Regenerative partition structures *

Alexander Gnedin
Utrecht University
e-mail gnedin@math.uu.nl
and
Jim Pitman
University of California, Berkeley
pitman@stat.Berkeley.EDU

Submitted: Aug 6, 2004; Accepted: Nov 4, 2004; Published: Jan 7, 2005
Mathematics Subject Classifications: 60G09, 60C05.
Keywords: partition structure, deletion kernel, regenerative composition structure

Abstract

A *partition structure* is a sequence of probability distributions for π_n, a random partition of n, such that if π_n is regarded as a random allocation of n unlabeled balls into some random number of unlabeled boxes, and given π_n some x of the n balls are removed by uniform random deletion without replacement, the remaining random partition of $n-x$ is distributed like π_{n-x}, for all $1 \le x \le n$. We call a partition structure *regenerative* if for each n it is possible to delete a single box of balls from π_n in such a way that for each $1 \le x \le n$, given the deleted box contains x balls, the remaining partition of $n-x$ balls is distributed like π_{n-x}. Examples are provided by the Ewens partition structures, which Kingman characterised by regeneration with respect to deletion of the box containing a uniformly selected random ball. We associate each regenerative partition structure with a corresponding regenerative composition structure, which (as we showed in a previous paper) is associated in turn with a regenerative random subset of the positive halfline. Such a regenerative random set is the closure of the range of a subordinator (that is an increasing process with stationary independent increments). The probability distribution of a general regenerative partition structure is thus represented in terms of the Laplace exponent of an associated subordinator, for which exponent an integral representation is provided by the Lévy-Khintchine formula. The extended Ewens family of partition structures, previously studied by Pitman and Yor, with two parameters (α, θ), is characterised for $0 \le \alpha < 1$ and $\theta > 0$ by regeneration with respect to deletion of each distinct part of size x with probability proportional to $(n-x)\tau + x(1-\tau)$, where $\tau = \alpha/(\alpha+\theta)$.

*Research supported in part by N.S.F. Grant DMS-0405779

1 Introduction and main results

This paper is concerned with sequences of probability distributions for random partitions π_n of a positive integer n. We may represent π_n as a sequence of integer-valued random variables

$$\pi_n = (\pi_{n,1}, \pi_{n,2}, \ldots) \quad \text{with} \quad \pi_{n,1} \geq \pi_{n,2} \geq \cdots \geq 0$$

so $\pi_{n,i}$ is the size of the ith largest part of π_n, and $\sum_i \pi_{n,i} = n$. We may also treat π_n as a multiset of positive integers with sum n, regarding π_n as a random allocation of n unlabeled balls into some random number of unlabeled boxes, with each box containing at least one ball. We call π_n *regenerative* if it is possible to delete a single box of balls from π_n in such a way that for each $1 \leq x \leq n$, given the deleted box contained x balls, the remaining partition of $n - x$ balls is distributed as if x balls had been deleted from π_n by uniform random sampling without replacement. We spell this out more precisely in Definition 1 below.

To be more precise, we assume that π_n is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is rich enough to allow various further randomisations considered below, including the choice of some *random part* $X_n \in \pi_n$, meaning that X_n is one of the positive integers in the multiset π_n with sum n. The distribution of π_n is then specified by some *partition probability function*

$$p(\lambda) := \mathbb{P}(\pi_n = \lambda) \qquad (\lambda \vdash n) \tag{1}$$

where the notation $\lambda \vdash n$ indicates that λ is a partition of n. The joint distribution of π_n and X_n is determined by the partition probability function p and some *deletion kernel* $d = d(\lambda, x), \lambda \vdash n, 1 \leq x \leq n$, which describes the conditional distribution of X_n given π_n, according to the formula

$$p(\lambda)d(\lambda, x) = \mathbb{P}(\pi_n = \lambda, X_n = x). \tag{2}$$

The requirement that X_n is a part of π_n makes $d(\lambda, x) = 0$ unless x is a part of λ, and

$$\sum_{x \in \lambda} d(\lambda, x) = 1 \tag{3}$$

for all partitions λ of n. Without loss of generality, we suppose further that π_n is the sequence of ranked sizes of classes of some random partition Π_n of the set $[n] := \{1, \ldots, n\}$, where conditionally given π_n all possible values of Π_n are equally likely. (Here and throughout the paper, we use the term *ranked* to mean that the terms of a sequence are weakly decreasing.) Equivalently, Π_n is an exchangeable random partition of $[n]$ as defined in [15]. For $1 \leq m \leq n$ let Π_m be the restriction of Π_n to $[m]$, and let π_m be the sequence of ranked sizes of classes of Π_m. We say that the random partition π_m of m is *derived from* π_n *by random sampling*, and call the distributions of the random partitions π_m for $1 \leq m \leq n$ *sampling consistent*. A *partition structure* is a function $p(\lambda)$ as in (1) for a sampling consistent sequence of distributions of π_n for $n = 1, 2, \ldots$. This concept was introduced by Kingman [10], who established a one-to-one correspondence between partition structures p and distributions for a sequence of nonnegative random variables

$V_1, V_2, \ldots$ with $V_1 \geq V_2 \geq \ldots$ and $\sum_i V_i \leq 1$. In Kingman's *paintbox representation* of p, the random partition π_n of n is constructed as follows from (V_k) and a sequence of independent random variables U_i with uniform distribution on $[0,1]$, with (U_i) and (V_k) are independent: π_n as in (1) is defined to be the sequence of ranked sizes of blocks of the partition of $[n]$ generated by a random equivalence relation $\sim$ on positive integers, with $i \sim j$ if and only if either $i = j$ or both U_i and U_j fall in I_k for some k, where the I_k are some disjoint random sub-intervals of $[0,1]$ of lengths V_k. See also [15] and papers cited there for further background.

Definition 1 Call a random partition π_n of n *regenerative*, if it is possible to select a random part X_n of π_n in such a way that for each $1 \leq x < n$, conditionally given that $X_n = x$ the remaining partition of $n-x$ is distributed according to the unconditional distribution of π_{n-x} derived from π_n by random sampling. Then π_n may also be called *regenerative with respect to deletion of* X_n, or *regenerative with respect to* d if the conditional law of X_n given π_n is specified by a deletion kernel d as in (2). Call a partition structure p *regenerative* if the corresponding π_n is regenerative for each $n = 1, 2, \ldots$.

According to this definition, π_n is regenerative with respect to deletion of some part $X_n \in \pi_n$ if and only if for each partition λ of n and each part $x \in \lambda$,

$$\mathbb{P}(\pi_n = \lambda, X_n = x) = \mathbb{P}(X_n = x)\,\mathbb{P}(\pi_{n-x} = \lambda - \{x\}) \qquad (\lambda \vdash n) \tag{4}$$

where $\lambda - \{x\}$ is the partition of $n - x$ obtained by deleting the part x from λ, and π_{n-x} is derived from π_n by sampling. Put another way, π_n is regenerative with respect to a deletion kernel d iff

$$p(\lambda)d(\lambda, x) = q(n, x)p(\lambda - \{x\}), \qquad x \in \lambda \qquad (\lambda \vdash n) \tag{5}$$

where $p(\mu) := \mathbb{P}(\pi_m = \mu)$ for $\mu \vdash m$ and $1 \leq m \leq n$ and

$$q(n, x) := \sum_{\{\lambda \vdash n \,:\, x \in \lambda\}} d(\lambda, x)p(\lambda) = \mathbb{P}(X_n = x) \qquad (1 \leq x \leq n) \tag{6}$$

is the unconditional probability that the deletion rule removes a part of size x from π_n.

