Bernoulli **11**(5), 2005, 847–861

Binary sequential representations of random partitions

JAMES E. YOUNG

Department of Mathematics, College of Charleston, Charleston SC 29424, USA.
E-mail: youngj@cofc.edu

Random partitions can be thought of as a consistent family of exchangeable random partitions of the sets $\{1, 2, \ldots, n\}$ for $n \geqslant 1$. Historically, random partitions were constructed by sampling an infinite population of types and partitioning individuals of the same type into a single class. A particularly tractable way to construct random partitions is via random sequences of 0s and 1s. The only random partition derived from an independent 0–1 sequence is Ewens' one-parameter family of partitions which plays a predominant role in population genetics. A two-parameter generalization of Ewens' partition is obtained by considering random partitions constructed from discrete renewal processes and introducing a convolution-type product on 0–1 sequences.

Keywords: combinatorial probability; combinatorial stochastic process; exchangeable; random partition; sequential construction

1. Introduction and motivation

Random partitions arise naturally in the study of random combinatorial structures which decompose into component parts. A general combinatorial theory of such composite structures, expressed in terms of generating functions, was developed in the early 1970s by Foata (1974) and others. There has been much recent interest in the asymptotics and probabilistic aspects of such random combinatorial structures (see Arratia and Tavaré 1992; Arratia *et al.* 2003). The recent St. Flour Lecture Notes by Pitman (2002) include a complete up-to-date overview of random partitions and other combinatorial stochastic processes.

Random partitions have applications in such diverse fields as population genetics (Ethier and Kurtz 1981; Ewens 1972, Hoppe 1987; Kingman 1978a; Watterson 1974), number theory (Vershik 1987; Donnelly and Grimmett 1993), combinatorics (Aldous and Pitman 1993; Donnelly *et al.* 1991a), fragmentation phenomena in physics (Mekjian and Lee 1991; Mekjian 1990; 1991) and computer science. Applications of random partitions have even found their way into the philosophical theory of inductive inference (Zabell 1992)! For example, random permutations decompose into cycles, random mappings decompose into tree components, which are important tools in computer science, and in population genetics theory the components correspond to subpopulations of individuals which have the same allelic type at a particular gene locus. Motivated by applications to fragmentation phenomena in physics, Mekjian and Lee (1991) consider Gibbs-type prescriptions for random partitions that include a class of partitions characterized in this paper via discrete renewal processes.

1350–7265

Motivated by the needs of population genetics theory, Kingman (1978a) introduced the idea of a *partition structure*. In this context, if a random sample of individuals are partitioned into groups having the same allele at a particular gene locus, then the distribution of this partition must be consistent with respect to the taking of further subsamples. This *consistency property* characterizes partition structures.

2. Random partitions

A *partition of n* is simply a collection of positive integers whose sum is n. Let ω_n denote the collection of partitions of n and encode a partition $\pi \in \omega_n$ via the *component counts* $(a_1, \ldots, a_n)$, where a_i is the number of components of size i, for $i = 1, \ldots, n$. A *partition structure* is a sequence of probability measures $P_1, P_2, \ldots$ defined on $\omega_1, \omega_2, \ldots$, respectively, which satisfies the consistency relation

$$P_n(a_1, \ldots, a_n) = P_{n+1}(a_1 + 1, \ldots, a_{n+1}) \frac{a_1 + 1}{n+1} + \sum_{r=2}^{n+1} P_{n+1}(a_1, \ldots, a_{r-1} - 1, a_r + 1, \ldots, a_{n+1}) \frac{r(a_r + 1)}{n+1} \tag{1}$$

for all $n \geqslant 1$. Partition structures are naturally constructed by sampling individuals of various types from hypothetical infinite populations. If we allow these populations to be random and have 'novel' types of individuals, then all partition structures may be constructed in this fashion. This is the content of *Kingman's representation theorem*.

Another interpretation of partition structures is in terms of *exchangeable random partitions* of the set of positive integers. A partition Π of the positive integers $\mathbb{N}$ induces a partition Π_n on $[n] := \{1, \ldots, n\}$ in the obvious way by restricting Π to $[n]$. A random partition Π of $\mathbb{N}$ is *exchangeable* if each Π_n is exchangeable and their distributions are consistent. More precisely, a random partition Π_n on $[n]$ is exchangeable if its distribution is invariant under the obvious action of the symmetric group on partitions of $[n]$ induced by the symmetry group acting as permutations of the set $[n]$. Both Kingman (1982) and Aldous (1985) consider this interpretation, and Aldous (1985) gives an elegant proof of Kingman's representation theorem as an application of de Finetti's theorem.

We use the term *random partition* (or simply *partition* if the context is clear) to always mean a partition structure on the level of set partitions, or equivalently, an exchangeable random partition of $\mathbb{N}$. This is the most natural definition of a random partition when considering partitions of sets of increasing size.

