i}$ can be interpreted as residual fractions in a *stick-breaking* scheme: start with a stick of length 1. Choose a point on the stick according to distribution $W_{n,1}$, 'break' the stick into two pieces, discard the piece of length $W_{n,1}$ and rescale the remaining half to have length 1. Repeating this procedure $k$ times, and (6) is the fraction broken off at step $k$ relative to the original stick length.

Together with $T_n \stackrel{d}{=} S_n$, one could use (6) to compute the marginal distribution for $X_n[k]$ in terms of the ratios $\overline{W}_{n,i}$. In general the $W_{n,i}$ are not necessarily independent, and their joint distributions need to be worked out from (5). However, when $F$ has gamma distribution, $T_n, W_{n,1}, \dots, W_{n,k}$ are independent, and (6) leads to the following result of Patil and Taillie [25].

**Proposition 3 (Patil–Taillie [25]).** *If $F$ has distribution* gamma$(a,\lambda)$ *for some* $a,\lambda>0$, *then* $T_n$ *and the* $W_{n,1},\dots,W_{n,n-1}$ *in* (6) *are mutually independent. In this case,*

$$\begin{aligned}
X_n[1] &= \gamma_0\beta_1,\\
X_n[2] &= \gamma_0\bar\beta_1,\beta_2,\\
&\cdots\\
X_n[n-1] &= \gamma_0\bar\beta_1\bar\beta_2\cdots\bar\beta_{n-2}\beta_{n-1},\\
X_n[n] &= \gamma_0\bar\beta_1\bar\beta_2\cdots\bar\beta_{n-1},
\end{aligned}$$

*where* $\gamma_0$ *has distribution* gamma$(an,\lambda)$, $\beta_k$ *has distribution* beta$(a+1,(n-k)a)$, $\bar\beta_k=1-\beta_k$ *for* $1\le k\le n-1$, *and the random variables* $\gamma_0,\beta_1,\dots,\beta_{n-1}$ *are independent.*

**Proof.** This statement appeared as a casual in-line statement without proof in [25], perhaps since there is an elementary proof (which we will outline later). For the sake of demonstrating previous computations, we shall start with (5). By assumption, $S_k$ has distribution gamma$(ak,\lambda)$. One substitutes the density of gamma$(ak,\lambda)$ for $\nu_k$ to obtain

$$\begin{aligned}
\mathbb{P}(T_{n-k}\in \mathrm{d}s\,|\,T_{n-k+1}=t) &= C\left(\frac{(t-s)^a}{t}\right)\left(\frac{s^{a(n-k)-1}}{t^{a(n-k)+a-1}}\right)\\
&= C\left(1-\frac{s}{t}\right)^a\left(\frac{s}{t}\right)^{a(n-k)-1}
\end{aligned}$$

for some normalizing constant $C$. By rearranging, we see that $\frac{T_{n-k}}{T_{n-k+1}}$ has distribution beta$(a+1,a(n-k))$, and is independent of $T_{n-k+1}$. Therefore $W_{n,1}$ is independent of $T_n$. By the stick-breaking construction, $W_{n,2}$ is independent of $T_{n-1}$ and $T_n$, and hence of $W_{n,1}$. The final formula follow from rearranging (6). □

Here is another direct proof. By the stick-breaking construction, it is sufficient to show that $T_n$ is independent of $W_{n,1}=\frac{X_n[1]}{T_n}$. Note that

$$\begin{aligned}
&\mathbb{P}\bigl(X_n[1]/T_n\in \mathrm{d}u, T_n\in \mathrm{d}t\bigr)\\
&\quad = nu\mathbb{P}\left(\frac{X_n(1)}{X_n(1)+(X_n(2)+\cdots+X_n(n))}\in \mathrm{d}u, T_n\in \mathrm{d}t\right).
\end{aligned}\tag{7}$$

Since $X_n(1)\overset{d}{=}\text{gamma}(a,1)$, $S_{n-1}=X_n(2)+\cdots+X_n(n)\overset{d}{=}\text{gamma}(a(n-1),1)$, independent of $X_n(1)$, the ratio $\frac{X_n(1)}{X_n(1)+S_{n-1}}$ has distribution beta$(a,a(n-1))$ and is independent of $T_n$. Thus,

$$\begin{aligned}
\mathbb{P}\bigl(X_n[1]/T_n\in \mathrm{d}u\bigr) &= nu\frac{\Gamma(a+a(n-1))}{\Gamma(a)\Gamma(a(n-1))}u^{a-1}(1-u)^{a(n-1)-1}\\
&= \frac{\Gamma(a+1+a(n-1))}{\Gamma(a+1)\Gamma(a(n-1))}u^a(1-u)^{a(n-1)-1}.
\end{aligned}$$

In other words, $X_n[1]/T_n \stackrel{d}{=} \text{beta}(a, a(n-1))$. This proves the claim.

Lukacs [22] proved that if $X, Y$ are non-degenerate, positive independent random variables, then $X+Y$ is independent of $\frac{X}{X+Y}$ if and only if both $X$ and $Y$ have gamma distributions with the same scale parameter. Thus, one obtains another characterization of the gamma distribution.

**Corollary 4 (Patil–Taillie converse).** *If $T_n$ is independent of $X_n[1]/T_n$, then $F$ is* gamma$(a, \lambda)$ *for some $a, \lambda > 0$.*

**Proof.** One applies Lukacs' theorem to $X_n(1)$ and $(X_n(2) + \cdots + X_n(n))$ in (7). □

## 3. Size-biased permutation as induced order statistics

When $n$ i.i.d. pairs $(X_n(i), Y_n(i))$ are ordered by their $Y$-values, the corresponding $X_n(i)$ are called the *induced order statistics* of the vector $Y_n$, or its *concomitants*. Gordon [17] first proved the following result for finite $n$ which shows that finite i.i.d. size-biased permutation is a form of induced order statistics. Here we state the infinite sequence version, which is a special case of [26], Lemma 4.4.

**Proposition 5 (Perman, Pitman and Yor [26]).** *Let $x$ be a fixed positive sequence with finite sum $t = \sum_{i=1}^{\infty} x(i)$, $\varepsilon$ a sequence of i.i.d. standard exponential random variables, independent of $x$. Let $Y$ be the sequence with $Y(i) = \varepsilon(i)/x(i)$, $i = 1, 2, \ldots$, $Y^{\uparrow}$ its sequence of increasing order statistics. Define $X^*(k)$ to be the value of the $x(i)$ such that $Y(i)$ is $Y^{\uparrow}(k)$. Then $(X^*(k), k = 1, 2, \ldots)$ is a size-biased permutation of the sequence $x$. In particular, the size-biased permutation of a positive i.i.d. sequence $(X_n(1), \ldots, X_n(n))$ is distributed as the induced order statistics of the sequence $(Y_n(i) = \varepsilon_n(i)/X_n(i), 1 \le i \le n)$ for an independent sequence of i.i.d. standard exponentials $(\varepsilon_n(1), \ldots, \varepsilon_n(n))$, independent of the $X_n(i)$.*

**Proof.** Note that the $Y(i)$ are independent exponentials with rates $x(i)$. Let $\sigma$ be the random permutation such $Y(\sigma(i)) = Y^{\uparrow}(i)$. Note that $X^*(k) = x(\sigma(k))$. Then

$$\mathbb{P}\big(\sigma(1) = i\big) = \mathbb{P}\big(Y(i) = \min\{Y(j), j = 1, 2, \ldots\}\big) = \frac{x(i)}{t},$$

thus $X^*(1) \stackrel{d}{=} x[1]$. In general, for distinct indices $i_1, \ldots, i_k$, by the memoryless property of the exponential distribution,

$$\begin{aligned}
&\mathbb{P}\big(\sigma(k) = i_k | \sigma(1) = i_1, \ldots, \sigma(k) = i_{k-1}\big) \\
&\quad = \mathbb{P}\big(Y(i_k) = \min\big\{Y\big(\sigma(j)\big), j \ge k\big\} | \sigma(1) = i_1, \ldots, \sigma(k) = i_{k-1}\big) = \frac{x(i_k)}{t - \sum_{j=1}^{k-1} x(i_j)}.
\end{aligned}$$

Induction on $k$ completes the proof. □

Proposition 5 readily supplies simple proofs for joint, marginal and asymptotic distributions of i.i.d. size-biased permutation. For instance, the proof of the following nesting property, which can be cumbersome, amounts to i.i.d. thinning.

**Corollary 6.** *Consider a finite i.i.d. size-biased permutation* $(X_n[1], \dots, X_n[n])$ *from a distribution* $F$. *For* $1 \le m \le n$, *select* $m$ *integers* $a_1 < \cdots < a_m$ *by uniform sampling from* $\{1, \dots, n\}$ *without replacement. Then the subsequence* $\{X_n[a_j], 1 \le j \le m\}$ *is jointly distributed as a finite i.i.d. size-biased permutation of length* $m$ *from* $F$.

In general, the induced order statistics representation of size-biased permutation is often useful in studying limiting distribution as $n \to \infty$, since one can consider the i.i.d. pair $(X_n(i), Y_n(i))$ and appeal to tools from empirical process theory. We shall demonstrate this in Sections 5 and 6.

## 3.1. Joint and marginal distribution revisited

We now revisit the results in Section 2 using induced order statistics. This leads to a different formula for the joint distribution, and an alternative proof of the Barouch–Kaufman formula (3).

**Proposition 7.** $(X_n[k], k = 1, \dots, n)$ *is distributed as the first coordinate of the sequence of pairs* $((X_n^*(k), U_n^{\downarrow}(k)), k = 1, \dots, n)$, *where* $U_n^{\downarrow}(1) \ge \cdots \ge U_n^{\downarrow}(n)$ *is a sequence of uniform order statistics, and conditional on* $(U_n^{\downarrow}(k) = u_k, 1 \le k \le n)$, *the* $X_n^*(k)$ *are independent with distribution* $(G_{u_k}(\cdot), k = 1, \dots, n)$, *where*

$$G_u(\mathrm{d}x) = \frac{x\mathrm{e}^{-\phi^{-1}(u)x} F(\mathrm{d}x)}{-\phi'(\phi^{-1}(u))}. \tag{8}$$

*Here* $\phi$ *is the Laplace transform of* $X$, *that is,* $\phi(y) = \int_0^\infty \mathrm{e}^{-yx} F(\mathrm{d}x)$, $\phi'$ *its derivative and* $\phi^{-1}$ *its inverse function.*

**Proof.** Let $X_n$ be the sequence of $n$ i.i.d. draws from $F$, $\varepsilon_n$ an independent sequence of i.i.d. standard exponentials, $Y_n(i) = \varepsilon_n(i)/X_n(i)$ for $i = 1, \dots, n$. Note that the pairs $\{(X_n(i), Y_n(i)), 1 \le i \le n\}$ is an i.i.d. sample from the joint distribution $F(\mathrm{d}x)[x\mathrm{e}^{-yx}\,\mathrm{d}y]$. Thus, $Y_n(i)$ has marginal density

$$P\bigl(Y_n(i) \in \mathrm{d}y\bigr) = -\phi'(y)\,\mathrm{d}y, \qquad 0 < y < \infty, \tag{9}$$

and its distribution function is $F_Y = 1 - \phi$. Given $\{Y_n(i) = y_i, 1 \le i \le n\}$, the $X_n^*(i)$ defined in Proposition 5 are independent with conditional distribution $\widetilde{G}(y_i, \cdot)$ where

$$\widetilde{G}(y, \mathrm{d}x) = \frac{x\mathrm{e}^{-yx} F(\mathrm{d}x)}{-\phi'(y)}. \tag{10}$$