A well known partition structure is obtained by letting π_n be the partition of n generated by the sizes of cycles of a uniformly distributed random permutation σ_n of $[n]$. If X_n is the size of the cycle of σ_n containing 1, then π_n is regenerative with respect to deletion of X_n, because given $X_n = x$ the remaining partition of $n - x$ is generated by the cycles of a uniform random permutation of a set of size $n - x$. In this example, the unconditional distribution $q(n, \cdot)$ of X_n is uniform on $[n]$. The deletion kernel is

$$d(\lambda, x) = \frac{x a_x(\lambda)}{n} \qquad (\lambda \vdash n) \tag{7}$$

where $a_x(\lambda)$ is the number of parts of λ of size x, so $\sum_{x=1}^{n} x a_x(\lambda) = n$. More generally, a part X_n is chosen from a random partition π_n of n according to (7) may be called a *size-biased part* of π_n. According to a well known result of Kingman [10], if a partition structure is regenerative with respect to deletion of a size-biased part, then it is governed by the Ewens sampling formula

$$p(\lambda) = \frac{n!\theta^\ell}{(\theta)_{n\uparrow 1}} \prod_r \frac{1}{r^{a_r} a_r!} \tag{8}$$

for some parameter $\theta \geq 0$, where λ is encoded by its multiplicities $a_r = a_r(\lambda)$ for $r = 1, 2, \ldots$, with

$$\ell = \Sigma a_r \, , n = \Sigma \, r a_r \tag{9}$$

and

$$(\theta)_{n\uparrow b} := \prod_{i=1}^{n} (\theta + (i-1)b).$$

The case $\theta = 1$ gives the distribution of the partition generated by cycles of a uniform random permutation. Pitman [11, 12] introduced a two-parameter extension of the Ewens family of partition structures, defined by the sampling formula

$$p(\lambda) = \frac{n!(\theta)_{\ell\uparrow\alpha}}{(\theta)_{n\uparrow 1}} \prod_r \left(\frac{(1-\alpha)_{r-1\uparrow 1}}{r!} \right)^{a_r} \frac{1}{a_r!} \tag{10}$$

for suitable parameters (α, θ), including

$$\{(\alpha, \theta) : 0 \leq \alpha \leq 1, \, \theta \geq 0\} \tag{11}$$

where boundary cases are defined by continuity. See [15] for a review of various applications of this formula. The result of [4, Theorem 8.1 and Corollary 8.2] shows that each (α, θ) partition structure with parameters subject to (11) is regenerative with respect to the deletion kernel

$$d(\lambda, r) = \frac{a_r}{n} \, \frac{(n-r)\tau + r(1-\tau)}{1 - \tau + (\ell - 1)\tau} \, , \tag{12}$$

where $\tau = \alpha/(\alpha + \theta) \in [0, 1]$, and (3) follows easily from (9). In Section 2 we establish:

Theorem 2 *For each $\tau \in [0,1]$, the only partition structures which are regenerative with respect to the deletion kernel* (12) *are the (α, θ) partition structures subject to* (11) *with $\alpha/(\alpha + \theta) = \tau$.*

The following three cases are of special interest:

Size-biased deletion This is the case $\tau = 0$: each part r is selected with probability proportional to r. Here, and in following descriptions, we assume that the parts of a partition are labeled in some arbitrary way, to distinguish parts of equal size. In particular, if π_n is the partition of n derived from an exchangeable random partition Π_n of $[n]$, then for each $i \in [n]$ the size $X_n(i)$ of the part of Π_n containing i defines a size-biased pick from the parts of π_n. Theorem 2 in this case reduces to Kingman's characterisation of the Ewens family of $(0, \theta)$ partition structures. Section 7 compares Theorem 2 with another characterisation of (α, θ) partition structures provided by Pitman [13] in terms of a size-biased random permutation of parts defined by iterated size-biased deletion.

Unbiased (uniform) deletion This is the case $\tau = 1/2$: given that π_n has ℓ parts, each part is chosen with probability $1/\ell$. Iteration of this operation puts the parts of π_n in an exchangeable random order. In this case, the conclusion of Theorem 2 is that the (α, α) partition structures for $0 \leq \alpha \leq 1$ are the only partition structures invariant under uniform deletion. This conclusion can also be drawn from Theorem 10.1 of [4]. As shown in [12, 14], the (α, α) partition structures are generated by sampling from the interval partition of $[0, 1]$ into excursion intervals of a Bessel bridge of dimension $2 - 2\alpha$. The case $\alpha = 1/2$ corresponds to excursions of a standard Brownian bridge.

Cosize-biased deletion In the case $\tau = 1$, each part of size r is selected with probability proportional to the size $n - r$ of the remaining partition. The conclusion of Theorem 2 in this case is that the $(\alpha, 0)$ partition structures for $0 \leq \alpha \leq 1$ are the only partition structures invariant under this operation. As shown in [12, 14], these partition structures are generated by sampling from the interval partition generated by excursion intervals of an unconditioned Bessel process of dimension $2 - 2\alpha$. The case $\alpha = 1/2$ corresponds to excursions of a standard Brownian motion.

The next theorem, which is proved in Section 3, puts Theorem 2 in a more general context:

Theorem 3 *For each probability distribution $q(n, \cdot)$ on $[n]$, there exists a unique joint distribution of a random partition π_n of n and a random part X_n of π_n such that X_n has distribution $q(n, \cdot)$ and π_n is regenerative with respect to deletion of X_n.*

Let $\pi_m, 1 \leq m \leq n$ be derived from π_n by random sampling. Then for each $1 \leq m \leq n$ the random partition π_m is regenerative with respect to deletion of some part X_m, whose distribution $q(m, \cdot)$ is that of H_m given $H_m > 0$, where H_m is the number of balls in the sample of size m which fall in some particular box containing X_n balls in π_n.

The main point of this theorem is its implication that if π_n is regenerative with respect to deletion of X_n according to some deletion kernel $d(\lambda, \cdot)$, which might be defined in the first instance only for partitions λ of n, then there is for each $1 \leq m \leq n$ an essentially unique way to construct $d(\lambda, \cdot)$ for partitions λ of m, so that formula (5) holds also for m instead of n. Iterated deletion of parts of π_n according to this extended deletion kernel puts the parts of π_n in a particular random order, call it the *order of deletion according to d*. This defines a random *composition of n*, that is a sequence of strictly positive integer random variables (of random length) with sum n. We may represent such a random composition of n as an infinite sequence of random variables, by padding with zeros. The various distributions involved in this representation of π_n are spelled out in the following corollary, which follows easily from the theorem.

Corollary 4 *In the setting of the preceding theorem,*

(i) *for each $1 \leq m \leq n$ the distribution $q(m, \cdot)$ of H_m is derived from $q(n, \cdot)$ by the formula*

$$q(m, k) = \frac{q_0(m, k)}{1 - q_0(m, 0)} \qquad (1 \leq k \leq m) \tag{13}$$

where
$$q_0(m,k) := \sum_{x=1}^{n} q(n,x) \frac{\binom{n-x}{m-k}\binom{x}{k}}{\binom{n}{m}} \qquad (0 \le k \le m).$$

(ii) *Let $X_{n,1}, X_{n,2}, \ldots$ be a sequence of non-negative integer valued random variables such that $X_{n,1}$ has distribution $q(n,\cdot)$, and for $j \ge 1$*
$$\mathbb{P}(X_{n,j+1} = \cdot \,|\, X_{n,1} + \cdots + X_{n,j} = r) = q(n-r,\cdot) \tag{14}$$
with $X_{n,j+1} = 0$ if $X_{n,1} + \cdots + X_{n,j} = n$, so $\mathcal{X}_n := (X_{n,1}, X_{n,2}, \ldots)$ is a random composition of n with
$$\mathbb{P}(\mathcal{X}_n = \lambda) = \prod_{j=1}^{\ell} q(\lambda_j + \cdots + \lambda_\ell, \lambda_j) \tag{15}$$
for each composition λ of n with ℓ parts of sizes $\lambda_1, \lambda_2, \ldots, \lambda_\ell$. Then (π_n, X_n) with the joint distribution described by Theorem 3 *can be constructed as follows: let $X_n = X_{n,1}$ and define π_n by ranking $\mathcal{X}_n$.*

(iii) *For each $1 \le m \le n$ the distribution of π_m is given by the formula*
$$\mathbb{P}(\pi_m = \lambda) = \sum_{\sigma} \prod_{j=1}^{\ell} q(\lambda_{\sigma(j)} + \cdots + \lambda_{\sigma(\ell)}, \lambda_{\sigma(j)}) \tag{16}$$
where λ is a partition of m into ℓ parts of sizes $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_\ell > 0$, and the summation extends over all $m!/\prod a_j(\lambda)!$ distinct permutations σ of the ℓ parts of λ, with $a_j(\lambda)$ being the number of parts of λ of size j.