3. Sampling construction of random partitions

A natural question to ask is how random partitions can be constructed. A complete answer is given by Kingman's representation theorem via sampling populations. Regard a partition of n as a partition obtained by sampling from a population of various types. In particular,

since a random partition is defined for all $n \geqslant 1$, we must sample from an infinite population. We define an infinite population as an element of the infinite-dimensional simplex

$$\Delta^{\infty} = \left\{ x = (x_1, x_2, \ldots) : x_i \geqslant 0 \quad \text{and} \quad \sum_{i=1}^{\infty} x_i = 1 \right\}.$$

We interpret x_i as the proportion of individuals of type i in the population. Let $P_n(\cdot|x)$ be the distribution of the partition of n of a random sample of size n taken from the population $x = (x_1, x_2, \ldots)$. Explicitly,

$$P_n(a_1, \ldots, a_n|x) = \frac{n!}{\prod_{i=1}^{\infty}(i!)^{a_i}} \sum_{A} x_1^{v_1} x_2^{v_2} \ldots, \qquad A = \{(v_1, v_2, \ldots)|\#\{v_j = i\} = a_i\}. \quad (2)$$

It is straightforward to check that P_n defined by (2) yields a random partition. More generally, a random partition may be obtained by allowing the population to be random. Give Δ^{∞} the relative topology inherited from the compact metrizable product space $[0, 1]^{\infty}$ and let μ be a probability measure on (the Borel σ-algebra of) Δ^{∞}. Then for a fixed $\pi \in \omega_n$, the map $P_n(\pi|\cdot) : \Delta^{\infty} \to [0, 1]$ defined by (2) is continuous and

$$P_n(\pi) := \int_{\Delta^{\infty}} P_n(\pi|x)\mu(\mathrm{d}x) \quad (3)$$

defines a probability measure on ω_n, which again yields a random partition (as we vary n). It would be nice if every random partition arose in this way. However, the sequence of probability measures (P_n) concentrated on the partitions of n given by the component counts $(n, 0, \ldots, 0)$ for each n forms a random partition that cannot be expressed in the form (3). This example shows that we must allow for the possibility of sampling novel types of individuals from a population.

To incorporate novel types into infinite populations, we extend the definition of population as follows. A population is an element of

$$\overline{\Delta}^{\infty} = \left\{ x = (x_1, x_2, \ldots) : x_i \geqslant 0 \text{ and } \sum_{i=1}^{\infty} x_i \leqslant 1 \right\}.$$

For $x \in \overline{\Delta}^{\infty}$ we define $x_0 = 1 - \sum_{i=1}^{\infty} x_i$ and interpret x_0 as the proportion of novel types in the population. More precisely, when sampling from such a population, the probability of selecting an individual whose type is different from all other types in the population and different from all previously sampled types is x_0.

As above, a random partition is obtained by sampling from a random population with novel types. Notice that this random partition is invariant with respect to permutations of the sequence $x = (x_1, x_2, \ldots)$. With this in mind, we normalize the population and define

$$\overline{\nabla}^{\infty} = \left\{ x = (x_{(1)}, x_{(2)}, \ldots) : x_{(1)} \geqslant x_{(2)} \geqslant \ldots \geqslant 0 \quad \text{and} \quad \sum_{i=1}^{\infty} x_{(i)} \leqslant 1 \right\}.$$

Every random partition can be realized by this construction for a unique measure μ on $\overline{\nabla}^{\infty}$.

Theorem 1 (Kingman's representation theorem). *For each random partition* (P_n) *there exists a unique representing measure* μ *on* $\overline{\nabla}^{\infty}$ *such that* $P_n(\pi) := \int_{\overline{\nabla}^{\infty}} P_n(\pi|x)\mu(\mathrm{d}x)$. *Furthermore, the unique representing measure is the limiting descending-order empirical distribution. Let* $X_r(n)$ *denote the relative frequency in a sample of size* n *(from a population corresponding to* P_n*) of the* r*th most frequent type. Then* $X(n) := (X_1(n), \ldots, X_n(n), 0, 0, \ldots)$ *is a random element of* $\overline{\nabla}^{\infty}$ *with distribution* μ_n *and* $\mu_n \overset{Law}{\longrightarrow} \mu$ *as* $n \to \infty$.

See Kingman (1978b; 1980) for a proof.

Remark. Kingman (1982) and Aldous (1985) interpret a random partition in terms of an *exchangeable random partition* of the set of positive integers $\mathbb{N}$, that is, a random partition of $\mathbb{N}$ whose restriction Π_n to $[n] := \{1, \ldots, n\}$ has the following property: the partition of n induced by Π_n has distribution P_n and, given π_n, Π_n is uniformly distributed over all partitions of $[n]$ with component sizes given by π_n. In this context, see Aldous (1985) for an elegant proof of Kingman's theorem *à la* de Finetti, and Pitman (1992) for further developments.

Although Kingman's result characterizes random partitions via sampling from their corresponding random populations, these random populations are explicitly known and easily constructed in only a few cases. In the remainder of this paper we consider more tractable ways to construct random partitions.

4. Sequential construction of random partitions

Random sequences of 0s and 1s naturally induce random partitions of n by considering the components formed by the blocks of the sequence between successive 1s. We show that the only random partition induced from Bernoulli sequences is the family of Ewens' partition structures originally discovered in population genetics by Ewens (1972). Similarly, there is only one family of random partitions associated with discrete homogeneous renewal processes. This family of random partitions is a particular example of a two-parameter generalization of Ewens' partition due to Pitman (1995), which we call *Pitman's random partition*.