Equation (8) follows from writing the order statistics as the inverse transforms of ordered uniform variables

$$\begin{aligned} \bigl(Y_n^{\uparrow}(1), \dots, Y_n^{\uparrow}(n)\bigr) &\stackrel{d}{=} \bigl(F_Y^{-1}\bigl(U_n^{\downarrow}(n)\bigr), \dots, F_Y^{-1}\bigl(U_n^{\downarrow}(1)\bigr)\bigr) \\ &\stackrel{d}{=} \bigl(\phi^{-1}\bigl(U_n^{\downarrow}(1)\bigr), \dots, \phi^{-1}\bigl(U_n^{\downarrow}(n)\bigr)\bigr), \end{aligned} \tag{11}$$

where $(U_n^{\downarrow}(k), k = 1, \dots, n)$ is an independent decreasing sequence of uniform order statistics. Note that the minus sign in (9) results in the reversal of the sequence $U_n$ in the second equality of (11). $\square$

**Corollary 8.** *For $1 \le k \le n$ and $0 < u < 1$, let*

$$f_{n,k}(u) = \frac{\mathbb{P}(U_n^{\uparrow}(k) \in \mathrm{d}u)}{\mathrm{d}u} = n\binom{n-1}{k-1} u^{n-k}(1-u)^{k-1} \tag{12}$$

*be the density of the kth largest of the n uniform order statistics $(U_n^{\uparrow}(i), i = 1, \dots, n)$. Then*

$$\frac{\mathbb{P}(X_n[k] \in \mathrm{d}x)}{x F(\mathrm{d}x)} = \int_0^{\infty} \mathrm{e}^{-xy} f_{n,k}\big(\phi(y)\big)\,\mathrm{d}y. \tag{13}$$

**Proof.** Equation (12) follows from known results on order statistics, see [11]. For $u \in [0,1]$, let $y = \phi^{-1}(u)$. Then $\frac{\mathrm{d}y}{\mathrm{d}u} = \frac{1}{\phi'(\phi^{-1}(u))}$ by the inverse function theorem. Apply this change of variable to (8), rearrange and integrate with respect to $y$ to obtain (13). $\square$

In particular, for the first and last values,

$$\frac{\mathbb{P}(X_n[1] \in \mathrm{d}x)}{x F(\mathrm{d}x)} = n\int_0^{\infty} \mathrm{e}^{-xy} \phi(y)^{n-1}\,\mathrm{d}y,$$
$$\frac{\mathbb{P}(X_n[n] \in \mathrm{d}x)}{x F(\mathrm{d}x)} = n\int_0^{\infty} \mathrm{e}^{-xy} \big(1-\phi(y)\big)^{n-1}\,\mathrm{d}y.$$

### 3.1.1. *Alternative derivation of the Barouch–Kaufman formula*

Write $\phi(y) = \mathbb{E}(\mathrm{e}^{-yX})$ for $X$ with distribution $F$. Then $\phi(y)^{n-1} = \mathbb{E}(\mathrm{e}^{-yS_{n-1}})$ where $S_{n-1}$ is the sum of $(n-1)$ i.i.d. random variables with distribution $F$. Since all integrals involved are finite, by Fubini's theorem

$$\begin{aligned}\frac{\mathbb{P}(X_n[1] \in \mathrm{d}x)}{x F(\mathrm{d}x)} &= n\int_0^{\infty} \mathrm{e}^{-xy} \phi(y)^{n-1}\,\mathrm{d}y = n\mathbb{E}\biggl(\int_0^{\infty} \mathrm{e}^{-xy}\mathrm{e}^{-yS_{n-1}}\,\mathrm{d}y\biggr)\\ &= n\mathbb{E}\biggl(\frac{1}{x+S_{n-1}}\biggr),\end{aligned}$$

which is a rearranged version of the Barouch–Kaufman formula (3) for $k = 1$. Indeed, one can derive the entire formula from Proposition 7. For simplicity, we demonstrate the case $k = 2$.

**Proof of (3) for $k = 2$.** The joint distribution of the two largest uniform order statistics $U_n^{\downarrow}(1), U_n^{\downarrow}(2)$ has density

$$f(u_1, u_2) = n(n-1)u_2^{n-1} \qquad \text{for } 0 \le u_2 \le u_1 \le 1.$$

Conditioned on $U_n^{\downarrow}(1) = u_1, U_n^{\downarrow}(2) = u_2$, $X_n[1]$ and $X_n[2]$ are independent with distribution (8). Let $y_1 = \phi^{-1}(u_1)$, $y_2 = \phi^{-1}(u_2)$, so $\mathrm{d}y_1 = \frac{\mathrm{d}u_1}{\phi'(\phi^{-1}(u_1))}$, $\mathrm{d}y_2 = \frac{\mathrm{d}u_2}{\phi'(\phi^{-1}(u_2))}$. Let $S_{n-2}$ denote the sum of $(n-2)$ i.i.d. random variables with distribution $F$. Apply this change of variable and integrate out $y_1, y_2$, we have

$$\begin{aligned}
\frac{\mathbb{P}(X_n[1] \in \mathrm{d}x_1, X_n[2] \in \mathrm{d}x_2)}{x_1 x_2 F(\mathrm{d}x_1) F(\mathrm{d}x_2)} &= n(n-1) \int_0^\infty \int_{y_1}^\infty \mathrm{e}^{-y_1 x_1} \mathrm{e}^{-y_2 x_2} \bigl(\phi(y_2)\bigr)^{n-2} \, \mathrm{d}y_2 \, \mathrm{d}y_1 \\
&= n(n-1) \int_0^\infty \int_{y_1}^\infty \mathrm{e}^{-y_1 x_1} \mathrm{e}^{-y_2 x_2} \mathbb{E}\bigl(\mathrm{e}^{-y_2 S_{n-2}}\bigr) \, \mathrm{d}y_2 \, \mathrm{d}y_1 \\
&= n(n-1) \mathbb{E}\biggl(\int_0^\infty \int_{y_1}^\infty \mathrm{e}^{-y_1 x_1} \mathrm{e}^{-y_2 (x_2 + S_{n-2})} \, \mathrm{d}y_2 \, \mathrm{d}y_1\biggr) \\
&= n(n-1) \mathbb{E}\biggl(\int_0^\infty \mathrm{e}^{-y_1 x_1} \frac{\mathrm{e}^{-y_1 (x_2 + S_{n-2})}}{x_2 + S_{n-2}} \, \mathrm{d}y_1\biggr) \\
&= n(n-1) \mathbb{E}\biggl(\frac{1}{(x_2 + S_{n-2})(x_1 + x_2 + S_{n-2})}\biggr),
\end{aligned}$$

where the swapping of integrals is justified by Fubini's theorem, since all integrals involved are finite. □

***Example 9.*** Suppose $F$ is gamma$(a, 1)$. Then $\phi(y) = (\frac{1}{1+y})^a$, and $\phi^{-1}(u) = u^{-1/a} - 1$. Hence $G_u$ in (8) is

$$G_u(\mathrm{d}x) = \frac{x}{a u^{(a+1)/a}} \mathrm{e}^{-(u^{-1/a} - 1)x} F(\mathrm{d}x) = \frac{x^a}{\Gamma(a+1)} u^{-(a+1)/a} \mathrm{e}^{-x u^{-1/a}}.$$

That is, $G_u$ is gamma$(a + 1, u^{-1/a})$.

### 3.1.2. *Patil–Taillie revisited*

When $F$ is gamma$(a, \lambda)$, Lemma 7 gives the following result, which is an interesting complement to the Patil–Taillie representation in Proposition 3.

**Proposition 10.** *Suppose $F$ is* gamma$(a, \lambda)$. *Then $G_u$ is* gamma$(a + 1, \lambda u^{-1/a})$, *and*

$$\bigl(X_n[k], k = 1, \dots, n\bigr) \stackrel{d}{=} \bigl(\bigl[U_n^{\downarrow}(k)\bigr]^{1/a} \gamma_k, k = 1, \dots, n\bigr), \tag{14}$$

*where $\gamma_1, \dots, \gamma_n$ are i.i.d.* gamma$(a + 1, \lambda)$ *random variables, independent of the sequence of decreasing uniform order statistics $(U_n^{\downarrow}(1), \dots, U_n^{\downarrow}(n))$. Alternatively, jointly for $k = 1, \dots, n$*

$$\begin{aligned}
X_n^{\mathrm{rev}}[1] &= \gamma_1 \beta_{an,1}, \\
X_n^{\mathrm{rev}}[2] &= \gamma_2 \beta_{an,1}, \beta_{an-a,1}, \\
&\cdots
\end{aligned}$$

$$X_n^{\mathrm{rev}}[n-1] = \gamma_{n-1}\beta_{an,1}\beta_{an-a,1}\cdots\beta_{2a,1},$$
$$X_n^{\mathrm{rev}}[n] = \gamma_n\beta_{an,1}\beta_{an-a,1}\cdots\beta_{a,1},$$

*where the $\beta_{an-ia,1}$ for $i = 0, \dots, n-1$ are distributed as* $\mathrm{beta}(an - ia, 1)$, *and they are independent of each other and the $\gamma_k$.*

**Proof.** The distribution $G_u$ is computed in the same way as in Example 9 and (14) follows readily from Proposition 7. □

A direct comparison of the two different representations in Propositions 3 and 10 creates $n$ distributional identities. For example, the equality $X_n[1] = X_n^{\mathrm{rev}}[n]$ shows that the following two means of creating a product of independent random variables produce the same result in law:

$$\beta_{a+1,(n-1)a}\gamma_{an,\lambda} \stackrel{d}{=} \beta_{an,1}\gamma_{a+1,\lambda}, \tag{15}$$

where $\gamma_{r,\lambda}$ and $\beta_{a,b}$ denote random variables with distributions $\mathrm{gamma}(r, \lambda)$ and $\mathrm{beta}(a, b)$, respectively. Indeed, this identity comes from the usual 'beta–gamma' algebra, which allows us to write

$$\beta_{a+1,(n-1)a} = \frac{\gamma_{a+1,\lambda}}{\gamma_{a+1,\lambda} + \gamma_{a(n-1),\lambda}}, \qquad \beta_{an,1} = \frac{\gamma_{an,\lambda}}{\gamma_{an,\lambda} + \gamma_{1,\lambda}}$$

for $\gamma_{a(n-1),\lambda}, \gamma_{1,\lambda}$ independent of all others. Thus, (15) reduces to

$$\gamma_{a+1,\lambda} + \gamma_{a(n-1),\lambda} \stackrel{d}{=} \gamma_{an,\lambda} + \gamma_{1,\lambda},$$

which is true since both sides have distribution $\mathrm{gamma}(an + 1, \lambda)$.