(iv) *Let $d(\lambda, x)$ for partitions λ of $m \le n$ and x a part of λ be derived from q and p via formula* (5)*, and let $\mathcal{X}_n$ be the random composition of n defined by the parts of π_n in order of deletion according to d. Then $\mathcal{X}_n$ has the distribution described in part* (ii).

Following [4], we call a transition probability matrix $q(m,j)$ indexed by $1 \le j \le m \le n$, with $\sum_{j=1}^{m} q(m,j) = 1$, a *decrement matrix.* A *random composition of n generated by* q is a sequence of random variables $\mathcal{X}_n := (X_{n,1}, X_{n,2}, \ldots)$ with distribution defined as in part (ii) of the previous corollary. Hoppe [7] called this scheme for generating a random composition of n a *discrete residual allocation model.*

Suppose now that $\mathcal{X}_n$ is the sequence of sizes of classes in a random ordered partition $\widetilde{\Pi}_n$ of the set $[n]$, meaning a sequence of disjoint non-empty sets whose union is $[n]$, and that conditionally given $\mathcal{X}_n$ all possible choices of $\widetilde{\Pi}_n$ are equally likely. Let $\mathcal{X}_m$ be the sequence of sizes of classes of the ordered partition of $[m]$ defined by restriction of $\widetilde{\Pi}_n$ to $[m]$. Then the $\mathcal{X}_m$ is said to be *derived from $\mathcal{X}_n$ by sampling*, and the sequence of distributions of $\mathcal{X}_m$ is called *sampling consistent.* A *composition structure* is a sampling consistent sequence of distributions of compositions $\mathcal{X}_n$ of n for $n = 1, 2, \ldots$.

Definition 5 Following [4], we call a random composition $\mathcal{X}_n = (X_{n,1}, X_{n,2}, \ldots)$ of n *regenerative*, if for each $1 \le x < n$, conditionally given that $X_{n,1} = x$ the remaining composition $(X_{n,2}, \ldots)$ of $n-x$ is distributed according to the unconditional distribution of $\mathcal{X}_{n-x}$ derived from $\mathcal{X}_n$ by random sampling. Call a composition structure $(\mathcal{X}_n)$ *regenerative* if $\mathcal{X}_n$ is regenerative for each $n = 1, 2, \ldots$.

Note the close parallel between this definition of regenerative compositions and Definition 1 of regenerative partitions. The regenerative property of a random partition is more subtle, because it involves random selection of *some* part to delete, and this selection process is allowed to be as general as possible, while for random compositions it is simply the first part that is deleted. The relation between the two concepts is provided by the following further corollary of Theorem 3:

Corollary 6 *If the parts of a regenerative partition π_n of n are put in deletion order to define a random composition of $\mathcal{X}_n$ of n, as in part* (iv) *of the previous corollary, then $\mathcal{X}_n$ is a regenerative composition of n.*

This reduces the study of regenerative partitions to that of regenerative compositions, for which a rather complete theory has already been presented in [4]. In particular, the basic results of [4], recalled here in Section 5, provide an explicit paintbox representation of regenerative partition structures, along with an integral representation of corresponding decrement matrices q. See also Section 4 for some variants of Corollary 6.

For obvious reasons, a partition structure π_n cannot be regenerative if π_n has at most m parts for every n, for some $m < \infty$. In particular, the two-parameter partition structures defined by (10) for $\alpha < 0$ and $\theta = -m\alpha > 0$ are not regenerative. Less obviously, the partition structures defined by (10) for $0 < \alpha < 1$ and $-\alpha < \theta < 0$, which have an unbounded number of parts, are also not regenerative. This follows from Corollary (6) and the discussion of [4], where it was shown that for this range of parameters the two-parameter partition structure cannot be associated with a regenerative composition structure.

2 Proof of Theorem 2

This is an extension of the argument of Kingman [10] in the case $\tau = 0$. Recall first that when partitions λ are encoded by their multiplicities, $a_r = a_r(\lambda)$ for $r = 1, 2, \ldots$, the sampling consistency condition on a partition probability function p is expressed by the formula

$$p(a_1, a_2, \ldots) = p(a_1 + 1, a_2, \ldots)\frac{a_1 + 1}{n + 1} + \sum_{r>1} p(\ldots, a_{r-1} - 1, a_r + 1, \ldots)\frac{r(a_r + 1)}{n + 1} \quad (17)$$

where p is assumed to vanish except when its arguments are non-negative integers, and $n = \sum_r r a_r$.

Assuming that p is a regenerative with respect to d, iterating (5) we have for parts $r, s \in \lambda$,

$$p(\lambda) = \frac{q(n,r)}{d(\lambda,r)} \frac{q(n-r,s)}{d(\lambda-\{r\},s)} p(\lambda-\{r,s\}), \tag{18}$$

which can clearly be expanded further. Since this expression is invariant under permutations of the parts, interchanging r and s we get

$$\frac{q(n,r)}{d(\lambda,r)} \frac{q(n-r,s)}{d(\lambda-\{r\},s)} = \frac{q(n,s)}{d(\lambda,s)} \frac{q(n-s,r)}{d(\lambda-\{s\},r)}.$$

Assume now that d is given by (12). Introducing

$$b(n,r) := \frac{q(n,r)n}{(n-r)\tau + r(1-\tau)}$$

formula (18) yields $b(n,r)b(n-r,s) = b(n,s)b(n-s,r)$. Taking $s = 1$ and abbreviating $f(n) := b(n,1)$ we obtain $b(n,r)/b(n-1,r) = f(n)/f(n-r)$, thus

$$b(n,r) = f(n-r+1)\cdots f(n-1)f(n)g(r)\,, \quad \text{for} \quad g(r) := \frac{b(r,r)}{f(1)\cdots f(r)}\,.$$

The full expansion of p now reads

$$p(\lambda) = \prod_{i=0}^{\ell-1}(1-\tau+i\,\tau)\prod_{k=1}^{n} f(k) \prod_{r} \frac{g(r)^{a_r}}{a_r!}$$

where a_r is the number of parts of λ of size r, with $\Sigma\, a_r = \ell$ and $\Sigma\, r a_r = n$. By homogeneity we can choose the normalisation $g(1) = 1$. Assuming that p is a partition structure, substituting into (17) and cancelling common terms gives

$$\frac{n+1}{f(n+1)} = (1-\tau+\ell\,\tau) + \sum_{r>1} r\, a_{r-1} \frac{g(r)}{g(r-1)}.$$

Now defining $h(r)$ by the substitution

$$\frac{g(r)}{g(r-1)} = -\frac{\tau}{r} + \frac{r-1}{r}\, h(r)$$

we obtain

$$\frac{n+1}{f(n+1)} = 1-\tau + \sum_{r>1}(r-1)a_{r-1}h(r)$$

which must hold for arbitrary partitions, hence $h(r) = \gamma$ for some constant. Therefore

$$f(n) = \frac{n}{1-\tau+(n-1)\gamma}\,, \qquad g(r) = \frac{(\gamma-\tau)_{r-1\uparrow\gamma}}{r!}.$$

It follows that

$$p(\lambda) = \frac{n!\,(1-\tau)_{\ell\uparrow\tau}}{(1-\tau)_{n\uparrow\gamma}} \prod_r \left(\frac{(\gamma-\tau)_{r-1\uparrow\gamma}}{r!}\right)^{a_r} \frac{1}{a_r!}$$

which is positive for all λ iff $\gamma > \tau$. The substitution

$$\alpha = \frac{\tau}{\gamma}\,, \qquad \theta = \frac{1-\tau}{\gamma}$$

reduces this expression to the two-parameter formula (10), and Theorem 2 follows.