We consider a 'convolution'-type operator on random partitions and show that if the partitions have sequential representations, then the convolution of these partitions has a sequential representation as well. This representing sequence for the convolution of random partitions is also given by a corresponding convolution defined for random sequences. This implies that there is a large collection of random partitions that admit sequential representations by starting with the above-mentioned representable partitions and closing up under the operations of convolution and the taking of limits. Finally, it is shown that the convolution of Ewens' partition with the partition induced by the above-mentioned discrete homogeneous renewal process is precisely Pitman's random partition.

Let $\xi = (\xi_1, \xi_2, \ldots)$ be a sequence with $\xi \in \{0, 1\}$ and $\xi_1 = 1$. Define for each $n \geqslant 1$ a partition of $[n]$ by $\Pi_n(\xi) = \Pi(\xi_{1:n}) = (a_1, \ldots, a_n) \in \omega_n$, where a_i is the number of *i-spacings* between successive 1s in the finite sequence $\xi_{1:n} = (\xi_1, \ldots, \xi_n)$. More precisely,

$$a_i = \#\{j : \xi_j = 1, \xi_{j+1} = 0, \xi_{j+2} = 0, \ldots, \xi_{j+i-1} = 0, \xi_{j+i} = 1; 1 \leqslant j \leqslant n\}$$
$$+ 1\{\xi_{n-i+1} = 1, \xi_{n-i+2} = 0, \ldots, \xi_n = 0\}.$$

For example,

$$\Pi(1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0) = (1, 2, 1, 1, 1) \in \omega_{17}.$$

Definition 1 *Representable random partition*. A random partition P_n is (*sequentially*) *representable* if there exists a random sequence $\xi = (\xi_1, \xi_2, \ldots)$ with $\xi_i \in \{0, 1\}$ and $\xi_1 = 1$ such that $P_n(\pi) = P[\Pi_n(\xi) = \pi]$ for all $\pi \in \omega_n$ and $n \geqslant 1$. The sequence $\xi = (\xi_1, \xi_2, \ldots)$ is called the *representing sequence* for the random partition P_n.

If P_n is representable, then the components of the partition of n can be 'built' one by one by reading the 0s and 1s of the corresponding representing sequence. It is not at all clear whether a given random partition admits a sequential representation. We take the opposite approach and ask if there are random partitions which are representable by certain natural random sequences of 0s and 1s. The random sequences considered are Bernoulli sequences and (discrete) homogeneous renewal processes.

5. Bernoulli sequences and Ewens' random partition

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent Bernoulli random variables with success probability $p_n = P(\xi_n = 1)$ and $p_1 = 1$. Given n, the sequence (ξ_i) induces a random partition of n. Consider the finite sequence (recall $\xi_1 = 1$ almost surely)

$$\xi_1 = 1, \xi_2, \xi_2, \ldots, \xi_n$$

and define a_i as the number of i-spacings between successive 1s, including the last $10 \ldots 0$. For example, the sequence

$$1\,0\,1\,0\,0\,1\,0\,1\,0\,0\,0\,0\,1\,0\,0\,1\,1\,0\,1\,0\,0$$

induces the partition of $n = 21$ with $a_1 = 1,\ a_2 = 3,\ a_3 = 3,\ a_4 = 0,\ a_5 = 1$.

Let $\Pi_n := \Pi_n(\xi_1, \xi_2, \ldots)$ denote the random partition of n defined by the sequence $\xi_1, \xi_2, \ldots, \xi_n$, and let P_n be the distribution of Π_n.

Theorem 2. *Let $\xi_1, \xi_2, \ldots$ be a sequence of independent Bernoulli random variables with success probability $p_n = P(\xi_n = 1)$ for $n \geqslant 1$. Then the sequence (P_n) of distributions of the induced partitions (Π_n) forms a random partition if and only if*

$$p_n = \frac{\theta}{\theta + n - 1}, \qquad n \geqslant 1, \tag{4}$$

where $\theta = p_2/(1 - p_2)$. In this case (P_n) is Ewens' random partition with parameter θ:

$$\mathrm{ERP}_\theta(a_1, \ldots, a_n) = \frac{n!}{\prod_{i=1}^{n} i^{a_i} a_i!} \cdot \frac{\theta^{\sum_{i=1}^{n} a_i}}{\theta(\theta+1)\cdots(\theta+n-1)},$$

where a_i is the number of component blocks of size i. This representing sequence is called the Chinese restaurant sequence with parameter θ.

Proof. The sequence (P_n) forms a random partition if and only if the following consistency relation holds:

$$P_n(a_1, \ldots, a_n) = P_{n+1}(a_1 + 1, \ldots, a_{n+1}) \frac{a_1 + 1}{n+1} + \sum_{r=2}^{n+1} P_{n+1}(a_1, \ldots, a_{r-1} - 1, a_r + 1, \ldots, a_{n+1}) \frac{r(a_r + 1)}{n+1}.$$

In particular,

$$P_n(0, 0, \ldots, 0, 1) = P_{n+1}(0, \ldots, 0, 1) + P_{n+1}(1, 0, \ldots, 1, 0) \cdot \frac{1}{n+1} \tag{5}$$

In terms of the p_i this becomes

$$(1 - p_2)\cdots(1 - p_n) = (1 - p_2)\cdots(1 - p_n)(1 - p_{n+1}) + \frac{1}{n+1}[p_2(1 - p_3)\cdots(1 - p_{n+1}) + (1 - p_2)\cdots(1 - p_n)p_{n+1}].$$