# 4. Limit in distributions of finite size-biased permutations

As hinted in the Introduction, our setup is a finite version of the size-biased permutation of ranked jumps of a subordinator, studied in [26]. In this section, we make this statement rigorous (see Proposition 15).

Let $\Delta = \{x = (x(1), x(2), \dots)\colon x(i) \geq 0, \sum_i x(i) \leq 1\}$ and $\Delta^{\downarrow} = \{x^{\downarrow}\colon x \in \Delta\}$ be closed infinite simplices, the later contains sequences with non-increasing terms. Denote their boundaries by $\Delta_1 = \{x \in \Delta\colon \sum_i x(i) = 1\}$ and $\Delta_1^{\downarrow} = \{x \in \Delta^{\downarrow}, \sum_i x(i) = 1\}$, respectively. Any finite sequence can be associated with an element of $\Delta_1$ after being normalized by its sum and extended with zeros. Thus, one can speak of convergence in distribution of sequences in $\Delta$.

We have to consider $\Delta$ and not just $\Delta_1$ because a sequence in $\Delta_1$ can converge to one in $\Delta$. For example, the sequence $(X_n, n \geq 1) \in \Delta_1$ with $X_n(i) = 1/n$ for all $i = 1, \dots, n$ converges to the elementwise zero sequence in $\Delta$. Thus, we need to define convergence in distribution of size-biased permutations in $\Delta$. We shall do this using Kingman's paintbox. In particular, with this definition, convergence of size-biased permutation is equivalent to convergence of order statistics. Our treatment in Section 4.1 follows that of Gnedin [15] with simplified assumptions. The proofs can be found in [15].

It then follows that size-biased permutation of finite i.i.d. sequences with almost sure finite sum converges to the size-biased permutation of the sequence of ranked jumps of a *subordinator*, roughly speaking, a non-decreasing process with independent and homogeneous increments. We give a review of Lévy processes and subordinators in Section 4.2. Many properties such as stick-breaking and the Markov property of the remaining sum have analogues in the limit. We explore these in Section 4.3.

## 4.1. Kingman's paintbox and some convergence theorems

*Kingman's paintbox* [20] is a useful way to describe and extend size-biased permutations. For $x \in \Delta$, let $s_k$ be the sum of the first $k$ terms. Note that $x$ defines a partition $\varphi(x)$ of the unit interval $[0,1]$, consisting of *components* which are intervals of the form $[s_k, s_{k+1})$ for $k = 1, 2, \ldots$, and the interval $[s_\infty, 1]$, which we call the *zero component*. Sample points $\xi_1, \xi_2, \ldots$ one by one from the uniform distribution on $[0,1]$. Each time a sample point discovers a new component that is not in $[s_\infty, 1]$, write down its size. If the sample point discovers a new point of $[s_\infty, 1]$, write 0. Let $X^* = (X^*(1), X^*(2), \ldots)$ be the random sequence of sizes. Since the probability of discovery of a particular (non-zero) component is proportional to its length, the non-zero terms in $X^*$ form the size-biased permutation of the non-zero terms in $x$ as defined by (1). In Kingman's paintbox terminology, the components correspond to different colors used to paint the balls with labels $1, 2, \ldots$. Two balls $i, j$ have the same paint color if and only if $\xi_i$ and $\xi_j$ fall in the same component. The size-biased permutation $X^*$ records the size of the newly discovered components, or paint colors. The zero component represents a continuum of distinct paint colors, each of which can be represented at most once.

By construction, the size-biased permutation of a sequence $x$ does not depend on the ordering of its terms. In particular, convergence in $\Delta^\downarrow$ implies convergence in distribution of the corresponding sequences of size-biased permutations. The converse is also true. Proofs of the following statements can be found in [15].

**Theorem 11 (Equivalence of convergence of order statistics and s.b.p. [15]).** *Suppose $X^\downarrow$, $X_1^\downarrow, X_2^\downarrow, \ldots$ are random elements of $\Delta^\downarrow$ and $X_n^\downarrow \stackrel{f.d.d.}{\to} X^\downarrow$. Then $(X_n^\downarrow)^* \stackrel{f.d.d.}{\to} (X^\downarrow)^*$.*

*Conversely, suppose $X_1, X_2, \ldots$ are random elements of $\Delta$ and $X_n^* \stackrel{f.d.d.}{\to} Y$ for some $Y \in \Delta$. Then $Y^* \stackrel{d}{=} Y$, and $X_n^\downarrow \stackrel{f.d.d.}{\to} Y^\downarrow$.*

We can speak of convergence of size-biased permutation on $\Delta$ without having to pass to order statistics. However, convergence of random elements in $\Delta$ implies neither convergence of order statistics nor of their size-biased permutations. To achieve convergence, we need to keep track of the sum of components. This prevents large order statistics from 'drifting' to infinity.

**Theorem 12 ([15]).** *Suppose $X, X_1, X_2, \ldots$ are random elements of $\Delta$ and*

$$\left(X_n, \sum_i X_n(i)\right) \stackrel{f.d.d.}{\to} \left(X, \sum_i X(i)\right).$$

*Then* $X_n^* \overset{f.d.d.}{\to} X^*$ *and* $X_n^{\downarrow} \overset{f.d.d.}{\to} X^{\downarrow}$.

## 4.2. Finite i.i.d. size-biased permutation and ranked jumps of a subordinator

Let $((X_n), n \geq 1)$ be an i.i.d. positive triangular array, that is, $X_n = (X_n(1), \dots, X_n(n))$, where $X_n(i), i = 1, \dots, n$ are i.i.d. and a.s. positive. Write $T_n$ for $\sum_{i=1}^n X_n(i)$. We ask for conditions under which the size-biased permutation of the sequence $(X_n, n \geq 1)$ converges to the size-biased permutation of some infinite sequence $X$. Let us restrict to the case $T_n \overset{d}{\to} T$ for some $T < \infty$ a.s. A classical result in probability states that $T_n \overset{d}{\to} T$ if and only if $T = \tilde{T}(1)$ for some *Lévy process* $\tilde{T}$, which in this case is a *subordinator*. For self-containment, we gather some necessary facts about Lévy processes and subordinators below. See [19], Section 15, for their proofs, [2] for a thorough treatment of subordinators.

***Definition 13.*** *A* Lévy process $\tilde{T}$ *in* $\mathbb{R}$ *is a stochastic process with right-continuous left-limits paths, stationary independent increments, and* $\tilde{T}(0) = 0$. *A* subordinator $\tilde{T}$ *is a Lévy process, with real, finite, non-negative increments.*

Following [19], we do not allow the increments to have infinite value. We suffer no loss of generality, since subordinators with jumps of possibility infinite size do not contribute to our discussion of size-biased permutation. Let $\tilde{T}$ be a subordinator, $T = \tilde{T}(1)$. For $t, \lambda \geq 0$, using the fact that increments are stationary and independent, one can show that

$$\mathbb{E}\big(\exp\big(-\lambda \tilde{T}(t)\big)\big) = \exp\big(-t\Phi(\lambda)\big),$$

where the function $\Phi: [0, \infty) \to [0, \infty)$ is called the *Laplace exponent* of $\tilde{T}$. It satisfies the *Lévy–Khinchine formula*

$$\Phi(\lambda) = \mathsf{d}\lambda + \int_0^\infty \big(1 - \mathrm{e}^{-\lambda x}\big)\Lambda(\mathrm{d}x), \qquad \lambda \geq 0,$$

where $\mathsf{d} > 0$ is the *drift coefficient*, and $\Lambda$ a unique measure on $(0, \infty)$ with $\Lambda([1, \infty)) < \infty$, called the *Lévy measure of* $\tilde{T}$. Assume $\int_0^1 x\Lambda(\mathrm{d}x) < \infty$, which implies a.s. $\tilde{T}(1) = T < \infty$. Then over $[0, 1]$, $\tilde{T}$ is the sum of a deterministic drift plus a Poisson point process (p.p.p.) with i.i.d. jumps

$$(\tilde{T})(t) = \mathsf{d}t + \sum_i X(i)\mathbf{1}_{\{\sigma(i) \leq t\}}$$

for $0 \leq t \leq 1$, where $\{(\sigma(i), X(i)), i \geq 1\}$ are points in a p.p.p. on $(0, \infty)^2$ with intensity measure $\mathrm{d}t\Lambda(\mathrm{d}x)$. The $X(i)$ are the *jumps* of $\tilde{T}$. Finally, we need a classical result on convergence of i.i.d. positive triangular arrays to subordinators (see [19], Section 15).

**Theorem 14.** *Let $(X(n), n \geq 1)$ be an i.i.d. positive triangular array, $T_n = \sum_{i=1}^n X_n(i)$. Then $T_n \xrightarrow{d} T$ for some random variable $T$, $T < \infty$ a.s. if and only if $T = \tilde{T}(1)$ for some subordinator $\tilde{T}$ whose Lévy measure $\Lambda$ satisfies $\int_0^1 x\Lambda(\mathrm{d}x) < \infty$. Furthermore, let $\mu_n$ be the measure of $X_n(i)$. Then on $\mathbb{R}_+$, the sequence of measures $(n\mu_n)$ converges vaguely to $\Lambda$, written*

$$n\mu_n \xrightarrow{v} \Lambda.$$

*That is, for all $f : \mathbb{R}_+ \to \mathbb{R}_+$ continuous with compact support, $n\mu_n(f) = n\int_0^\infty f(x)\mu_n(\mathrm{d}x)$ converges to $\Lambda(f) = \int_0^\infty f(x)\Lambda(\mathrm{d}x)$. In particular, if $\mu_n, \Lambda$ have densities $\rho_n, \rho$, respectively, then we have pointwise convergence for all $x > 0$*

$$n\rho_n(x) \to \rho(x).$$

**Proposition 15.** *Let $(X_n, n \geq 1)$ be an i.i.d. positive triangular array, $T_n = \sum_{i=1}^n X_n(i)$. Suppose $T_n \xrightarrow{d} T$ for some $T$ a.s. finite. Let $X$ be the sequence of ranked jumps of $T$ arranged in any order, $(X/T)^*$ be the size-biased permutation of the sequence $(X/T)$ as defined using Kingman's paintbox, $(X^*)' = T \cdot (X/T)^*$. Then*

$$(X_n)^* \xrightarrow{f.d.d.} \left(X^*\right)'.$$

**Proof.** The sequence of decreasing order statistics $X_n^\downarrow$ converges in distribution to $X^\downarrow$ [19]. Since $T_n, T > 0$ a.s. and $T_n \xrightarrow{d} T$, $X_n^\downarrow/T_n \xrightarrow{\text{f.d.d.}} X^\downarrow/T$. Theorem 11 combined with multiplying through by $T$ prove the claim. □

For subordinators without drift, $\mathsf{d} = 0$, $\sum_i X(i) = T$, hence $(X^*)' = X^*$. When $\mathsf{d} > 0$, the sum of the jumps $\sum_i X_i$ is strictly less than $T$, so $(X^*)' \neq X^*$. In this case, there is a non-trivial zero component coming from an accumulation of mass at 0 of $n\mu_n$ in the limit as $n \to \infty$. At each finite, large $n$, we have a significant number of jumps with 'microscopic' size.