3 Fragmented permutations

We use the term *fragmented permutation of* $[n]$ for a pair $\gamma = (\sigma, \lambda) \in S_n \times C_n$, where S_n is the set of all permutations of $[n]$, and C_n is the set of all compositions of n. We interpret a fragmented permutation γ as a way to first arrange n balls labeled by $[n]$ in a sequence, then fragment this sequence into some number of *boxes*. We may represent a fragmented permutation in an obvious way, e.g.

$$\gamma = 2, 3, 9 \,|\, 1, 8 \,|\, 6, 7, 5 \,|\, 4$$

describes the configuration with balls 2, 3 and 9 in that order in the first box, balls 1 and 8 in that order in the second box, and so on, that is $\gamma = (\sigma, \lambda)$ for $\sigma = (2, 3, 9, 1, 8, 6, 7, 5, 4)$ and $\lambda = (3, 2, 3, 1)$.

We now define a transition probability matrix on the set of all fragmented permutations of $[n]$. We assume that some probability distribution $q(n, \cdot)$ is specified on $[n]$. Given some initial fragmented permutation γ,

- let X_n be a random variable with distribution $q(n, \cdot)$, meaning

$$\mathbb{P}(X_n = x) = q(n, x), \quad 1 \leq x \leq n;$$

- given $X_n = x$, pick a sequence of x different balls uniformly at random from the

$$n(n-1)\cdots(n-x+1)$$

  possible sequences;

- remove these x balls from their boxes and put them, in the order they are chosen, into a new box to the left of the remaining $n - x$ balls in boxes.

To illustrate for $n = 9$, if the initial fragmented permutation is $\gamma = 2, 3, 9 \,|\, 1, 8 \,|\, 6, 7, 5 \,|\, 4$ as above, $X_9 = 4$ and the sequence of balls chosen is $(7, 4, 8, 1)$, then the new fragmented permutation is

$$7, 4, 8, 1 \,|\, 2, 3, 9 \,|\, 6, 5.$$

Definition 7 Call the Markov chain with this transition mechanism the *$q(n,\cdot)$-chain on fragmented permutations of* $[n]$.

To prepare for the next definition, we recall a basic method of transformation of transition probability functions. Let Q be a transition probability matrix on a finite set S, and let $f : S \to T$ be a surjection from S onto some other finite set T. Suppose that the $Q(s,\cdot)$ distribution of f depends only on the value of $f(s)$, that is

$$\sum_{x: f(x)=t} Q(s,x) = \widehat{Q}(f(s),t), \qquad (t \in T) \tag{19}$$

for some matrix $\widehat{Q}$ on T. The following consequences of this condition are elementary and well known:

- if $(Y_n, n = 0,1,2,\ldots)$ is a Markov chain with transition matrix Q and starting state x_0, then $(f(Y_n), n = 0,1,2,\ldots)$ is a Markov chain with transition matrix $\widehat{Q}$ and starting state $f(x_0)$;
- if Q has a unique invariant probability measure π, then $\widehat{Q}$ has unique invariant probability measure $\widehat{\pi}$ which is the π distribution of f.

To decribe this situation, we may say that $\widehat{Q}$ is the *push-forward of Q by f.*

Definition 8 The *$q(n,\cdot)$-chain on permutations of* $[n]$ is the $q(n,\cdot)$-chain on fragmented permutations of $[n]$ pushed forward by projection from (σ,λ) to σ. Similarly, pushing forward from (σ,λ) to λ defines the *$q(n,\cdot)$-chain on compositions of n* and pushing forward further from compositions to partitions, by ranking, defines the *$q(n,\cdot)$-chain on partitions of n.*

In terms of shuffling a deck of cards, the $q(n,\cdot)$-chain on permutations of $[n]$ can be represented as a random to top shuffle in which a number X is first picked at random according to $q(n,\cdot)$, then X cards are picked one by one from the deck and put in uniform random order to form a packet which is then placed on top of the deck. This is the inverse of the top X to random shuffle studied by Diaconis, Fill and Pitman [3], in which X cards are cut off the top of the deck, then inserted one by one uniformly at random into the bottom of the deck. Keeping track of packets of cards in this shuffle leads naturally to the richer state space of fragmented permutations.

The mechanism of the $q(n,\cdot)$-chain on compositions of n is identical to that described above for fragmented permutations, except that the labels of the balls are ignored. The mechanism of the $q(n,\cdot)$-chain on partitions of n is obtained by further ignoring the order of boxes in the composition. The following lemma connects these Markov chains to the basic definitions of regenerative partitions and regenerative compositions which we made in Section 1.

Lemma 9 *Let $q(n,\cdot)$ be a probability distribution on $[n]$. Then*

(i) *a random composition* $\mathcal{X}_n = (X_{n,1}, X_{n,2}, \ldots)$*, with* $X_{n,1}$ *distributed according to* $q(n, \cdot)$*, is regenerative if and only if the distribution of* $\mathcal{X}_n$ *is an invariant distribution for the* $q(n, \cdot)$*-chain on compositions of* n*,*

(ii) *a random partition* π_n *is regenerative with respect to deletion of some part* X_n *with distribution* $q(n, \cdot)$ *if and only if the distribution of* π_n *is an invariant distribution for the* $q(n, \cdot)$*-chain on partitions of* n*.*

Proof. The proofs of the two cases are similar, so we provide details only in case (ii). The condition that π_n is regenerative with respect to deletion of X_n can be written as follows:

$$\pi_n - \{X_n\} \stackrel{d}{=} \hat{\pi}_{n-X_n}$$

where

(a) $\stackrel{d}{=}$ denotes equality in distribution of two random elements of the set of partitions of m for some $0 \leq m \leq n$, allowing a trivial partition of 0,

(b) on the left side $\pi_n - \{X_n\}$ denotes the random partition of $n - X_n$ derived from π_n by deletion of the part X_n of π_n,

(c) on the right side $(\hat{\pi}_m, 0 \leq m \leq n)$ is a sampling consistent sequence of random partitions, independent of X_n, with $\hat{\pi}_n \stackrel{d}{=} \pi_n$.