Solving this equation for p_{n+1} yields

$$p_{n+1} = \frac{p_2}{p_2 + n(1 - p_2)}, \qquad \text{for } n \geqslant 0.$$

Setting $\theta = p_2/(1 - p_2)$ yields

$$p_{n+1} = \frac{\theta}{\theta + n}, \qquad \text{for } n \geqslant 0.$$

This is only a necessary condition, since (1) must hold for all $(a_1, \ldots, a_n) \in \omega_n$ and $n \geqslant 1$. To see that (4) is sufficient, proceed as in the Chinese restaurant process (Aldous 1985; Joyce and Tavaré 1987). Read the sequence $\xi_1, \xi_2, \ldots, \xi_n$ backwards to construct a random permutation of $\{1, 2, \ldots, n\}$ by building cycles one by one. Start the first cycle with a 1. If $\xi_n = 1$, close the cycle and start a new cycle with the next available (smallest) integer. If $\xi_n = 0$, choose an integer uniformly at random from the remaining $n - 1$ integers and place it to the right of 1. Continue in this way, by next looking at ξ_{n-1} and closing the cycle when $\xi_{n-1} = 1$ and continuing to build the cycle when $\xi_{n-1} = 0$, etc. A random permutation so constructed has the following distributional properties:

$$P(\text{permutation has } k \text{ cycles}) = \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)}$$

and

$$\begin{aligned} P[\Pi_n = (a_1, \ldots, a_n)] &= P[\text{permutation has cycle structure } (a_1, \ldots, a_n)] \\ &= \#\{\text{permutations with cycle structure } (a_1, \ldots, a_n)\} \\ &\quad \times \left(\frac{\theta^{\Sigma a_i}}{\theta(\theta+1)\cdots(\theta+n-1)}\right) \\ &= \frac{n!}{\prod_{i=1}^{n} i^{a_i} a_i!} \cdot \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \end{aligned}$$

where $k = \sum_{i=1}^{n} a_i$ is the number of cycles of the random permutation. (A permutation has cycle structure $(a_1, \ldots, a_n)$ if there are a_i cycles of length i for $i \geqslant 1$.) □

Remark. For $\theta = 1$, Feller (1945) used this Bernoulli sequence representation to obtain central limit type asymptotics for the number of cycles in random permutations. Arratia and Tavaré (1992) extend the use of the Bernoulli representation to general θ and obtain Poisson asymptotics for the component counts $(a_1, \ldots, a_n)$ which are sharp enough to imply central limit asymptotics as well. See also Donnelly *et al.* (1991b) and Arratia *et al.* (1992).

6. Discrete renewal processes

Instead of considering spacings in independent sequences of 0s and 1s, consider the waiting time between successive 1s in a discrete renewal process. This gives a (homogeneous) discrete renewal process characterization of a two-parameter extension of Ewens' partition with $0 < \alpha < 1$ and $\theta = 0$, originally discovered by Pitman (1995) in a different context. We can recover the full two-parameter case by combining the independent 0–1 sequence with the discrete renewal process as discussed below.

Let $T_1, T_2, \ldots$ be independent and identically distributed (i.i.d.) integer-valued waiting-time random variables with distribution $P(T = n) = p_n$ for $n \geqslant 1$. We think of the Ts as the waiting time between successive 1s in a random sequence which begins with 1, almost surely, followed by $(T_1 - 1)$ 0s, then another 1 followed by $(T_2 - 1)$ 0s, ... For example, the ξ-sequence corresponding to

$$T_1 = 3,\ T_2 = 2,\ T_3 = 5,\ T_4 = 1,\ T_5 = 3,\ T_6 = 4$$

is

$$1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0.$$

The partition of n may be defined directly in terms of the T-sequence as follows.

Let

$$k = k(n) = \min\left\{ j : \sum_{i=1}^{j} T_i \geqslant n \right\}$$

and define

$$a_j = \#\{i : T_i = j,\ 1 \leqslant i \leqslant k-1\} + 1\{T_k^* = j\},$$

where

$$T_k^* = n - \sum_{i=1}^{k-1} T_i.$$

Let $\Pi_n = \Pi_n(T_1, T_2, \ldots)$ denote this random partition of n and let P_n be the distribution of Π_n.

Theorem 3. *Let $T_1, T_2, \ldots$ be a sequence of i.i.d. positive integer-valued random variables with distribution $P(T = n) = p_n$ for $n \geqslant 1$. Then the sequence (P_n) of distributions of the induced partitions (Π_n) forms a random partition if and only if*

$$p_n = (-1)^{n-1}\binom{p_1}{n}, \qquad \text{for } n \geqslant 1. \tag{6}$$

Letting $\alpha = p_1$, the corresponding random partition is given by

$$P[\Pi_n = (a_1, \ldots, a_n)] = \frac{n!}{\prod_{i=1}^{n}(i!)^{a_i} a_i!} \cdot \frac{(k-1)!\alpha^{k-1}}{(n-1)!} \prod_{i=1}^{n}[(1-\alpha)(2-\alpha)\cdots(i-1-\alpha)]^{a_i}.$$