The case without drift was studied by Perman, Pitman and Yor in [26] with the assumption $\Lambda(0, \infty) = \infty$ to ensure that the sequence of jumps has infinite length. We shall re-derive some of their results as limits of results for finite i.i.d. size-biased permutation using Theorem 14 in the next section. One can obtain another finite version of the Perman–Pitman–Yor setup by letting $\Lambda(0, \infty) < \infty$, but this can be reduced to finite i.i.d. size-biased permutation by conditioning. Specifically, $\tilde{T}$ is now a compound Poisson process, where the subordinator waits for an exponential time with rate $\Lambda(0, \infty)$ before making a jump $X$, whose length is independent of the waiting time and distributed as $\mathbb{P}(X \leq t) = \Lambda(0, t]/\Lambda(0, \infty)$ [2]. If $(X(1), X(2), \ldots)$ is the sequence of successive jumps of $(\tilde{T}_s, s \geq 0)$, then $(X(1), X(2), \ldots, X(N))$ is the sequence of successive jumps of $(\tilde{T}_s, 0 \leq s \leq 1)$, where $N$ is a Poisson random variable with mean $\Lambda(0, \infty)$, independent of the jump sequence $(X(1), X(2), \ldots)$. For $N > 0$, properties of the size-biased permutation of $(X(1), \ldots, X(N))$ can be deduced from those of a finite i.i.d. size-biased permutation by conditioning on $N$.

## 4.3. Markov property in the limit

Results in [26] can be obtained as limits of those in Section 2, including the Markov property and the stick-breaking representation. Consider a subordinator with Lévy measure $\Lambda$, drift $\mathsf{d} = 0$. Let $\tilde{T}_0$ be the subordinator at time 1. Assume $\Lambda(1, \infty) < \infty$, $\Lambda(0, \infty) = \infty$, $\int_0^1 x\Lambda(\mathrm{d}x) < \infty$, and $\Lambda(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$ for some density $\rho$. Note that $\tilde{T}_0 < \infty$ a.s., and it has a density determined by $\rho$ via its Laplace transform, which we denote $\nu$. Let $\tilde{T}_k$ denote the remaining sum after removing the first $k$ terms of the size-biased permutation of the sequence $X^{\downarrow}$ of ranked jumps.

**Proposition 16 ([26]).** *The sequence* $(\tilde{T}_0, \tilde{T}_1, \ldots)$ *is a Markov chain with stationary transition probabilities*

$$\mathbb{P}(\tilde{T}_1 \in \mathrm{d}t_1 | \tilde{T}_0 = t) = \frac{t - t_1}{t} \cdot \rho(t - t_1) \frac{\nu(t_1)}{\nu(t)}\,\mathrm{d}t_1.$$

Note the similarity to (5). Starting with (4) and send $n \to \infty$, for any finite $k$, we have $\nu_{n-k} \to \nu$ pointwise, and by Theorem 14, $(n-k)\nu_1 \to \rho$ pointwise over $\mathbb{R}$, since there is no drift term. Thus, the analogue of (4) in the limit is

$$\mathbb{P}(\tilde{T}_0 \in \mathrm{d}t_0, \ldots, \tilde{T}_k \in \mathrm{d}t_k) = \left(\prod_{i=0}^{k-1} \frac{t_i - t_{i+1}}{t_i} \rho(t_i - t_{i+1})\right) \nu(t_k)\,\mathrm{d}t_0 \cdots \mathrm{d}t_k.$$

Rearranging gives the transition probability in Proposition 16.

Conditionally given $\tilde{T}_0 = t_0, \tilde{T}_1 = t_1, \ldots, \tilde{T}_n = t_n$, the sequence of remaining terms in the size-biased permutation $(X[n+1], X[n+2], \ldots)$ is distributed as $(X^{\downarrow}(1), X^{\downarrow}(2), \ldots)$ conditioned on $\sum_{i \geq 1} X^{\downarrow}(i) = t_n$, independent of the first $n$ size-biased picks [26], Theorem 4.2. The stick-breaking representation in (6) now takes the form

$$\frac{X[k]}{\tilde{T}_0} = W_k \prod_{i=1}^{k-1} \overline{W}_i, \tag{16}$$

where $X[k]$ is the $k$th size-biased pick, and $W_i = \frac{X[i]}{\tilde{T}_{i-1}}$, $\overline{W}_i = 1 - W_i = \frac{\tilde{T}_i}{\tilde{T}_{i-1}}$. Proposition 3 and Corollary 4 parallel the following result.

**Proposition 17 (McCloskey [23] and Perman–Pitman–Yor [26]).** *The random variables* $\tilde{T}_0$ *and* $W_1, W_2, \ldots$ *in* (16) *are mutually independent if and only if* $\tilde{T}_0$ *has distribution* $\mathrm{gamma}(a, \lambda)$ *for some* $a, \lambda > 0$. *In this case, the* $W_i$ *are i.i.d. with distribution* $\mathrm{beta}(1, a)$ *for* $i = 1, 2, \ldots$.

## 4.4. Invariance under size-biased permutation

We take a small detour to explain some results related to Propositions 3 and 17 on characterization of size-biased permutations. For a random discrete distribution prescribed by its probability mass function $P \in \Delta_1$, let $P^*$ be its size-biased permutation. (Recall that $\Delta$ is the closed infinite simplex, $\Delta_1$ is its boundary. These are defined at the beginning of Section 4.) Given $P \in \Delta_1$, one may ask when is there a $Q \in \Delta_1$ such that $Q = P^*$. Clearly $(P^*)^* = P^*$ for any $P \in \Delta_1$, thus this question is equivalent to characterizing random discrete distributions on $\mathbb{N}$ which are invariant under size-biased permutation (ISBP). One such characterization is the following [27], Theorem 4: suppose $P \in \Delta_1$, $P_1 > 0$ a.s. Then $P = P^*$ if and only if for each $k = 2, 3, \ldots,$ the function of $k$-tuples of positive integers

$$(n_1, \ldots, n_k) \mapsto \mathbb{E}\left(\prod_{i=1}^{k} P_i^{n_i-1} \prod_{i=1}^{k-1}\left(1 - \sum_{j=1}^{i} P_j\right)\right) \tag{17}$$

is a symmetric function of $n_1, \ldots, n_k$. Here is the interpretation of this function using Kingman's paintbox. Sample $n = n_1 + \cdots + n_k$ points $\xi_1, \ldots, \xi_n$ one by one from the uniform distribution on $[0, 1]$. Each time a sample point discovers a new component, write down its size. Conditioned on the event $P = p = (p_1, p_2, \ldots)$, then

$$\mathbb{E}\left(\prod_{i=1}^{k} P_i^{n_i-1} \prod_{i=1}^{k-1}\left(1 - \sum_{j=1}^{i} P_j\right) \middle| P = p\right)$$

is the probability that we discovered $k$ distinct paint boxes, where the first box $p_1$ is rediscovered $n_1 - 1$ times, then we discover the box $p_2$, which is then rediscovered $n_2 - 1$ times, and so on. Thus if $P = P^*$, then the value of the function in (17) is the probability of the event $\xi_1 = \cdots = \xi_{n_1}, \xi_{n_1+1} = \cdots = \xi_{n_1+n_2}$ and so on up to $\xi_{n-n_k+1} = \cdots = \xi_n$, and that $\xi_1, \xi_{n_1+1}, \ldots, \xi_n$ are distinct.

Consider the stick-breaking representation of size-biased permutation, that is, $P_n = W_1 \cdots W_{n-1}\overline{W}_n$. Suppose we want to find ISBP distributions $P$ such that the $W_i$'s are independent. By the above characterization, this is equivalent to finding such $P$ where

$$\mathbb{E}\left(W_1^r \bar{W}_1^{s+1}\right)\mathbb{E}\left(W_2^s\right) = \mathbb{E}\left(W_1^s \bar{W}_1^{r+1}\right)\mathbb{E}\left(W_2^r\right)$$

for all pairs of non-negative integers $r$ and $s$. By analyzing this equation, Pitman [27] proved a complete characterization of ISBP in this case.

**Theorem 18 ([27]).** *Let $P \in \Delta_1$, $P_1 < 1$, and $P_n = W_1 \cdots W_{n-1}\overline{W}_n$ for independent $W_i$. Then $P = P^*$ if and only if one of the four following conditions holds.*

1. *$P_n \geq 0$ a.s. for all $n$, in which case the distribution of $W_n$ is*

$$\text{beta}(1 - \alpha, \theta + n\alpha)$$

   *for every $n = 1, 2, \ldots,$ for some $0 \leq \alpha < 1$, $\theta > -\alpha$.*
2. *For some integer constant $m$, $P_n \geq 0$ a.s. for all $1 \leq n \leq m$, and $P_n = 0$ a.s. otherwise. Then either*
   (a) *For some $\alpha > 0$, $W_n$ has distribution* beta$(1 + \alpha, m\alpha - n\alpha)$ *for $n = 1, \ldots, m$; or*
   (b) *$W_n = 1/(m - n + 1)$ a.s., that is, $P_n = 1/m$ a.s. for $n = 1, \ldots, m$; or*

(c) $m = 2$, *and the distribution* $F$ *on* $(0, 1)$ *defined by* $F(\mathrm{d}w) = \bar{w}\mathbb{P}(W_1 \in \mathrm{d}w)/\mathbb{E}(\bar{W}_1)$ *is symmetric about* $1/2$.

The McCloskey case of Proposition 17 is case 1 with $\alpha = 0, \theta > 0$, and Patil–Taillie case of Proposition 3 is case 2(a). These two cases are often written in the form $W_i$ has distribution $\mathrm{beta}(1 - \alpha, \theta + i\alpha)$, $i = 1, 2, \ldots$ for pairs of real numbers $(\alpha, \theta)$ satisfying either $(0 \le \alpha < 1, \theta > -\alpha)$ (case 1), or $(\alpha < 0, \theta = m\alpha)$ for some $m = 1, 2, \ldots$ (case 2(a)). In both settings, such a distribution $P$ is known as the $\mathrm{GEM}(\alpha, \theta)$ distribution. The abbreviation GEM was introduced by Ewens, which stands for Griffiths–Engen–McCloskey. If $P$ is $\mathrm{GEM}(\alpha, \theta)$, then $P^{\downarrow}$ is called a Poisson–Dirichlet distribution with parameters $(\alpha, \theta)$, denoted $\mathrm{PD}(\alpha, \theta)$ [26].

In the McCloskey case of Proposition 17, the function (17) is the Donnelly–Tavare–Griffiths formula. If one changes variables from $n_i$'s to $s_j$, where $s_j$ is the number of $n_i$'s equal to $j$, then (17) is the Ewens' sampling formula [13]. In studying this formula, Kingman [20] initiated the theory of partition structures; see [14] for recent developments. Subsequent authors have studied partition structures and their representations in terms of exchangeable random partitions, random discrete distributions, random trees and associated random processes of fragmentation and coalescence, Bayesian statistics, and machine learning. See [28] and references therein.