Consider the random partition

$$\pi_n^* := \hat{\pi}_{n-X_n} \cup \{X_n\}$$

obtained from $\hat{\pi}_n$ by first removing X_n balls from $\hat{\pi}_n$ by random sampling, then putting all these balls in a new box. The conditional distribution of π_n^* given $\hat{\pi}_n$ defines a transition probability matrix on the set of partitions of n, which is the transition matrix of the $q(n, \cdot)$-chain on partitions of n. If π_n is regenerative with respect to deletion of X_n, then $\pi_n \stackrel{d}{=} \hat{\pi}_n \stackrel{d}{=} \pi_n^*$. That is to say, the distribution of π_n is an invariant probability measure for the $q(n, \cdot)$-chain on partitions of n. Conversely, if the distribution of $\hat{\pi}_n$ is invariant for the $q(n, \cdot)$-chain on partitions of n, so $\hat{\pi}_n \stackrel{d}{=} \pi_n^*$, we can set $\pi_n := \pi_n^*$, and then by construction

$$\pi_n - \{X_n\} = \pi_n^* - \{X_n\} = \hat{\pi}_{n-X_n}.$$

So π_n is regenerative with respect to deletion of X_n with distribution $q(n, \cdot)$. □

Theorem 3 and its corollaries now follow easily from the previous lemma and the following lemma:

Lemma 10 *For each probability distribution* $q(n, \cdot)$ *on* $[n]$*, the* $q(n, \cdot)$*-chain on fragmented permutations of* $[n]$ *has a unique stationary distribution. Under this distribution,*

(i) *the permutation of* $[n]$ *has uniform distribution;*

(ii) *the permutation and the composition are independent;*

(iii) *the composition of n is generated by q according to* Corollary 4.

Hence, the distribution on compositions described by Corollary 4 (ii) *is the unique stationary probability distribution for the $q(n,\cdot)$-chain on compositions of n, and distribution of π_n described by formula* (16) *is the unique stationary probability distribution for the $q(n,\cdot)$-chain on partitions of n.*

Proof. Let $m := \max\{x : q(n,x) > 0\}$, and write $n = km + r$ for positive integers k and r with $1 \le r < m$. We argue that whatever the initial state γ, there is a strictly positive probability that after $k+1$ steps the $q(n,\cdot)$-chain on fragmented permutations reaches the state

$$1, 2, \ldots, m \,|\, \cdots \,|\, (k-1)m+1, \ldots, km \,|\, km+1, \ldots, n$$

in which the permutation is the identity and the composition is $(m, \cdots, m, r)$. To see this, note that the transition mechanism ensures that after one step from γ it is possible to reach a state of the form

$$n-m+1, n-m+2, \ldots, n \,|\, \ldots\ldots$$

for some $\ldots\ldots$ determined by the initial configuration γ. Then after two steps it is possible to reach a state of the form

$$(k-1)m+1, \ldots, km \,|\, km+1, \ldots, n \,|\, \ldots\ldots$$

for some $\ldots\ldots$, and so on. Since there is a state which can be reached eventually no matter what the initial state, it follows from the elementary theory of Markov chains that a stationary distribution exists and is unique.

Let $P_{q(n,\cdot)}$ denote the push-forward of this stationary distribution to compositions of n, that is the stationary distribution of the $q(n,\cdot)$-chain on compositions of n. By definition of the $q(n,\cdot)$-chain on fragmented permutations,

- under $P_{q(n,\cdot)}$ the number of balls in the first box has distribution $q(n,\cdot)$;
- under $P_{q(n,\cdot)}$, for each $1 \le m < n$, given that the first box contains $n-m$ balls, the remaining composition of m has the distribution on compositions of m derived from $P_{q(n,\cdot)}$ by taking a random sample of m out of the n balls, to be denoted $(P_{q(n,\cdot)})^{n\to m}$.

To complete the proof of the lemma, it just remains to check the key identity

$$(P_{q(n,\cdot)})^{n\to m} = P_{q(m,\cdot)} \tag{20}$$

where $q(m,\cdot)$ is derived from $q(n,\cdot)$ by formula (13). Due to independence of the composition and the permutation, a composition of m with distribution $(P_{q(n,\cdot)})^{n\to m}$ is obtained from the stationary distribution of the $q(n,\cdot)$-chain on fragmented permutations by ignoring balls $m+1, \ldots, n$, and considering the composition of m induced by the balls labeled by $[m]$. But when the fragmented permutation evolves according to the $q(n,\cdot)$-chain, it is clear that at each step, no matter what the initial state, for $0 \le s \le m$, the probability that exactly s of the m balls get moved to the left is $q_0(m,s)$ as in (13). Let $q(m,\cdot)$ be as in formula (13). Since the probability of moving at least one of the first m balls is $1 - q_0(m,0)$, no matter what the initial state, the $q(n,\cdot)$-chain on fragmented permutations of $[n]$ pushes forward to a Markov chain on fragmented permutations of $[m]$. The transition matrix of this chain is a mixture of the identity matrix and the matrix of $q(m,\cdot)$-chain on fragmented permutations of $[m]$, with weights $q_0(m,0)$ and $1 - q_0(m,0)$ respectively. Hence the equilibrium distribution of this chain with state space fragmented permutations of $[m]$ is identical to the equilibrium distribution of the $q(m,\cdot)$-chain on fragmented permutations of $[m]$, whose projection onto compositions of $[m]$ is $P_{q(m,\cdot)}$. This proves (20). $\square$

4 Some corollaries

This section spells out some further corollaries of Theorem 3.

Corollary 11 *The distribution of every regenerative partition π_n of n is obtained by ranking the components of some regenerative composition $\mathcal{X}_n$ of n, whose distribution is uniquely determined by that of π_n. Then π_n is regenerative with respect to deletion of X_n distributed like the first component of $\mathcal{X}_n$. This correspondence, made precise by the formulae of* Corollary 4*, establishes bijections between the following sets of probability distributions:*

(i) *probability distributions $q(n,\cdot)$ on $[n]$;*

(ii) *distributions of regenerative compostions $\mathcal{X}_n$ of n;*

(iii) *distributions of regenerative partitions π_n on n.*

An explicit link from (iii) to (i) is provided by a recursive formula [4, Equation (34)] expressing $q(n,\cdot)$ via the probabilities of one-part partitions $(p(j),\ j = 1,\dots,n)$. There is also an explicit formula expressing $q(n,\cdot)$ as a rational function of probabilities $(p(1^m),\ m = 1,\dots,n)$ where $1^m \vdash m$ is the partition with only singleton parts.

As a variant of the above corollary, we record also:

Corollary 12 *Given a decrement matrix $q = (q(m,j), 1 \le j \le m \le n)$ for some fixed n, for $1 \le m \le n$ let $\mathcal{X}_m$ be the random composition of m generated by q, and π_m the random partition of m obtained by ranking $\mathcal{X}_m$. The following conditions are equivalent:*

(i) *the entire decrement matrix q is determined by $q(n,\cdot)$ according to formula* (13)*;*

(ii) *the sequence of compositions $(\mathcal{X}_m, 1 \le m \le n)$ derived from q is sampling consistent;*

(iii) *the sequence of partitions $(\pi_m, 1 \le m \le n)$ derived from q as in* (16) *is sampling consistent.*

The equivalence of (i) and (ii) can also be read from [4], where formula (13) was given only for $m = n - 1$ as a means of recursively computing $q(m, \cdot)$ from $q(n, \cdot)$ for $m < n$. That (ii) implies (iii) is obvious. That (iii) implies (ii) is not obvious, but this is an immediate consequence of Theorem 3 and Corollary 4, because (iii) means that π_n is regenerative with respect to deletion of the first term of $\mathcal{X}_n$.