Proof of necessity. Equation (5) is a necessary condition for the sequence (P_n) to be a random partition. In terms of the T_i this becomes

$$P(T_1 > n-1) = P(T_1 > n) + [P(T_1 = 1,\ T_2 > n-1) + P(T_1 = n)]\frac{1}{n+1},$$

$$p_n = \frac{1}{n+1}[p_1(1 - p_1 - p_2 - \ldots - p_{n-1}) + p_n] \tag{7}$$

$$= \frac{p_1(1 - p_1 - p_2 - \ldots - p_{n-1})}{n}$$

$$p_n = \frac{p_1(1 - p_1 - \ldots - p_{n-2})}{n} - \frac{p_1 p_{n-1}}{n}. \tag{8}$$

Iterating (8), we have

$$p_n = \frac{(n-1)p_{n-1} - p_1 p_{n-1}}{n}$$
$$= \frac{(n-1-p_1)}{n} \cdot p_{n-1} = \frac{(n-1-p_1)}{n} \cdot \frac{(n-2-p_1)}{n-1} \cdot p_{n-2} = \cdots$$
$$= \frac{(n-1-p_1)(n-2-p_1)\cdots(1-p_1)(p_1)}{n!}$$
$$p_n = (-1)^{n-1}\binom{p_1}{n}. \tag{9}$$

For convenience of notation, we set $\alpha = p_1$, so that (9) becomes

$$p_n = (-1)^{n-1}\binom{\alpha}{n}, \qquad \text{for } n \geqslant 1 \text{ and } 0 < \alpha < 1. \tag{10}$$

Remark. Now, a priori, it is not clear that the solution to (7) defines a proper distribution for T, that is, $\sum_{n=1}^{\infty} p_n = 1$. But this is confirmed by recalling that (10) is a familiar discrete distribution in the theory of random walks: (10) is the distribution of the ladder indices $(\min\{n : S_n > 0\})$ for a real random walk S_n with $P(S_n > 0) = \alpha$ for all $n \geqslant 1$, for example, a walk with stable increments of any index where $P(\text{increment} > 0) = \alpha$. For $\alpha = 1/2$, (10) is the distribution of half the return time to zero for a simple symmetric random walk (Feller 1971).

To see that (10) is also sufficient for (P_n) to be a random partition, we turn now to partitions derived from excursions of recurrent Bessel processes.

7. Partitions of Bessel processes

Let $(B_t,\ t \geqslant 0)$ with $B_0 = 0$ be a Bessel process of dimension $\delta = 2 - 2\alpha,\ 0 < \alpha < 1$; see Revuz and Yor (1999) and Rogers and Williams (2000) for background. The zero set of the process $(B_t,\ 0 \leqslant t \leqslant 1)$ is a random closed subset of [0, 1]. The interval components of the complement of this zero set will be called *excursion intervals*. Note that this allows a final *meander interval* of the form $(L_1, 1]$, where L_1 is the last zero of B_t before time 1. Let $U_1, U_2, \ldots$ be a sequence of i.i.d. uniform random variables on [0, 1], independent of B_t. Define a random equivalence relation $\sim$ on the positive integers N by $i \sim j$ if and only if U_i and U_j fall in the same excursion interval. The collection of $\sim$-equivalence classes is then an exchangeable random partition of N. In this context, Kingman's representation theorem (Kingman 1978b) says that every random partition can be associated with an exchangeable random partition of N obtained by the above construction for some random closed subset of [0, 1]. See Aldous (1985) for a proof and Pitman (1995) for further developments.

When the random closed subset is the zero set of a recurrent Bessel process, the corresponding random partition is the $(\alpha, 0)$ random partition.

Theorem 4 (Pitman (1997). *Fix $n \geqslant 1$, and for $1 \leqslant i \leqslant n$, let a_i be the number of excursion intervals of $(B_t, 0 \leqslant t \leqslant 1)$ that contain exactly i of the n points $U_1, U_2, \ldots, U_n$, so that $(a_1, \ldots, a_n) \in \omega_n$. Then*

$$P(a_1, \ldots, a_n) = \frac{n!}{\prod_{i=1}^{n} (i!)^{a_i} a_i!} \cdot \frac{(k-1)! \alpha^{k-1}}{(n-1)!} \prod_{i=1}^{n} [(1-\alpha)(2-\alpha) \cdots (i-1-\alpha)]^{a_i} \tag{11}$$

forms a random partition.

With this theorem in hand, we return to the proof of Theorem 3.

Proof of sufficiency of Theorem 3. Let $(B_t, t \geqslant 0)$ with $B_0 = 0$ be a Bessel process of dimension $\delta = 2 - 2\alpha$, $0 < \alpha < 1$, and let $U_1, U_2, \ldots, U_n$ be i.i.d. random variables uniformly distributed on [0, 1] and independent of B_t. Define a sequence of indicators

$$Z_i^n = 1\{B_t = 0 \text{ for some } U_{(i)} < t < U_{(i+1)}\}, \qquad \text{for } i = 1, \ldots, n-1,$$

$$Z_0^n = 1\{B_t = 0 \text{ for some } 0 < t < U_{(1)}\} \overset{\text{a.s.}}{=} 1,$$

$$Z_n^n = 1\{B_t = 0 \text{ for some } U_{(n)} < t < 1\},$$

where $U_{(1)} < U_{(2)} < \ldots < U_{(n)}$ are the order statistics of the U_i.