## 5. Asymptotics of the last $u$ fraction of the size-biased permutation

In this section, we derive Glivenko–Cantelli and Donsker-type theorems for the distribution of the last $u$ fraction of terms in a finite i.i.d. size-biased permutation. It is especially convenient to work with the induced order statistics representation since we can appeal to tools from empirical process theory. In particular, our results are special cases of more general statements which hold for arbitrary induced order statistics in $d$ dimensions (see Section 5.2). Features pertaining to i.i.d. size-biased permutation are presented in Lemma 19. The proof is a direct computation. We first discuss the interesting successive sampling interpretation of Lemma 19, quoting some results needed to make the discussion rigorous. We then derive the aforementioned theorems and conclude with a brief historical account of induced order statistics.

**Lemma 19.** *Suppose* $F$ *has support on* $(0, \infty)$, *finite mean. For* $u \in (0, 1)$, *define*

$$F_u(\mathrm{d}x) = \frac{\mathrm{e}^{-x\phi^{-1}(u)}}{u} F(\mathrm{d}x) \tag{18}$$

*and extend the definition to* $\{0, 1\}$ *by continuity, where* $\phi$ *is the Laplace transform of* $F$ *as in Proposition* 7. *Then* $F_u$ *is a probability distribution on* $(0, \infty)$ *for all* $u \in [0, 1]$, *and* $G_u$ *in* (8) *satisfies*

$$G_u(\mathrm{d}x) = xF_u(\mathrm{d}x)/\mu_u, \tag{19}$$

*where* $\mu_u = \int x F_u(\mathrm{d}x) = \frac{-\phi'(\phi^{-1}(u))}{u}$. *Furthermore*,

$$\int_0^u G_s(\mathrm{d}x)\,\mathrm{d}s = F_u(\mathrm{d}x) \tag{20}$$

*for all* $s \in [0,1]$. *In other words, the density*

$$f(u,x) = F_u(\mathrm{d}x)/F(\mathrm{d}x) = u^{-1}\mathrm{e}^{-x\phi^{-1}(u)}$$

*of* $F_u$ *with respect to* $F$ *solves the differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}u}\bigl[uf(u,x)\bigr] = \frac{-xf(u,x)}{\mu_u} \tag{21}$$

*with boundary condition* $f(1,x) \equiv 1$.

For any distribution $F$ with finite mean $\mu$ and positive support, $xF(\mathrm{d}x)/\mu$ defines its *size-biased distribution*. If $F$ is the empirical distribution of $n$ positive values $x_n(1),\ldots,x_n(n)$, for example, one can check that $xF(\mathrm{d}x)/\mu$ is precisely the distribution of the first size-biased pick $X_n[1]$. For continuous $F$, the name size-biased distribution is justified by the following lemma.

**Lemma 20.** *Consider an i.i.d. size-biased permutation* $(X_n[1],\ldots,X_n[n])$ *from a distribution* $F$ *with support on* $[0,\infty)$ *and finite mean* $\mu$. *Then*

$$\lim_{n\to\infty} \mathbb{P}\bigl(X_n[1] \in \mathrm{d}x\bigr) = \frac{xF(\mathrm{d}x)}{\mu}.$$

Since the $\lfloor nu\rfloor$th smallest out of $n$ uniform order statistics converge to $u$ as $n \to \infty$, $G_u$ is the limiting distribution of $X_n^{\mathrm{rev}}[\lfloor nu\rfloor]$, the size-biased pick performed when a $u$-fraction of the sequence is left. By (19), $G_u$ is the size-biased distribution of $F_u$. Thus, $F_u$ can be interpreted as the limiting distribution of the remaining $u$-fraction of terms in a successive sampling scheme. This intuition is made rigorous by Corollary 21 below. In words, it states that $F_u$ is the limit of the empirical distribution function (e.d.f.) of the last $u > 0$ fraction in a finite i.i.d. size-biased permutation.

**Corollary 21.** *For* $u \in (0,1]$, *let* $F_{n,u}(\cdot)$ *denote the empirical distribution of the last* $\lfloor nu\rfloor$ *values of an i.i.d. size-biased permutation with length* $n$. *For each* $\delta \in (0,1)$, *as* $n \to \infty$,

$$\sup_{u\in[\delta,1]} \sup_{\mathbf{I}} \bigl|F_{n,u}(\mathbf{I}) - F_u(\mathbf{I})\bigr| \xrightarrow{a.s.} 0, \tag{22}$$

*where* $\mathbf{I}$ *ranges over all subintervals of* $(0,\infty)$.

Therefore in the limit, after removing the first $1-u$ fraction of terms in the size-biased permutation, we are left with an (infinitely) large sequence of numbers distributed like i.i.d. draws from $F_u$, from which we do a size-biased pick, which has distribution $G_u(\mathrm{d}x) = xF_u(\mathrm{d}x)/\mu_u$ as specified by (19).

Since $X_n^{\mathrm{rev}}[\lfloor nu \rfloor]$ converges in distribution to $G_u$ for $u \in [0, 1]$, Corollary 21 lends a sampling interpretation to Lemma 19. Equation (21) has the heuristic interpretation as characterizing the evolution of the mass at $x$ over time $u$ in a successive sampling scheme. To be specific, consider a successive sampling scheme on a large population of $N$ individuals, with species size distribution $H$. Scale time such that at time $u$, for $0 \le u \le 1$, there are $Nu$ individuals (from various species) remaining to be sampled. Let $H_u$ denote the distribution of species sizes at time $u$, and fix the bin $(x, x + \mathrm{d}x)$ of width $\mathrm{d}x$ on $(0, \infty)$. Then $NuH_u(\mathrm{d}x)$ is the number of individuals whose species size lie in the range $(x, x + \mathrm{d}x)$ at time $u$. Thus, $\frac{\mathrm{d}}{\mathrm{d}u}NuH_u(\mathrm{d}x)$ is the rate of individuals to be sampled from this range of species size at time $u$. The probability of an individual whose species size is in $(x, x + \mathrm{d}x)$ being sampled at time $u$ is $\frac{xH_u(\mathrm{d}x)}{\int_0^\infty xH_u(\mathrm{d}x)}$. As we scaled time such that $u \in [0, 1]$, in time $\mathrm{d}u$ we sample $N\,\mathrm{d}u$ individuals. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}u}NuH_u(\mathrm{d}x) = -N\frac{xH_u(\mathrm{d}x)}{\int_0^\infty xH_u(\mathrm{d}x)}.$$

Let $f(u, x) = H_u(\mathrm{d}x)/H(\mathrm{d}x)$, then as a function in $u$, the above equation reduces to (21).

***Example 22.*** Suppose $F$ puts probability $p$ at $a$ and $1 - p$ at $b$, with $a < b$. Let $p(u) = F_u(a)$ be the limiting fraction of $a$ left when proportion $u$ of the sample is left. Then the evolution equation (21) becomes

$$\begin{aligned}
p'(u) &= u^{-1}\left(\frac{a}{ap(u) + b(1 - p(u))} - 1\right)p(u) \\
&= u^{-1}\left(\frac{a - ap(u) - b(1 - p(u))}{b - (b - a)p(u)}\right)p(u) \\
&= u^{-1}\frac{(1 - p(u))p(u)(a - b)}{b - (b - a)p(u)}, \qquad 0 \le u \le 1,
\end{aligned}$$

with boundary condition $p(0) = p$. To solve for $p(u)$, let $y$ solve $u = p\mathrm{e}^{-ay} + (1 - p)\mathrm{e}^{-by}$. Then $p(u) = p\mathrm{e}^{-ay}/u$.

## 5.1. A Glivenko–Cantelli theorem

We now state a Glivenko–Cantelli-type theorem which applies to size-biased permutations of finite deterministic sequences. Versions of this result are known in the literature [4,17,18,31], see discussions in Section 5.2. We offer an alternative proof using induced order statistics.

**Theorem 23.** *Let $(x_n, n = 1, 2, \dots)$ be a deterministic triangular array of positive numbers with corresponding c.d.f. sequence $(E_n, 1 \le n)$. Suppose*

$$\sup_x \left|E_n(x) - F(x)\right| \to 0 \qquad \text{as } n \to \infty \tag{23}$$

*for some distribution $F$ on $(0,\infty)$. Let $u \in (0,1]$. Let $E_{n,u}(\cdot)$ be the empirical distribution of the last $\lfloor nu \rfloor$ terms in a size-biased permutation of the sequence $x_n$. Then for each $\delta \in (0,1)$,*

$$\sup_{u\in[\delta,1]} \sup_{\mathbf{I}} \bigl|E_{n,u}(\mathbf{I}) - F_u(\mathbf{I})\bigr| \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty, \tag{24}$$

*where $\mathbf{I}$ ranges over all subintervals of $(0,\infty)$.*

We state the theorem in terms of convergence in distribution of the last $u$ fraction of terms in a successive sampling scheme. Since $E_n \to F$ uniformly, an analogous result holds for the distribution of the first $(1-u)$ fraction. The last $u$ fraction is more interesting due to the heuristic interpretation of $F_u$ in the previous section.

**Proof of Theorem 23.** Define $Y_n(i) = \varepsilon_n(i)/x_n(i)$ for $i = 1,\dots,n$ where $\varepsilon_n(i)$ are i.i.d. standard exponentials as in Proposition 5. Let $H_n$ be the empirical distribution function (e.d.f.) of the $Y_n(i)$,

$$H_n(y) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Y_n(i)<y\}}.$$

Let $J_n$ denote the e.d.f. of $(x_n(i), Y_n(i))$. By Proposition 7,

$$\begin{aligned} E_{n,u}(\mathbf{I}) &= \frac{n}{\lfloor nu \rfloor}\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{x_n(i)\in\mathbf{I}\}}\mathbf{1}_{\{Y_n(i)>H_n^{-1}(1-u)\}} \\ &= \frac{n}{\lfloor nu \rfloor} J_n\bigl(\mathbf{I}\times\bigl[H_n^{-1}(1-u), H_n^{-1}(1)\bigr]\bigr). \end{aligned} \tag{25}$$

Fix $\delta \in (0,1)$, and let $u \in [\delta,1]$. Let $\phi$ be the Laplace transform of $F$ and $J$ the joint law of $(X, \varepsilon/X)$, where $X$ is a random variable with distribution $F$, and $\varepsilon$ is an independent standard exponential. Note that $\frac{1}{u}J(\mathbf{I}\times[\phi^{-1}(u),\infty)) = F_u(\mathbf{I})$. Thus,

$$\begin{aligned} \mathbb{E}_{n,u}(\mathbf{I}) - F_u(\mathbf{I}) = {}& \biggl(\frac{n}{\lfloor nu \rfloor} J_n\bigl(\mathbf{I}\times\bigl[H_n^{-1}(1-u), H_n^{-1}(1)\bigr]\bigr) - \frac{n}{\lfloor nu \rfloor} J_n\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr)\biggr) \\ &+ \biggl(\frac{n}{\lfloor nu \rfloor} J_n\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr) - \frac{1}{u} J\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr)\biggr). \end{aligned} \tag{26}$$

Let us consider the second term. Note that

$$J_n\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr) = \int_0^\infty \mathrm{e}^{-t\phi^{-1}(u)}\mathbf{1}_{\{t\in\mathbf{I}\}} E_n(\mathrm{d}t).$$