Corollary 13 *The following two conditions on a random composition of n*

$$\mathcal{X}_n := (X_{n,1}, X_{n,2}, \ldots)$$

are equivalent:

(i) $\mathcal{X}_n$ *is regenerative;*

(ii) *for each* $1 \leq j \leq k < n$*, conditionally given* $X_{n,1}, \ldots, X_{n,j}$ *with* $X_{n,1} + \cdots + X_{n,j} = k$*, the partition of* $n-k$ *obtained by ranking* $X_{n,j+1}, X_{n,j+2}, \ldots$ *has the same distribution as* π_{n-k}*, the partition of* $n - k$ *obtained by sampling from* $\mathcal{X}_n$*.*

Proof. That (i) implies (ii) is obvious. Conversely, condition (ii) for $j = 1$ states that the partition π_n derived from $\mathcal{X}_n$ is regenerative with respect to deletion of $X_{n,1}$. Let $q(n, \cdot)$ be the distribution of $X_{n,1}$, and let d denote the corresponding deletion kernel, extended to partitions of m for $1 \leq m \leq n$ in accordance with Theorem 3. Condition (ii) for $j = 2$ implies that for each i such that $q(n, i) > 0$, the partition π_{n-i} is regenerative with respect to deletion of a part whose unconditional distribution equals the conditional distribution of $X_{n,2}$ given $X_{n,1} = i$. According to the uniqueness statement of Theorem 3, this distribution must be the distribution $q(n - i, \cdot)$ determined by $q(n, \cdot)$ via (13). So the joint law of $(X_{n,1}, X_{n,2})$ is identical to that described in Corollary 4 (ii). Continuing in this way, it is clear that the distribution of the entire sequence $\mathcal{X}_n$ is that described in Corollary 4 (ii). The conclusion now follows from Corollary 6. $\square$

5 Paintbox representations

Gnedin's paintbox representation of composition structures [6] uses a random closed set $\mathcal{R} \subset [0, 1]$ to separate points of a uniform sample into clusters. Given $\mathcal{R}$, define an interval partition of $[0, 1]$ comprised of *gaps*, that is open interval components of $[0, 1] \setminus \mathcal{R}$, and of individual points of $\mathcal{R}$. A random ordered partition of $[n]$ is then constructed from $\mathcal{R}$ and independent uniform sample points $U_1, \ldots, U_n$ by grouping the indices of sample points which fall in the same gap, and letting the points which hit $\mathcal{R}$ to be singletons. A random composition $\mathcal{X}_n$ of n is then constructed as the sequence of block sizes in this partition of $[n]$, ordering the blocks from left to right, according to the location of the corresponding sample points in $[0, 1]$. Gnedin showed that every composition structure $(\mathcal{X}_n)$ can be so represented. As in Kingman's representation of partition structures, $\mathcal{R}$ can be interpreted as an asymptotic shape of $\mathcal{X}_n$, provided $\mathcal{X}_n$ is properly encoded as an element of the metric space of closed subsets of $[0, 1]$ with the Hausdorff distance function.

According to the main result of [4], each regenerative composition structure $(\mathcal{X}_n)$ is associated in this way with an $\mathcal{R}$ which is *multiplicatively regenerative* in the following sense: for $t \in [0,1]$ let $D_t := \inf([t,1] \cap \mathcal{R})$, and given that $D_t < 1$ let

$$\mathcal{R}^{[D_t,1]} := \{(z - D_t)/(1 - D_t), z \in \mathcal{R} \cap [D_t, 1]\}$$

which is the restriction of $\mathcal{R}$ to $[D_t, 1]$ scaled back to $[0,1]$; then

$$(\mathcal{R}^{[D_t,1]} \,|\, D_t \text{ with } D_t < 1\,, \mathcal{R} \cap [0, D_t]) \stackrel{d}{=} \mathcal{R} \tag{21}$$

meaning that the conditional distribution of $\mathcal{R}^{[D_t,1]}$, given D_t with $D_t < 1$ and given $\mathcal{R} \cap [0, D_t]$, is identical to the unconditional distribution of $\mathcal{R}$. This condition holds if and only if $-\log(1-\mathcal{R})$ is a regenerative subset of $[0,\infty[$ in the usual (additive) sense. So by a result of Maisonneuve, the most general multiplicatively regenerative subset $\mathcal{R}$ can be constructed as the closure of $\{1 - \exp(-S_t), t \geq 0\}$ for some *subordinator* $(S_t, t \geq 0)$, that is an increasing process with stationary independent increments [1]. Thus regenerative composition structures are parameterised by a pair $(\tilde{\nu}, \mathtt{d})$ where $\tilde{\nu}$ is a measure on $]0,1]$ with finite first moment and $\mathtt{d} \geq 0$. The measure $\tilde{\nu}(\mathrm{d}u)$ is the image of the Lévy measure $\nu(\mathrm{d}s)$ of the subordinator via the transformation from s to $1-\exp(-s)$. and $\mathtt{d}$ is the drift parameter of the subordinator. So the Laplace exponent of the subordinator, evaluated at a positive integer n, is

$$\Phi(n) = n\mathtt{d} + \int_{]0,1]} (1 - (1-x)^n)\tilde{\nu}(\mathrm{d}x)\,.$$

The decrement matrix of the regenerative composition is then

$$q(n,r) = \frac{\Phi(n,r)}{\Phi(n)}\,, \qquad 1 \leq r \leq n\,,\; n = 1, 2, \ldots$$

where

$$\Phi(n,r) = n\mathtt{d}\,1(r=1) + \binom{n}{r} \int_{]0,1]} x^r (1-x)^{n-r} \tilde{\nu}(\mathrm{d}x)\,.$$

Uniqueness of the parameterisation is achieved by a normalisation condition, e.g. $\Phi(1) = 1$.

The partition structure derived by sampling from a random closed subset $\mathcal{R}$ of $[0,1]$ depends only on the distribution of the sequence of *ranked lengths induced by* $\mathcal{R}$

$$V(\mathcal{R}) := (V_1(\mathcal{R}), V_2(\mathcal{R}), \ldots)$$

where $V_i(\mathcal{R})$ is the length of the ith longest interval component of $[0,1] \setminus \mathcal{R}$. Our consideration of regenerative partition structures suggests the following definition. Call $\mathcal{R}$ *weakly multiplicatively regenerative* if for each $t \in [0,1]$

$$(V(\mathcal{R}^{[D_t,1]}) \,|\, D_t \text{ with } D_t < 1\,, \mathcal{R} \cap [0, D_t]) \stackrel{d}{=} V(\mathcal{R}) \tag{22}$$

meaning that the conditional distribution of relative ranked lengths induced by $\mathcal{R} \cap [D_t, 1]$, given D_t with $D_t < 1$ and given the restricted set $\mathcal{R} \cap [0, D_t]$, is identical to the unconditional distribution of ranked lengths induced by $\mathcal{R}$. From Theorem 3 we easily deduce:

Corollary 14 *A random closed subset $\mathcal{R}$ of $[0,1]$ is weakly multiplicatively regenerative if and only if it is multiplicatively regenerative.*

Proof. The "if" part is obvious by measurability of the map from $\mathcal{R}$ to $V(\mathcal{R})$. To argue the converse, suppose that $\mathcal{R}$ is weakly multiplicatively regenerative. Without loss of generality, it can be supposed that $\mathcal{R}$ is defined on the same probability space as a sequence of independent uniform $[0,1]$ variables U_i for $i = 1, 2, \ldots$. Let $\mathcal{X}_n = (X_{n,1}, X_{n,2}, \ldots)$ for $n = 1, 2, \ldots$ be the sequence of compositions of n derived from $\mathcal{R}$ by sampling with these independent uniform variables, and let π_n be the partition of n defined by ranking $\mathcal{X}_n$. By consideration of (22) with t replaced by $U_{n,1}$, where $U_{n,k}$ is the kth order statistic of $U_1, \ldots, U_n$, it is easily argued that π_n is regenerative with respect to deletion of $X_{n,1}$. If $X_{n,1} < n$, we apply (22) at $t = U_{n,X_{n,1}+1}$ and repeat the argument to show that $\pi_n - \{X_{n,1}, X_{n_2}\}$ is a distributional copy of π_{n-m} given $X_{n,1}$ and $X_{n,2}$ with $X_{n,1} + X_{n,2} = m$. Iterating further we see that $\mathcal{X}_n$ satisfies condition (ii) of Corollary 13. Hence $\mathcal{X}_n$ is regenerative. Thus $(\mathcal{X}_n)$ defines a regenerative composition structure, and it follows the main result of [4] that the set $\mathcal{R}$ is regenerative. $\square$

6 Uniqueness and positivity

For each n equation (5) is the basic algebraic relation between the functions $p(\lambda)$, $d(\lambda)$ for $\lambda \vdash n$ and $q(n, \cdot)$. According to Corollary 4 $q(n, \cdot)$ uniquely determines p, for each n, but d satisfying (5) need not be unique because $p(\lambda)$ may assume zero value for some λ, as illustrated by the following example.