The joint distribution of the random variables $Z_0^n, Z_1^n, \ldots, Z_n^n$ can be obtained using the standard representation of uniform order statistics in terms of a Poisson process. Let $\tau_0 = 0$ and $\tau_n = \eta_1 + \eta_2 + \ldots + \eta_n$, where $\eta_1, \eta_2, \ldots$ is a sequence of i.i.d. random variables with exponential mean 1 distribution, defined on the same probability space as the Bessel process B_t, and independent of B_t. Define analogous indicators

$$Z_i = 1\{B_t = 0 \text{ for some } \tau_i < t < \tau_{i+1}\}, \qquad \text{for } i = 0, 1, \ldots.$$

It is a standard fact that $(U_{(1)}, \ldots, U_{(n)})$ has the same distribution as $(\tau_1/\tau_{n+1}, \ldots, \tau_n/\tau_{n+1})$, and since the zero set of B_t is invariant under scaling, $(Z_0^n, \ldots, Z_n^n)$ and $(Z_0, \ldots, Z_n)$ have the same distribution.

Now consider an excursion interval that contains at least one of the times τ_i. Call such an interval a *hit excursion interval*. Let T_j be the number of τ_i that fall in the jth hit excursion interval. Then clearly

$$Z_i = 1\{T_1 + \ldots + T_m = i, \text{ for some } m\}$$

are renewal indicators. The fact that the T_i are i.i.d. follows from the strong Markov property of B_t at the right endpoint of such a hit excursion interval together with the memoryless property of the exponential distribution. From results of standard renewal theory (Feller 1971), it follows that the distribution of the T_i is given precisely by $P(T = n) = (-1)^{n-1} \binom{\alpha}{n}$.

Therefore, the partition distribution of $\Pi_n = \Pi_n(T_1, T_2, \ldots)$ is given by (11) and hence defines a random partition. □

The random partition given by (11) is a special case of a two-parameter generalization of Ewens' random partition due to Pitman (1995).

Definition 2 *Pitman's random partition*. Pitman's random partition with parameters (α, θ), where $0 \leqslant \alpha < 1$ and $\theta > -\alpha$, is given by

$$P_{(\alpha,\theta)}(a_1, \ldots, a_n) = \frac{n!}{\prod_{i=1}^{n}(i!)^{a_i} a_i!} \cdot \frac{[\theta + \alpha]_{k-1;\alpha}}{[\theta + 1]_{n-1}} \prod_{i=1}^{n} ([1 - \alpha]_{i-1})^{a_i}, \tag{12}$$

where

$$[x]_{m;a} := \begin{cases} x(x + a) \cdots (x + (m - 1)a), & \text{for } m = 1, 2, \ldots, \\ 1, & \text{for } m = 0, \end{cases}$$

and

$$[x]_m := [x]_{m;1}.$$

Remark. Ewens' partition with parameter θ corresponds to $\alpha = 0$, $\theta > 0$, and the random partition derived from discrete homogeneous renewal processes corresponds to $0 < \alpha < 1$, $\theta = 0$. See Pitman (1995) for various descriptions and asymptotics of the (α, θ) partition. For a full treatment of the law of the corresponding ranked relative frequencies of Pitman's random partition, see Pitman and Yor. Also see Perman *et al.* (1992) and Pitman (1996) for the residual allocation model for the relative frequency of classes in order of appearance.

8. Convolution of random partitions

In this section, a convolution-type product is defined on the family of all random partitions. This convolution operator preserves (sequentially) representable partitions. The representing sequence corresponding to the convolution of representable random partitions is itself a convolution-type product of random 0–1 sequences. As a concrete example, we show that the random partition obtained by convolving Ewens' partition with parameter θ with the partition with parameter $(\alpha, 0)$ is the random partition with parameter (α, θ), provided both α and θ are strictly positive.

Let $P = (P_n)$ and $Q = (Q_n)$ be the distributions of two random partitions and define a new partition distribution $P * Q = R$ as follows. First partition $[n]$ according to P_n, and then partition each of the component blocks of this partition using Q_k for $k \leqslant n$, independently given the blocks of the first partition of $[n]$ via P_n.

We now show that convolution preserves the representing sequences X and Y corresponding to P and Q, respectively. To construct the corresponding representing

sequence $(X * Y)_n$ for $(P * Q)_n$, first lay down the sequence $(X_1, X_2, \ldots)$, and then in between 1s of the X-sequence lay down independent copies of $(Y_2, Y_3, \ldots)$. (Recall that the first term in representing sequences equals 1 almost surely.) For example, suppose X is given by

$$1000101001000001001100 0100 \ldots .$$

Then the sequence $(X * Y)_n$ will look like

$$1Y_2^1Y_3^1Y_4^11Y_2^21Y_2^3Y_3^31Y_2^4Y_3^4Y_4^4Y_5^4Y_6^41Y_2^5Y_3^511Y_2^6Y_3^6Y_4^61Y_2^7Y_3^7 \ldots,$$

where the blocks (Y_n^k) are independent copies of the sequence Y_n. More precisely, define stopping times $\iota_1 = 1$ and $\iota_k = \min\{n > \iota_{k-1} : X_n = 1\}$ and define a new sequence $Z_n = (X * Y)_n$ block by block as follows:

$$(Z_1, \ldots, Z_{\iota_2-1}) = (Y_1^1, \ldots, Y_{\iota_2-1}^1),$$

$$(Z_{\iota_2}, \ldots, Z_{\iota_3-1}) = (Y_1^2, \ldots, Y_{\iota_3-1}^2), \ldots$$

$$(Z_{\iota_k}, \ldots, Z_{\iota_{k+1}-1}) = (Y_1^k, \ldots, Y_{\iota_{k+1}-1}^k), \ldots,$$

where the blocks (Y_n^k) are independent copies of the sequence Y_n. It is clear from the above construction that $Z_n = (X * Y)_n$ is the representing sequence for the random partition $(P * Q)_n$. We summarize this in the following theorem:

Theorem 5. *Let P_n and Q_n be partition distributions with respective representing sequences X_n and Y_n. Then $(P * Q)_n$ has representing sequence $(X * Y)_n$.*

By a straightforward calculation of partitions of small n, say $n \leqslant 5$, the convolution product is easily seen to be non-commutative. In one non-trivial case, we can explicitly compute the convolution product.

Theorem 6. *Let P_n be Ewens' partition with parameter $\theta > 0$ and Q_n be the partition with parameter $(\alpha, 0)$, $\alpha > 0$. Then $(P * Q)_n$ is the random partition with parameter (α, θ).*

Proof. By an urn scheme construction due to Pitman (1997), we may interpret the (α, θ) partition as an exchangeable random partition $\Pi^{(\alpha,\theta)}$ of $\mathbb{N}$ as follows. The probability that $\Pi_n^{(\alpha,\theta)}$ equals any particular partition of $\{1, \ldots, n\}$ into k blocks A_i of sizes n_i, $i = 1, \ldots, k$, is

$$P(\Pi_n^{(\alpha,\theta)} = \{A_i\}_1^k) = \frac{(\theta+\alpha)\cdots(\theta+(k-1)\alpha)}{(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k} (1-\alpha)\cdots(n_i-1-\alpha).$$

When $\alpha = 0$, this gives the partition of $\{1, \ldots, n\}$ induced by the Blackwell–McQueen–Hoppe urn scheme (Blackwell and MacQueen 1973). Let P_n be the $(0, \theta)$ partition and Q_n be the $(\alpha, 0)$ partition. The partition $(P * Q)_n$ of n corresponds to the following partition of $\{1, \ldots, n\}$. First, partition $\{1, \ldots, n\}$ according to a $(0, \theta)$ random partition. Next,

independently partition each component of the $(0, \theta)$ random partition according to an $(\alpha, 0)$ random partition. Let Π_n denote the resulting random partition of $\{1, \ldots, n\}$. Then the probability that Π_n equals any particular partition of $\{1, \ldots, n\}$ into k classes A_i of sizes n_i, $i = 1, \ldots, k$, is

$$P(\Pi_n = \{A_i\}_1^k) = \sum_{j=1}^{k} \sum_{\{C_i, 1 \leqslant i \leqslant j\}} \frac{\theta^j}{[\theta]_{n-1}} \prod_{i=1}^{j} (m_i - 1)! \prod_{i=1}^{j} \frac{[\#(C_i) - 1]! \alpha^{\#(C_i)-1}}{(m_i - 1)!} \prod_{i=1}^{k} [1 - \alpha]_{n_i - 1}$$

$$= \frac{1}{[\theta]_{n-1}} \sum_{j=1}^{k} \theta^j \sum_{\{C_i, 1 \leqslant i \leqslant j\}} \prod_{i=1}^{j} [\#(C_i) - 1]! \alpha^{\#(C_i)-1} \prod_{i=1}^{k} [1 - \alpha]_{n_i - 1},$$

where $\sum_{\{C_i, 1 \leqslant i \leqslant j\}}$ denotes the sum over all partitions $\{C_i, 1 \leqslant i \leqslant j\}$ of the set $\{1, \ldots, k\}$ into j parts. But by a standard identity (see, for example, Goulden and Jackson 1983) for the generating function of the number of cycles in a random permutation,

$$\sum_{j=1}^{k} \theta^j \sum_{\{C_i, 1 \leqslant i \leqslant j\}} \prod_{i=1}^{j} [\#(C_i) - 1]! \alpha^{\#(C_i)-1} = \theta(\theta + \alpha) \cdots (\theta + (k-1)\alpha),$$

we obtain

$$P(\Pi_n = \{A_i\}_1^k) = \frac{(\theta + \alpha) \cdots (\theta + (k-1)\alpha)}{(\theta + 1) \cdots (\theta + n - 1)} \prod_{i=1}^{k} (1 - \alpha) \cdots (n_i - 1 - \alpha),$$

which is the (α, θ) random partition of $\{1, \ldots, n\}$ corresponding to the (α, θ) partition of n.

If we consider the corresponding representing sequences for these random partitions we get the following result.