Since $E_n$ converges to $F$ uniformly and $\mathrm{e}^{-t\phi^{-1}(u)}$ is bounded for all $t \in (0,\infty)$ and $u \in [\delta,1]$,

$$\sup_{u\in[\delta,1]} \sup_{\mathbf{I}} \biggl|\frac{n}{\lfloor nu \rfloor} J_n\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr) - \frac{1}{u} J\bigl(\mathbf{I}\times\bigl[\phi^{-1}(u),\infty\bigr)\bigr)\biggr| \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Let us consider the first term. Recall that $H_n^{-1}(1) = \max_{i=1,\dots,n} Y_i$, thus $H_n^{-1}(1) \to \infty$ as $n \to \infty$ a.s. Since $J_n$ is continuous in the second variable, it is sufficient to show that

$$\sup_{u\in[\delta,1]} \left|H_n^{-1}(1-u) - \phi^{-1}(u)\right| \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty. \tag{27}$$

To achieve this, let $A_n$ denote the 'average' measure

$$A_n(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\bigl(Y_n(i) < y\bigr) = 1 - \int_0^\infty \mathrm{e}^{-xy}\,\mathrm{d}E_n(x).$$

A theorem of Wellner [34], Theorem 1, states that if the sequence of measures $(A_n, n \geq 1)$ is tight, then the Prohorov distance between $H_n$ and $A_n$ converges a.s. to 0 and $n \to \infty$. In this case, since $E_n$ converges to $F$ uniformly, $A_n$ converges uniformly to $1-\phi$. Thus $H_n$ converges uniformly to $1-\phi$, and (27) follows. □

**Proof of Corollary 21.** When $E_n$ is the e.d.f. of $n$ i.i.d. picks from $F$, then (23) is satisfied a.s. by the Glivenko–Cantelli theorem. Thus, Theorem 23 implies Corollary 21. □

**Proof of Lemma 20.** Let $\phi$ be the Laplace transform of $F$. For $y > 0$,

$$\frac{\mathrm{d}\phi(y)^n}{\mathrm{d}y} = n\phi(y)^{n-1}\phi'(y).$$

By Corollary 8, we have

$$\frac{\mathbb{P}(X_n[1] \in \mathrm{d}x)}{xF(\mathrm{d}x)} = n\int_0^\infty \mathrm{e}^{-xy}\phi(y)^{n-1}\,\mathrm{d}y = \int_0^\infty \frac{\mathrm{e}^{-xy}}{\phi'(y)}\frac{\mathrm{d}}{\mathrm{d}y}\bigl(\phi(y)^n\bigr)\,\mathrm{d}y.$$

Apply integration by parts, the constant term is

$$\frac{\mathrm{e}^{-xy}}{\phi'(y)}\phi(y)^n\Big|_0^\infty = -\frac{1}{\phi'(0)} = \frac{1}{\mu}.$$

The integral term is

$$\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}y}\bigl(\mathrm{e}^{-xy}\phi'(y)\bigr)\bigl(\phi(y)\bigr)^n\,\mathrm{d}y.$$

The integrand is integrable for all $n$, thus

$$\lim_{n\to\infty}\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}y}\bigl(\mathrm{e}^{-xy}\phi'(y)\bigr)\bigl(\phi(y)\bigr)^n\,\mathrm{d}y = \int_0^\infty \lim_{n\to\infty}\frac{\mathrm{d}}{\mathrm{d}y}\bigl(\mathrm{e}^{-xy}\phi'(y)\bigr)\bigl(\phi(y)\bigr)^n\,\mathrm{d}y = 0.$$

Since $n \to \infty$, $\phi(y)^n \to 0$ for all $y > 0$. Therefore, $\lim_{n\to\infty} \frac{\mathbb{P}(X_n[1]\in\mathrm{d}x)}{xF(\mathrm{d}x)} = \frac{1}{\mu}$. □

### 5.2. Historical notes on induced order statistics and successive sampling

Induced order statistics were first introduced by David [8] and independently by Bhattacharya [3]. Typical applications stem from modeling an indirect ranking procedure, where subjects are ranked based on their $Y$-attributes although the real interest lies in ranking their $X$-attributes, which are difficult to obtain at the moment where the ranking is required.[1] For example, in cattle selection, $Y$ may represent the genetic makeup, for which the cattle are selected for breeding, and $X$ represents the milk yields of their female offspring. Thus a portion of this literature focuses on comparing distribution of induced order statistics to that of usual order statistics [9,16,24,35]. The most general statement on asymptotic distributions is obtained by Davydov and Egorov [10], who proved the functional central limit theorem and the functional law of the iterated logarithm for the process $S_{n,u}$ under tight assumptions. The functional central limit theorem for i.i.d. size-biased permutation is a special case of their results. Various versions of results in Section 5, including the functional central limit theorem, are also known in the successive sampling community [4,17,18,31,32]. For example, Bickel, Nair and Wang [4] proved Theorem 23 with convergence in probability when $E_n$ and $F$ have the same discrete support on finitely many values.

## 6. Poisson coupling of size-biased permutation and order statistics

Comparisons between the distribution of induced order statistics and order statistics of the same sequence have been studied in the literature [9,16,24,35]. However, finite i.i.d. size-biased permutation has the special feature that there exists an explicit coupling between these two sequences as described in Proposition 5. Using this fact, we now derive Theorem 25, which gives a Poisson coupling between the *last* $k$ size-biased terms $X_n^{\mathrm{rev}}[1], \ldots, X_n^{\mathrm{rev}}[k]$ and the $k$ smallest order statistics $X_n^{\uparrow}(1), \ldots, X_n^{\uparrow}(k)$ as $n \to \infty$. The existence of a Poisson coupling is not surprising, since the increasing sequence of order statistics $(X_n^{\uparrow}(1), X_n^{\uparrow}(2), \ldots)$ converges to points in a Poisson point process (p.p.p.) whose intensity measure depends on the behavior of $F$ near the infimum of its support, which is 0 in our case. This standard result in order statistics and extreme value theory dates back to Rényi [30], and can be found in [11].

### 6.1. Random permutations from Poisson scatter

Let $N(\cdot)$ be a Poisson scatter on $(0, \infty)^2$. Suppose $N(\cdot)$ has intensity measure $m$ such that for all $s, t \in (0, \infty)$

$$m\big((0, s) \times (0, \infty)\big) < \infty, \qquad m\big((0, \infty) \times (0, t)\big) < \infty.$$

[1] One often uses $X$ for the variable to be ordered, and $Y$ for the induced variable, with the idea that $Y$ is to be predicted. Here we use $X$ for the induced order statistics since $X_n[k]$ has been used for the size-biased permutation. The role of $X$ and $Y$ in our case is interchangeable, as evident when one writes $X_n(i)Y_n(i) = \varepsilon_n(i)$.

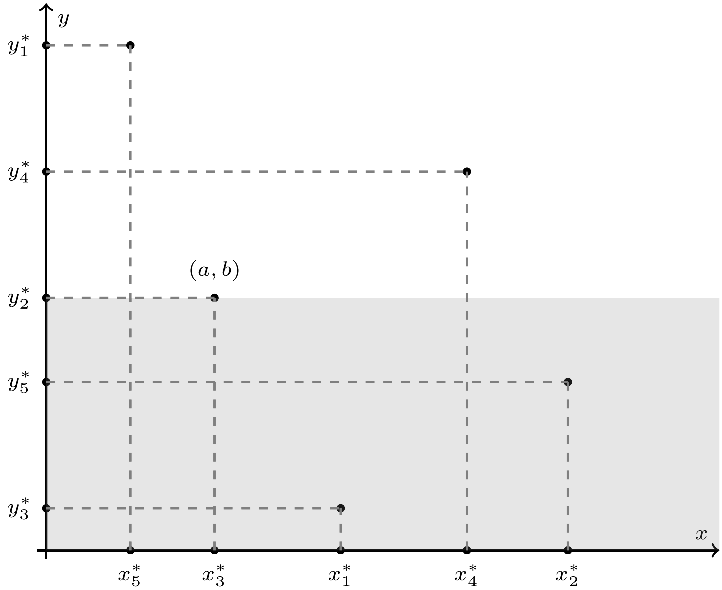


**Figure 1.** A point scatter on the plane. Here $J_1 = 5, J_2 = 3, J_3 = 1, J_4 = 4, J_5 = 2$, and $K_1 = 3$, $K_2 = 5, K_3 = 2, K_4 = 4, K_5 = 1$. The permutations $J$ and $K$ are inverses. Conditioned on $x^{\uparrow}(2) = a$ and $y^*(2) = b$, the number of points lying in the shaded region determines $J_2 - 1$.

Then one obtains a random permutation of $\mathbb{N}$ from ranking points $(x(i), y(i))$ in $N$ according to either the $x$ or $y$ coordinate. Let $x^*$ and $y^*$ denote the induced order statistics of the sequence $x$ and $y$ obtained by ranking points by their $y$ and $x$ values in increasing order, respectively. For $j, k = 1, 2, \ldots$, define sequences of integers $(K_j)$, $(J_k)$ such that $x^{\uparrow}(J_k) = x^*(k)$, $y^{\uparrow}(K_j) = y^*(j)$; see Figure 1.

For $j \geq 1$, conditioned on $x(j) = s$, $y^*(j) = t$,

$$K_j - 1 \stackrel{d}{=} \operatorname{Poisson}\bigl(m\bigl((s, \infty) \times (0, t)\bigr)\bigr) + \operatorname{Binomial}\left(j - 1, \frac{m((0, s) \times (0, t))}{m((0, s) \times (0, \infty))}\right), \tag{28}$$

where the two random variables involved are independent. Similarly, for $k \geq 1$, conditioned on $x_k^* = s$, $y(k) = t$,

$$J_k - 1 \stackrel{d}{=} \operatorname{Poisson}\bigl(m\bigl((0, s) \times (t, \infty)\bigr)\bigr) + \operatorname{Binomial}\left(k - 1, \frac{m((0, s) \times (0, t))}{m((0, \infty) \times (t, \infty))}\right), \tag{29}$$

where the two random variables involved are independent. When $m$ is a product measure, it is possible to compute the marginal distribution of $K_j$ and $J_k$ explicitly for given $j, k \geq 1$.

Random permutations from Poisson scatters appeared in [26], Section 4. When $X^{\downarrow}$ is the sequence of ranked jumps of a subordinator, these authors noted that one can couple the size-biased permutation with the order statistics via the following p.p.p.

$$N(\cdot) := \sum_{k\geq 1} \mathbf{1}\big[\big(X[k], Y^{\uparrow}(k)\big) \in \cdot\big] = \sum_{k\geq 1} \mathbf{1}\big[\big(X^{\downarrow}(k), Y(k)\big) \in \cdot\big], \tag{30}$$

where $Y$ is an independent sequence of standard exponentials. Thus, $N(\cdot)$ has intensity measure $m(\mathrm{d}x\,\mathrm{d}y) = x\mathrm{e}^{-xy}\Lambda(\mathrm{d}x)\,\mathrm{d}y$. The first expression in (30) defines a scatter of $(x, y)$ values in the plane listed in increasing $y$ values, and the second represents the same scatter listed in decreasing $x$ values. Since $\sum_{i\geq 1} X^{\downarrow}(i) < \infty$ a.s., the $x$-marginal of the points in (30) has the distribution of the size-biased permutation $X^*$, since it prescribes the joint distribution of the first $k$ terms $X[1], \ldots, X[k]$ of $X^*$ for any finite $k$. Perman, Pitman and Yor used this p.p.p. representation to generalize size-biased permutation to $h$-biased permutation, where the 'size' of a point $x$ is replaced by an arbitrary strictly positive function $h(x)$; see [26], Section 4.