Example (Regenerative hook partition structures.) A partition λ of n is called a *hook* if it has at most one part larger than 1. The only regenerative partition structures p such that π_n is a hook with probability one for every n are those which can be generated by $\tilde{\nu} = \delta_1$, the Dirac mass at 1 and $\mathtt{d} \geq 0$, including the trivial boundary case with $d = \infty$. Then for $n > 1$

$$q(n,n) = \frac{1}{1 + n\mathtt{d}}, \qquad q(n,1) = \frac{n\mathtt{d}}{1 + n\mathtt{d}}.$$

This implies that the associated *hook composition structure* gives positive probability only to compositions of the type (n) or $(1^m, n-m)$, $1 \leq m \leq n$, for every n. For these hook composition structures the deletion kernel is an arbitrary kernel with the property

$$d(\lambda,\, 1) = 1 \ \text{ if } 1 \in \lambda\,. \tag{23}$$

This property is characteristic, that is each partition structure regenerative according to such deletion kernel d is derived from a hook composition structure. Indeed, if a composition structure compatible with such d gave positive probability to a composition $(x, y, \ldots)$ with $x > 1$ then, by sampling consistency, the composition $(2, 1)$ would also have positive probability, contradicting $q(3, 2) = 0$.

The next lemma shows that only in the hook case can there be any ambiguity about the deletion kernel generating some partition structure.

Lemma 15 *Let (π_n) be a regenerative partition structure. Then*

(i) *either (π_n) is a hook partition structure,*

(ii) *or $p > 0$, $d > 0$ and $q > 0$ for all admissible values of arguments.*

Moreover, (i) *holds iff $p(2, 2) = 0$, while* (ii) *holds iff $p(2, 2) > 0$.*

Proof. If V_2 in Kingman's representation is strictly positive with nonzero probability, then $p(n, n) > 0$ for all n. By virtue of $p(n, n) = q(2n, n)p(n)$ follows $q(2n, n) > 0$, but then by Corollary 4(i) $q(n, \cdot) > 0$ and this implies $p > 0$ and $d > 0$. In this case $p(2, 2) > 0$. Alternatively, if $\mathbb{P}(V_2 = 0) = 1$ then $p(2, 2) = 0$ and we are in the hook case. $\square$

Thus, if d is not a kernel satisfying (23), then $d(\lambda, x) = 0$ for some partition λ and some part $x \in \lambda$ implies that a partition structure regenerative according to d is trivial, (*i.e.* is either the pure-singleton partition with $p(1^n) \equiv 1$, or the one-block partition with $p(n) \equiv 1$).

The positivity condition in lemma rules out nontrivial partition structures, which have an absolute bound on the number of parts for all n. For example, none of the members of the two-parameter family of partition structures (10) is regenerative for $\alpha < 0$ and $-\theta/\alpha = k \in \{2, 3, \ldots\}$ because each π_n has at most k parts.

Corollary 16 *If a partition structure is regenerative and satisfies $p(2, 2) > 0$ then q uniquely determines p and d, and p uniquely determines q and d. Thus if a regenerative partition structure is not hook the corresponding deletion kernel is unique.*

Checking if a given partition structure is regenerative according to an unknown deletion kernel can be done by first computing q, by some algebraic manipulations, from a given partition probability function p, then computing a partition probability function p_* for the regenerative partition structure related to this q, and finally checking if $p = p_*$. When this method is applied to a partition from the two-parameter family with $0 < \alpha < 1$, $-\alpha < \theta < 0$, the resulting q recorded in [4, Equation (39)] is not everywhere positive, hence the partition structures with such parameters are not regenerative.

7 Generalisations and related work

Given some deletion kernel d, it is of interest to consider pairs of partition structures (p_0, p_1) such that d *reduces* p_0 *to* p_1, meaning that the following extension of formula (5) holds:

$$p_0(\lambda)d(\lambda, x) = q(n, x)p_1(\lambda - \{x\}), \qquad x \in \lambda \qquad (\lambda \vdash n) \tag{24}$$

where

$$q(n,x) := \sum_{\{\lambda \vdash n \,:\, x \in \lambda\}} d(\lambda,x) p_0(\lambda) \qquad (1 \le x \le n) \tag{25}$$

is the unconditional probability that the deletion rule removes a part of size x from π_n distributed according to p_0.

Pitman [13] showed that if $p_0, p_1, p_2, \ldots$ is a sequence of partition structures such that size-biased deletion reduces p_i to p_{i+1} for each $i \ge 0$, and p_0 can be represented in terms of random sampling from $(V_1, V_2, \ldots)$ with $V_i > 0$ for each i, then p_0 is an (α, θ) partition as in (10) for some $0 \le \alpha < 1$ and $\theta > -\alpha$, in which case p_j is the $(\alpha, \theta + j\alpha)$ partition structure. This result and Theorem 2 are two different two-parameter generalisations of Kingman's characterisation of $(0, \theta)$ partition structures. In the result of [13], the deletion kernel is still defined by size-biased sampling, and repeated deletions generate a succession of partition structures. Whereas in Theorem 2 the deletion kernel is modified, and repeated deletions generate the same partition structure.

A class of partition structures satisfying (24) is associated with Markovian compositions introduced in [5]. A composition structure of this type is derived from a set $\mathcal{R}$ which has a special leftmost interval $[0, X]$, and otherwise $\mathcal{R} \cap [X, 1]$ is a scaled copy of some other multiplicatively regenerative set $\mathcal{R}'$, which is independent of X and has the property (21). For example, the members of the two-parameter family with $0 < \alpha < 1, \ -\alpha < \theta < 0$ can be associated with such Markovian composition structures. In general, such $\mathcal{R}$ can be represented as a transformed range of a finite-mean subordinator with arbitrary initial distribution.

This discussion leaves open a number of interesting questions. One posed in [11, 15], and apparently still open, is the problem of describing all pairs of partition structures (p_0, p_1) such that p_0 reduces to p_1 by size-biased deletion. One could ask the same question for other deletion kernels too. But size-biased deletion is of special interest because of its natural interpretation in terms of Kingman's paintbox construction from ranked random frequencies $(V_1, V_2, \ldots)$: if X_n is a size-biased pick from π_n derived by sampling from such (V_i) associated with the partition structure p_0, then X_n/n converges in distribution to $\tilde{V}_1$ which is a size-biased pick from the limiting frequencies, meaning that

$$\mathbb{P}(\tilde{V} = V_i \,|\, V_1, V_2, \ldots) = V_i \,,$$

$$\mathbb{P}(\tilde{V} = 0 \,|\, V_1, V_2, \ldots) = 1 - \Sigma_i V_i \,,$$

where it is assumed for simplicity that the V_i are almost surely distinct. The probability distribution of $\tilde{V}$ on $[0, 1]$, known as the *structural distribution* encodes many important features of the partition structure p_0. In particular, $p_0(n) = \mathbb{E}(\tilde{P}^{n-1})$ for $n = 1, 2, \ldots$ (see [13, 15, 5]). See also [16, 5] for related work.