Corollary. *Let X_n denote the Chinese restaurant sequence* (4) *with parameter θ and let Y_n denote the discrete homogeneous renewal sequence* (10) *corresponding to the* $(\alpha, 0)$ *partition. Then $(X * Y)_n$ is the representing sequence for the* (α, θ) *partition, provided $\alpha, \theta > 0$.*

We have seen that the full two-parameter (α, θ) random partition admits a sequential representation. However, aside from trivial partitions, this is the only explicit family known which can be constructed via 0–1 sequences. It would be nice to find other random partitions which allow such sequential constructions or, conversely, to construct novel random 0–1 sequences representing random partitions.

An ultimate objective remains to determine natural necessary and sufficient conditions for random partitions to admit sequential constructions like the random 0–1 sequences discussed above.

References

Aldous, D.J. (1985) Exchangeability and related topics. In P.-L. Hennequin (ed.), *Ecole d'Été de Probabilités de Saint-Flour XIII*. Lecture Notes in Math. 1117. Berlin: Springer-Verlag.

Aldous, D.J. and Pitman, J. (1993) Brownian bridge asymptotics for random mappings. *Random Structures Algorithms*, **5**, 487–512.

Arratia, R.A. and Tavaré, S. (1992) Limit theorems for combinatorial structures via discrete process approximations. *Random Structures Algorithms*, **3**, 321–345.

Arratia, R.A., Barbour, A.D. and Tavaré, S. (1992) Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.*, **2**, 519–535.

Arratia, R.A., Barbour, A.D. and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach*. Zurich: European Mathematical Society.

Blackwell, D. and MacQueen, J.B. (1973) Ferguson distributions via Pólya urn schemes. *Ann. Statist*., **1**, 353–355.

Donnelly, P. and Grimmett, G. (1993) On the asymptotic distribution of large prime factors. *J. London Math. Soc. (2)*, **47**, 395–404.

Donnelly, P., Ewens, W.J. and Padmadisastra, S. (1991a) Functionals of random mappings: Exact and asymptotic results. *Adv. Appl. Probab.*, **23**, 437–455.

Donnelly, P., Kurtz, T.G. and Tavaré, S. (1991b) On the functional central limit theorem for the Ewens sampling formula. *Ann. Appl. Probab.*, **1**, 539–545.

Ethier, S.N. and Kurtz, T.G. (1981) The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.*, **13**, 429–452.

Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoret. Popul. Biol.*, **3**, 87–112.

Feller, W. (1945) The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.*, **51**, 800-832.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications, Volume II* (2nd edn). New York: Wiley.

Foata, D. (1974) *La série génératrice exponentielle dans les problèmes d'énumération*, Séminaire de Mathématiques Supérieures 54. Montreal: Presses de l'Université de Montréal.

Goulden, I.P. and Jackson, D.M. (1983) *Combinatorial Enumeration*. New York: Wiley.

Hoppe, F.M. (1987) The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.*, **25**, 123–159.

Joyce, P. and Tavaré, S. (1987) Cycles, permutations, and the structure of the Yule process with immigration. *Stochastic Process. Appl.*, **25**, 309–314.

Kingman, J.F.C. (1978a) Random partitions in population genetics. *Proc. Roy Soc. Lond. Ser. A*, **361**, 1–20.

Kingman, J.F.C. (1978b) The representation of partition structures. *J. London Math. Soc. (2)*, **18**, 374–380.

Kingman, J.F.C. (1980) *Mathematics of Genetic Diversity.* Philadelphia: Society of Industrial and Applied Mathematics.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Process. Appl.*, **13**, 253–248.

Mekjian, A.Z. (1990) Distribution of cluster sizes from evaporation to total multifragmentation. *Phys. Rev. C*, **41**, 2103–2117.

Mekjian, A.Z. (1991) Cluster distributions in physics and genetic diversity. *Phys. Rev. A*, **44**, 8361–8374.

Mekjian, A.Z. and Lee, S.J. (1991) Models of fragmentation and partitioning phenomena based on the symmetric group S_n and combinatorial analysis. *Phys. Rev. A*, **44**, 6294–6312.

Perman, M., Pitman, J. and Yor, M. (1992) Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields*, **92**, 21–39.

Pitman, J. (1992) The two-parameter generalization of Ewens' random partition structure. Technical Report 345, Department of Statistics, University of California, Berkeley.

Pitman, J. (1995) Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, **102**, 145–158.

Pitman, J. (1996) Random discrete distributions invariant under size-biased permutation. *Adv. Appl. Probab.*, **28**, 525–539.

Pitman, J. (1997) Partition structures derived from Brownian motion and stable subordinators. *Bernoulli*, **3**, 79–96.

Pitman, J. (2002) *Combinatorial Stochastic Processes.* Lecture notes for St. Flour course.

Pitman, J. and Yor, M. (1997) The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**, 855–900.

Revuz, D. and Yor, M. (1999) *Continuous Martingales and Brownian Motion* (3rd edn). Berlin: Springer-Verlag.

Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov Processes, and Martingales, Vol. 2: Itô Calculus* (2nd edn). Cambridge: Cambridge University Press.

Vershik, A.M. (1987) The asymptotic distribution of factorizations of natural numbers into prime divisors. *Soviet Math. Dokl.*, **34**, 57–61.

Watterson, G.A. (1974) The sampling theory of selectively neutral alleles. *Adv. Appl. Probab.*, **6**, 463–488.

Zabell, S.L. (1992) Predicting the unpredictable. *Synthese*, **90**, 205–232.

Received August 2004 and revised February 2005