## 6.2. Poisson coupling in the limit

Our theorem states that in the limit, finite i.i.d. size-biased permutation is a form of random permutation obtained from a Poisson scatter with a certain measure, which, under a change of coordinate, is given by (36). Before stating the theorem, we need some technical results. The distribution of the last few size-biased picks depends on the behavior of $F$ near 0, the infimum of its support. We shall consider the case where $F$ has 'power law' near 0, like that of a Gamma distribution.

**Lemma 24.** *Suppose $F$ is supported on $(0, \infty)$ with Laplace transform $\phi$. Let $u = \phi(y)$, $X_u$ a random variable distributed as $G_u(\mathrm{d}x)$ defined in (8). For $\lambda, a > 0$,*

$$F(x) \sim \frac{\lambda^a x^a}{\Gamma(a+1)} \qquad \text{as } x \to 0, \tag{31}$$

*if and only if,*

$$\phi(y) \sim \lambda^a / y^a \qquad \text{as } y \to \infty. \tag{32}$$

*Furthermore*, (31) *implies*

$$u^{-1/a} X_u \xrightarrow{d} \mathrm{gamma}(a+1, \lambda) \qquad \text{as } u \to 0. \tag{33}$$

**Proof.** The equivalence of (31) and (32) follows from a version of Karamata Tauberian theorem [5], Section 1.7. Assume (31) and (32). We shall prove (33) by looking at the Laplace transform of the non-size-biased version $X'_u$, which has distribution $F_u$. For $\theta \geq 0$,

$$E\big(\exp\big(-\theta X'_u\big)\big) = \int_0^\infty u^{-1} \exp(-yx - \theta x) F(\mathrm{d}x) = u^{-1}\phi(y+\theta) = \frac{\phi(y+\theta)}{\phi(y)}. \tag{34}$$

Now as $y \to \infty$ and $u = \phi(y) \to 0$, for each fixed $\eta > 0$, (32) implies

$$E\bigl(\exp\bigl(-\eta u^{-1/a} X'_u\bigr)\bigr) = \frac{\phi(y + \eta\phi(y)^{-1/a})}{\phi(y)} \sim \frac{\lambda^a (y + \eta\lambda^{-1} y)^{-a}}{\lambda^a y^{-a}} = \left(\frac{\lambda}{\lambda+\eta}\right)^a.$$

That is to say

$$u^{-1/a} X'_u \xrightarrow{d} \text{gamma}(a, \lambda). \tag{35}$$

Since $\phi$ is differentiable, (34) implies $E(X'_u) = \phi'(y)/\phi(y)$. Now $\phi$ has an increasing derivative $\phi'$, thus (32) implies $\phi'(y) \sim a\lambda^a / y^{a+1}$ as $y \to \infty$. Therefore,

$$u^{-1/a} E\bigl(X'_u\bigr) = \frac{\phi'(y)}{\phi(y)^{1+1/a}} \to \frac{a}{\lambda},$$

which is the mean of a gamma$(a, \lambda)$ random variable. Thus, the random variables $u^{-1/a} X'_u$ are uniformly integrable, so for any bounded continuous function $h$, we can compute

$$E\bigl(h\bigl(u^{-1/a} X_u\bigr)\bigr) = \frac{E[(u^{-1/a} X'_u) h(u^{-1/a} X'_u)]}{u^{-1/a} E(X'_u)} \to \frac{E[\gamma_{a,\lambda} h(\gamma_{a,\lambda})]}{E(\gamma_{a,\lambda})} = E\bigl(h(\gamma_{a+1,\lambda})\bigr),$$

where $\gamma_{b,\lambda}$ is a gamma$(b, \lambda)$ random variable. This proves (33). $\square$

We now present the analogue of (43) for the last few size-biased picks $X_n^{\text{rev}}[1], \dots, X_n^{\text{rev}}[k]$ and the promised Poisson coupling.

**Theorem 25.** *Suppose that* (31) *holds for some* $\lambda, a > 0$. *Let* $N(\cdot)$ *be a Poisson scatter on* $(0,\infty)^2$ *with intensity measure*

$$\frac{\mu(\mathrm{d}s\,\mathrm{d}t)}{\mathrm{d}s\,\mathrm{d}t} = \frac{1}{a}\Gamma(a+1)^{1/a}(s/t)^{1/a}\exp\bigl\{-\bigl(\Gamma(a+1)s/t\bigr)^{1/a}\bigr\}. \tag{36}$$

*By ranking points in either increasing* $T$ *or* $S$ *coordinate, one can write*

$$N(\cdot) = \sum_k \mathbf{1}\bigl[\bigl(S(k), T^{\uparrow}(k)\bigr) \in \cdot\bigr] = \sum_j \mathbf{1}\bigl[\bigl(S^{\uparrow}(j), T(j)\bigr) \in \cdot\bigr]. \tag{37}$$

*Define* $\Psi_{a,\lambda}(s) = s^{1/a}\Gamma(a+1)^{1/a}/\lambda$. *Define a sequence of random variables* $\xi$ *via*

$$\xi(k) = \Psi_{a,\lambda}\bigl(S^{\uparrow}(k)\bigr), \tag{38}$$

*and let* $\xi^*$ *be its reordering defined* $\xi^*(k) = \Psi_{a,\lambda}(S(k))$. *Then jointly as* $n \to \infty$,

$$n^{1/a} X_n^{\uparrow} \xrightarrow{f.d.d.} \xi, \tag{39}$$

$$n^{1/a}\bigl(X_n^*\bigr)^{\text{rev}} \xrightarrow{f.d.d.} \xi^*. \tag{40}$$

*In particular, for each* $n$, *let* $J_n = (J_{nk}, 1 \le k \le n)$ *be the permutation of* $\{1, \dots, n\}$ *defined by* $X_n^{\text{rev}}[k] = X_n(J_{nk})$. *As* $n \to \infty$,

$$(J_{nk}, 1 \le k \le n) \xrightarrow{f.d.d.} (J_k : 1 \le k < \infty), \tag{41}$$

*where* $J_k$ *is the random permutation of* $\{1, 2, \dots\}$, *defined by*

$$\xi^*(k) = \xi(J_k) \tag{42}$$

*for* $k = 1, 2, \dots,$ *and the f.d.d. convergence in* (39), (40), (41) *all hold jointly.*

In other words, the Poisson point process $N(\cdot)$ defined in (37) with measure (36) defines a random permutation $(J_k)$ of $\mathbb{N}$ and its inverse $(K_j)$. Theorem 25 states that $(J_k)$ is precisely the limit of the random permutation induced by the size-biased permutation of a sequence of $n$ i.i.d. terms from $F$. Furthermore, to obtain the actual sequence of size-biased permutation, one only needs to apply the *deterministic* transformation $\Psi_{a,\lambda}$ to the sequence of $s$-marginals of points in $N(\cdot)$, ranked according to their $t$-values. The sequence of increasing order statistics can be obtained by applying the transformation $\Psi_{a,\lambda}$ to the $s$-marginals ranked in increasing order.

**Proof of Theorem 25.** By Lemma 24, it is sufficient to prove the theorem for the case $F$ is gamma$(a, \lambda)$. First, we check that the sequence on the right-hand side of (39) and (40) have the right distribution. Indeed, by standard results in order statistics [11], Theorem 2.1.1, as $n \to \infty$, the sequence $n^{1/a} X_n^{\uparrow}$ converges (f.d.d.) to the sequence $\widetilde{\xi}$, where

$$\widetilde{\xi}(k) = \bigl(S^{\uparrow}(k)\bigr)^{1/a} \Gamma(a+1)^{1/a}/\lambda = \Psi_{a,\lambda}\bigl(S^{\uparrow}(k)\bigr), \tag{43}$$

where $S^{\uparrow}(k) = \varepsilon_1 + \cdots + \varepsilon_k$ for $\varepsilon_i$ i.i.d. standard exponentials. Similarly, by Proposition 10 and law of large numbers, the sequence $n^{1/a}(X_n^*)^{\text{rev}}$ converges (f.d.d.) to the sequence $\widetilde{\xi}^*$, where

$$\widetilde{\xi}^*(k) = \bigl(T^{\uparrow}(k)\bigr)^{1/a} \gamma_k/\lambda,$$

where $T^{\uparrow}(k) = \varepsilon'_1 + \cdots + \varepsilon'_k$ for i.i.d. standard exponentials $\varepsilon'_i$, and $\gamma_k$, $k = 1, \dots, n$ are i.i.d. gamma$(a+1, 1)$, independent of the $T(k)$. By direct computation, we see that $\widetilde{\xi} \stackrel{d}{=} \xi$ and $\widetilde{\xi}^* \stackrel{d}{=} \xi^*$. The dependence between the two sequences $S$ and $T$ comes from Proposition 5, which tells us that $S(k)$ is the term $S^{\uparrow}(j)$ that is paired with $T^{\uparrow}(k)$ in our Poisson coupling. Observe $\Psi_{a,\lambda}$ has inverse function $\Psi_{a,\lambda}^{-1}(x) = \lambda^a x^a / \Gamma(a+1)$. Thus applying (43), we have

$$S(k) = \Psi_{a,\lambda}^{-1}\bigl(\widetilde{\xi}^*(k)\bigr) = \lambda^a \bigl[\widetilde{\xi}^*(k)\bigr]^a / \Gamma(a+1) = T(k)\gamma_k^a / \Gamma(a+1). \tag{44}$$

Comparing (43) and (44) gives a pairing between $S(k)$ and $S^{\uparrow}(k)$, and hence $T^{\uparrow}(k)$ and $S^{\uparrow}(k)$, via $\widetilde{\xi}(k)$ and $\widetilde{\xi}^*(k)$. Hence, we obtain another definition of $J_k$ equivalent to (42):

$$S(k) = S^{\uparrow}(J_k).$$

Let $T(j)$ be the $T$ value corresponding to the order statistic $S^{\uparrow}(j)$ of the sequence $S$. That is,

$$T(j) = T^{\uparrow}(K_j),$$

where $(K_j)$ is a random permutation of the positive integers. By (44), $(J_k)$ is the inverse of $(K_j)$. Together with (39) and (40), this implies (41), proving the last statement. The intensity measure $\mu$ comes from direct computation. $\square$

Marginal distributions of the random permutation $(J_k)$ and its inverse $(K_j)$ are given in (28) and (29). Note that for $k = 1, 2, \ldots,$

$$S(k) = T^{\uparrow}(k)\gamma_k^a / \Gamma(a+1)$$

for i.i.d. $\gamma_k$ distributed as gamma$(a+1, 1)$, independent of the sequence $T^{\uparrow}$, and

$$T(k) = \Gamma(a+1) S^{\uparrow}(k) \tilde{\varepsilon}_k^{-a}$$

for i.i.d. standard exponentials $\varepsilon_k$, independent of the sequence $(S(k))$ but not of the $\gamma_k$. Since the projection of a Poisson process is Poisson, the $s$ and $t$-marginal of $\mu$ is just Lebesgue measure, as seen in the proof. Thus by conditioning on either $S^{\uparrow}(k)$ or $T^{\uparrow}(k)$, one can evaluate (28) and (29) explicitly. In particular, by a change of variable $r = \Gamma(a+1)^{1/a}(s/t)^{1/a}$, one can write $\mu$ in product form. This leads to the following.