Central measures We comment briefly on how our results link to the potential theory on graded graphs, as developed by the Russian school in connection with the asymptotic representation theory of the symmetric group, see [8] for a survey.

For $\lambda \vdash n,\ \mu \vdash n+1$ the number of ways to extend a given partition of $[n]$ with shape λ to some partition of $[n+1]$ with shape μ does not depend on the choice of partition of $[n]$. Denote this number $\kappa(\lambda, \mu)$. Letting $\mathcal{P}_n$ denote the set of partitions of n, consider the graded graph with the vertex set $\mathcal{P} = \bigcup_n \mathcal{P}_n$ and multiplicity $\kappa(\lambda, \mu)$ for the edge connecting λ and μ. A partition structure is a nonnegative solution of the recursion

$$p(\lambda) = \sum_{\mu \vdash n+1} \kappa(\lambda, \mu) p(\mu)\,, \qquad \lambda \vdash n\,, \quad n = 1, 2, \ldots,$$

which is another form of (17), with $p(1) = 1$.

For every $\lambda \vdash n$ a directed path in $\mathcal{P}$ connecting the root (1) and λ encodes a partition of $[n]$ with shape λ. Thus p defines a a *central measure* on the space of infinite paths of $\mathcal{P}$, with the property that all paths of length n, connecting (1) and $\lambda \vdash n$, have the same probability depending only on λ, for every n and $\lambda \in \mathcal{P}_n$. Each infinite path corresponds to a partition of the set $\mathbb{N}$ and the 'centrality' of a measure just means that the random partition of $\mathbb{N}$ is exchangeable.

Originally, the concept of central measure emerged in connection with the Young lattice $\mathcal{Y}$, which has the same vertex set as $\mathcal{P}$, but a different multiplicity function (dictated by the branching rule for characters of the symmetric group). According to a classical result of Thoma, central measures on $\mathcal{Y}$ can be described in terms of two random decreasing sequences $(V_j), (W_j)$ (as to be compared with a single decreasing sequence (V_j) in Kingman's representation). See [9] for particular examples of central measures on $\mathcal{Y}$, and [2] for a general construction of central measures on graphs which generalise both $\mathcal{Y}$ and $\mathcal{P}$.

Although the graphs $\mathcal{Y}$ and $\mathcal{P}$ are different, there is a way to push a partition structure on $\mathcal{P}$ to a central measure on $\mathcal{Y}$, analogous to the transition from monomial symmetric functions to Schur symmetric functions. It would be interesting to study the image of regenerative partition structures under this mapping in some detail.

General deletion kernels For partitions μ and λ, interpreted as distributions of unlabeled balls in unlabeled boxes, write $\mu \subset \lambda$ if μ can be obtained from λ by deletion of one or more boxes. Call a function d on pairs of partitions $(\lambda,\ \mu)$ with $\mu \subset \lambda$, a *deletion kernel* if it satisfies

$$0 \le d \le 1\,, \qquad \sum_{\lambda^- :\, \lambda^- \subset \lambda} d(\lambda\,,\ \lambda^-) = 1\,.$$

For each λ the deletion kernel defines a random split in two subpartitions $\lambda^+ \subset \lambda$ (which remains) and $\lambda^- \subset \lambda$ (which is removed). We can also regard d as transition probability function for a Markov chain which jumps from larger partition λ to a smaller λ^+.

Given d, for a sequence of random partitions $(\pi_n \vdash n,\ n = 1, 2, \ldots)$ – which in general need not satisfy any consistency condition besides that the π_n's are defined on the same probability space – with each π_n we associate a split into π_n^+ and π_n^- (with π_n^- removed). This suggests various concepts of invariance, depending on how much conditioning is allowed. Thus we could distinguish a *weak deletion property*

$$\mathbb{P}(\pi_n^+ = \lambda^+ \,|\, \pi_n^- = \lambda^-) = \mathbb{P}(\pi_{n-r} = \lambda^+)\,, \qquad r = |\lambda^-|\,, \tag{26}$$

and a *strong deletion property*

$$\mathbb{P}(\pi_n^+ = \lambda^+ \,|\, |\pi_n^-| = r) = \mathbb{P}(\pi_{n-r} = \lambda^+)\,,$$

where we write $|\lambda|$ for $\Sigma\lambda_j$. It would be interesting to extend results of this paper to such general deletion operations.

8 Acknowledgement

The work on this paper started during our joint visit to the Laboratoire de Probabilités, University Paris VI, in March 2004. We thank Jean Bertoin and Marc Yor for their hospitality and support.

References

[1] J. Bertoin. Subordinators: examples and applications. Lectures on probability theory and statistics (Saint-Flour, 1997), 1–91, Lecture Notes in Math., 1717, Springer, Berlin, 1999. MR 1746300

[2] A. Borodin and G. Olshanski. Harmonic functions on multiplicative graphs and interpolation polynomials. *Electronic J. Combinatorics*, 7: paper #R28, 2000. MR 1758654

[3] P. Diaconis, J. Fill, and J. Pitman. Analysis of top to random shuffles. *Combinatorics, Probability and Computing*, 1:135–155, 1992. MR 1179244

[4] A. Gnedin and J. Pitman. Regenerative composition structures. Technical Report 644, Dept. Statistics, U.C. Berkeley, 2003. To appear in *Ann. Probab.*. arXiv:math.PR/0307307 .

[5] A. Gnedin and J. Pitman. Composition structures derived from self-similar random sets, 2004. In preparation.

[6] A. V. Gnedin. The representation of composition structures. *Ann. Probab.*, 25(3):1437–1450, 1997. MR 1457625

[7] F. M. Hoppe. The sampling theory of neutral alleles and an urn model in population genetics. *Journal of Mathematical Biology*, 25:123 – 159, 1987. MR 896430

[8] S. Kerov. The boundary of Young lattice and random Young tableaux. *DIMACS Ser. Discr. Math. Theor. Comp. Sci.*, 24:133-158, Amer. Math. Soc. Providence, RI, 1996. MR 1363510

[9] S. Kerov, G. Olshanski and A. Vershik. Harmonic analysis on the infinite symmetric group. arXiv:math.RT/0312270, 2003.

[10] J. F. C. Kingman. Random partitions in population genetics. *Proc. R. Soc. Lond. A.*, 361:1–20, 1978. MR 526801

[11] J. Pitman. The two-parameter generalization of Ewens' random partition structure. Technical Report 345, Dept. Statistics, U.C. Berkeley, 1992. Available at `http://www.stat.berkeley.edu/tech-reports/`.

[12] J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Th. Rel. Fields*, 102:145–158, 1995. MR 1337249

[13] J. Pitman. Random discrete distributions invariant under size-biased permutation. *Adv. Appl. Prob.*, 28:525–539, 1996. MR 1387889

[14] J. Pitman. Partition structures derived from Brownian motion and stable subordinators. *Bernoulli*, 3:79–96, 1997. MR 1466546

[15] J. Pitman. Combinatorial stochastic processes. Technical Report 621, Dept. Statistics, U.C. Berkeley, 2002. Lecture notes for St. Flour course, July 2002. To appear in Springer Lecture Notes in Mathematics. Available at `http://www.stat.berkeley.edu/tech-reports/`.

[16] J. Pitman and M. Yor. Random discrete distributions derived from self-similar random sets. *Electron. J. Probab.*, 1:Paper 4, 1–28, 1996. `http://www.math.washington.edu/~ejpecp/EjpVol1/paper4.abs.html` MR 1386296