**Proposition 26.** *For $j \geq 1$, conditioned on $S^{\uparrow}(j) = s$, $T(j) = \Gamma(a+1)sr^{-a}$ for some $r > 0$, $K_j - 1$ is distributed as*

$$\mathrm{Poisson}\big(m(s,r)\big) + \mathrm{Binomial}\big(j-1, p(s,r)\big) \tag{45}$$

*with*

$$m(s,r) = asr^{-a} \int_r^{\infty} x^{a-1} \mathrm{e}^{-x}\, \mathrm{d}x \tag{46}$$

*and*

$$p(s,r) = as^{2/a-2} r^{a-2} \int_0^r x^{a-1} \mathrm{e}^{-x}\, \mathrm{d}x, \tag{47}$$

*where the* Poisson *and* Binomial *random variables are independent. Similarly, for $k \geq 1$, conditioned on $T^{\uparrow}(k) = t$, $S(k) = tr^a / \Gamma(a+1)$ for some $r > 0$, $J_k - 1$ is distributed as*

$$\mathrm{Poisson}\big(m'(t,r)\big) + \mathrm{Binomial}\big(k-1, p'(t,r)\big) \tag{48}$$

*with*

$$m'(t,r) = t\left(\frac{r^a + a\int_r^{\infty} x^{a-1}\mathrm{e}^{-x}\,\mathrm{d}x}{\Gamma(a+1)} - 1\right) \tag{49}$$

*and*

$$p'(t,r) = \Gamma(a+1)^{1-2/a} a t^{2/a-2} r^{a-1/a} \int_0^r x^{a-1} \mathrm{e}^{-x}\, \mathrm{d}x, \tag{50}$$

*where the* Poisson *and* Binomial *random variables are independent.*

**Proposition 27 (Marginal distributions of $K_1$ and $J_1$).** *Suppose that* (31) *holds for some $\lambda > 0$ and $a = 1$. Then the distribution of $K_1$, the $k$ such that $\xi(1) = \xi^*(k)$, is a mixture of geometric distributions, and so is that for $J_1$, the $j$ such that $\xi^*(1) = \xi(j)$. In particular,*

$$P(K_1 = k) = \int_0^\infty p_r q_r^{k-1} \mathrm{e}^{-r}\, \mathrm{d}r, \tag{51}$$

*where $p_r = r/(r + \mathrm{e}^{-r})$, $q_r = 1 - p_r$, and*

$$P(J_1 = j) = \int_0^\infty \tilde{p}_r \tilde{q}_r^{j-1} r \mathrm{e}^{-r}\, \mathrm{d}r, \tag{52}$$

*where $\tilde{p}_r = 1/(r + \mathrm{e}^{-r})$, $\tilde{q}_r = 1 - \tilde{p}_r$.*

**Proof.** When $a = 1$, $\int_r^\infty t^{a-1} \mathrm{e}^{-t}\, \mathrm{d}t = \mathrm{e}^{-r}$. Substitute to (46) and (49) give

$$m(s,r) = s r^{-1} \mathrm{e}^{-r}, \qquad m'(t,r) = t\bigl(r - 1 + \mathrm{e}^{-r}\bigr).$$

By a change of variable, (36) becomes

$$\frac{\mu(\mathrm{d}s\, \mathrm{d}r)}{\mathrm{d}s\, \mathrm{d}r} = s\mathrm{e}^{-r}, \qquad \frac{\mu(\mathrm{d}t\, \mathrm{d}r)}{\mathrm{d}t\, \mathrm{d}r} = t r \mathrm{e}^{-r}.$$

Thus, *conditioned on $s$ and $r$*, $K_1 - 1$ is distributed as the number of points in a p.p.p. with rate $r^{-1}\mathrm{e}^{-r}$ before the first point in a p.p.p. with rate 1. This is the geometric distributions on $(0, 1, \ldots)$ with parameter $p_r = 1/(1 + r^{-1}\mathrm{e}^{-r})$. Since the marginal density of $r$ is $\mathrm{e}^{-r}$, integrating out $r$ gives (51). The computation for the distribution of $J_1$ is similar. $\square$

One can check that each of (51) and (52) sum to 1. We conclude with a 'fun' computation. Suppose that (31) holds for some $\lambda > 0$ and $a = 1$. That is, $F$ behaves like an exponential c.d.f. near 0. By Proposition 27, $E(J_1) = 9/4$ and $E(K_1) = \infty$. That is, the last size-biased pick is expected to be almost the second smallest order statistic, while the smallest order statistic is expected to be picked infinitely earlier on in a successive sampling scheme(!). The probability that the last species to be picked in a successive sampling scheme is also the one of smallest species size is

$$\lim_{n\to\infty} P\bigl(X_n^{\mathrm{rev}}[1] = X_n(1)\bigr) = P\bigl(\xi^*(1) = \xi(1)\bigr) = P(J_1 = 1) = P(K_1 = 1)$$
$$= \int_0^\infty \frac{r\mathrm{e}^{-r}}{r + \mathrm{e}^{-r}}\, \mathrm{d}r = 1 - \int_0^1 \frac{u}{u - \log u}\, \mathrm{d}u \approx 0.555229.$$

## 7. Summary

This paper reviewed and complemented results on the exact and asymptotic distribution of the size-biased permutation of finitely many independent and identically distributed positive terms. Our setting lies in the intersection between induced order statistics, size-biased permutation of ranked jumps of a subordinator, and successive sampling. We discussed size-biased permutation from these different viewpoints and obtained simpler proofs of known results. Our main contribution, Theorem 25, gives a Poisson coupling between the asymptotic distribution of the last few terms of a size-biased permutation and its few smallest order statistics.

We thank two anonymous referees for their careful reading and constructive comments.

## References

[1] Barouch, E. and Kaufman, G.M. (1976). Probabilistic modelling of oil and gas discovery. In *Energy: Mathematics and Models* 133–150. Philadelphia: SIAM.

[2] Bertoin, J. (1999). Subordinators: Examples and applications. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1997). *Lecture Notes in Math*. **1717** 1–91. Berlin: Springer. MR1746300

[3] Bhattacharya, P.K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist*. **2** 1034–1039. MR0386100

[4] Bickel, P.J., Nair, V.N. and Wang, P.C.C. (1992). Nonparametric inference under biased sampling from a finite population. *Ann. Statist*. **20** 853–878. MR1165596

[5] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871

[6] Chaumont, L. and Yor, M. (2003). *Exercises in Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics* **13**. Cambridge: Cambridge Univ. Press. MR2016344

[7] Cheng, W., Dembczyński, K. and Hüllermeier, E. (2010). Label ranking methods based on the Plackett–Luce model. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 215–222, Haifa, Israel, June 2010. Omnipress.

[8] David, H.A. (1973). Concomitants of order statistics. *Bull. Inst. Internat. Statist*. **45** 295–300. MR0373149

[9] David, H.A., O'Connell, M.J. and Yang, S.S. (1977). Distribution and expected value of the rank of a concomitant of an order statistic. *Ann. Statist*. **5** 216–223. MR0445673

[10] Davydov, Y. and Egorov, V. (2001). Functional CLT and LIL for induced order statistics. In *Asymptotic Methods in Probability and Statistics with Applications* (*St. Petersburg*, 1998). *Stat. Ind. Technol*. 333–349. Boston, MA: Birkhäuser. MR1890337

[11] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156

[12] Engen, S. (1978). *Stochastic Abundance Models with Emphasis on Biological Communities and Species Diversity*. London: Chapman & Hall. MR0515721

[13] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biology* **3** 87–112; erratum, ibid. 3 (1972), 240; erratum, ibid. 3 (1972), 376. MR0325177

[14] Gnedin, A., Haulk, C. and Pitman, J. (2010). Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In *Probability and Mathematical Genetics*. *London Mathematical Society Lecture Note Series* **378** 264–298. Cambridge: Cambridge Univ. Press. MR2744243

[15] Gnedin, A.V. (1998). On convergence and extensions of size-biased permutations. *J. Appl. Probab*. **35** 642–650. MR1659532

[16] Goel, P.K. and Hall, P. (1994). On the average difference between concomitants and order statistics. *Ann. Probab.* **22** 126–144. MR1258869
[17] Gordon, L. (1983). Successive sampling in large finite populations. *Ann. Statist.* **11** 702–706. MR0696081
[18] Holst, L. (1973). Some limit theorems with applications in sampling theory. *Ann. Statist.* **1** 644–658. MR0365836
[19] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1876169
[20] Kingman, J.F.C. (1978). Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A* **361** 1–20. MR0526801
[21] Kingman, J.F.C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* (2) **18** 374–380. MR0509954
[22] Lukacs, E. (1955). A characterization of the Gamma distribution. *Ann. Math. Statist.* **26** 319–324. MR0069408
[23] McCloskey, J.W. (1965). A model for the distribution of individuals by species in an environment. Ph.D. thesis, Michigan State Univ.
[24] Nagaraja, H.N. and David, H.A. (1994). Distribution of the maximum of concomitants of selected order statistics. *Ann. Statist.* **22** 478–494. MR1272095
[25] Patil, G.P. and Taillie, C. (1977). Diversity as a concept and its implications for random communities. *Bull. Inst. Internat. Statist.* **47** 497–515. MR0617593
[26] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448
[27] Pitman, J. (1996). Random discrete distributions invariant under size-biased permutation. *Adv. in Appl. Probab.* **28** 525–539. MR1387889
[28] Pitman, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Berlin: Springer. MR2245368
[29] Plackett, R.L. (1968). Random permutations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **30** 517–534. MR0247695
[30] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.* **4** 191–231. MR0061792
[31] Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I, II. *Ann. Math. Statist.* **43** 373–397; ibid. 43 (1972), 748–776. MR0321223
[32] Sen, P.K. (1979). Invariance principles for the coupon collector's problem: A martingale approach. *Ann. Statist.* **7** 372–380. MR0520246
[33] Sturmfels, B. and Welker, V. (2012). Commutative algebra of statistical ranking. *J. Algebra* **361** 264–286. MR2921622
[34] Wellner, J.A. (1981). A Glivenko–Cantelli theorem for empirical measures of independent but non-identically distributed random variables. *Stochastic Process. Appl.* **11** 309–312. MR0622172
[35] Yang, S.S. (1977). General distribution theory of the concomitants of order statistics. *Ann. Statist.* **5** 996–1002. MR0501519

*Received October 2012 and revised March 2014*