

REVERSIBLE MARKOV STRUCTURES ON DIVISIBLE SET PARTITIONS

HARRY CRANE,* *Rutgers University*

PETER MCCULLAGH,** *University of Chicago*

Abstract

We study k -divisible partition structures, which are families of random set partitions whose block sizes are divisible by an integer $k = 1, 2, \dots$. In this setting, exchangeability corresponds to the usual invariance under relabeling by arbitrary permutations; however, for $k > 1$, the ordinary deletion maps on partitions no longer preserve divisibility, and so a random deletion procedure is needed to obtain a partition structure. We describe explicit Chinese restaurant-type seating rules for generating families of exchangeable k -divisible partitions that are consistent under random deletion. We further introduce the notion of *Markovian partition structures*, which are ensembles of exchangeable Markov chains on k -divisible partitions that are consistent under a random process of *Markovian deletion*. The Markov chains we study are reversible and refine the class of Markov chains introduced in *J. Appl. Probab.* **48**(3):778–791.

Keywords: Markovian partition structure; exchangeable partition structure; Ewens–Pitman partition; Chinese restaurant process; divisible partition; group-divisible association scheme

2010 Mathematics Subject Classification: Primary 60C05, 60J10

Secondary 60B99

* Postal address: Rutgers University

Department of Statistics

110 Frelinghuysen Road

Piscataway, NJ 08854, USA

** Postal address: University of Chicago

Department of Statistics

Eckhart Hall

5734 S. University Ave.

Chicago, IL 60637

1. Introduction

A *partition* π of $[n] := \{1, \dots, n\}$ is a collection $B_1/\dots/B_m$ of non-empty, disjoint subsets, called blocks, for which $\bigcup_{1 \leq j \leq m} B_j = [n]$. For $n, k \geq 1$, we call a partition $\pi = B_1/\dots/B_m$ of $[nk]$ *k-divisible*, or just *divisible*, if the cardinality of each block B_1, \dots, B_m is divisible by k . When $k = 2$, we call π an *even partition*.

Divisible partitions are natural in ecological applications as well as randomization in experimental design. For example, in experimental design, each of nk individuals is assigned one of k treatments. If individuals are further grouped into blocks so that every treatment is assigned the same number of times within each block, then the block structure of the design is a divisible partition of $[nk]$. In this setting, divisible partitions are related to group-divisible association schemes; see Bailey [1] for further connections between experimental design and the theory of partitions. Our study of probabilistic structures of divisible random partitions is motivated by the above heuristic as well as the appeal of partition models to applications in clustering and classification [5, 17], population genetics [12, 13], and linguistics [10, 11].

We study ensembles of random divisible partitions whose distributions are consistent under a random deletion operation. Kingman [15, 16] first studied the deletion properties of random integer partitions. He defined a *partition structure* as a collection $\mathbb{P} := (\mathbb{P}_1, \mathbb{P}_2, \dots)$ of probability distributions on the spaces $(\mathcal{P}_n, n \in \mathbb{N})$ of integer partitions of each $n = 1, 2, \dots$ such that if $\lambda' \in \mathcal{P}_{n-1}$ is obtained by choosing a part of $\lambda \sim \mathbb{P}_n$ with probability proportional to its size and reducing it by one, then $\lambda' \sim \mathbb{P}_{n-1}$. On set partitions, consistency of a family of distributions \mathbb{P} is defined through the usual non-random restriction operation. Any consistent collection determines a unique probability measure on partitions of $\mathbb{N} := \{1, 2, \dots\}$. Gnedin, et al [14] study further deletion properties of random partitions.

In our main theorems, we extend the Chinese restaurant process [18] and the Ewens–Pitman Markov chain [4] to the space of k -divisible partitions, and we obtain the finite-dimensional distributions of these processes. By reversing these procedures, we describe natural deletion mechanisms under which the prescribed finite-dimensional distributions are consistent. In the Markov chain case, this produces a *Markovian partition structure*, that is, a family $\{(\varepsilon^{(n)}, \mathcal{E}^{(n)})\}_{n \geq 1}$ of initial distributions $\varepsilon^{(n)}$ and transition

probability measures $\mathcal{E}^{(n)}$ on k -divisible partitions of $[nk]$ that are exchangeable and consistent under a Markovian deletion scheme. Under this operation, Markovian partition structures have the form

$$\begin{array}{ccc} \mathcal{P}_{[nk] | k} & \longrightarrow & \mathcal{P}_{[nk] | k} \\ \downarrow & & \downarrow \\ \mathcal{P}_{[(n-1)k] | k} & \longrightarrow & \mathcal{P}_{[(n-1)k] | k}, \end{array}$$

where horizontal arrows denote Markov transitions in time and vertical arrows represent randomized projections by the Markovian deletion scheme. In particular, the marginal distributions at fixed times, as n varies, are consistent; and the marginal distribution of each sequence, for fixed n , is a Markov chain. For a single time t , the marginal behavior of the Markovian deletion scheme coincides with the deletion scheme for k -divisible partition structures. We provide the details in Sections 3.2 and 3.4.

To obtain a characteristic measure $\varepsilon^{(\infty)}$ on the limit space of partitions of \mathbb{N} , we need a deterministic deletion operation; but there is no such operation in the k -divisible setting with $k > 1$. When $k > 1$, simple deletion of the highest labeled group $\{nk + 1, \dots, (n+1)k\}$ from an exchangeable divisible partition of $[(n+1)k]$ does not preserve divisibility; therefore, a random deletion scheme is needed. For example, the partition $\pi = 1468/27/35$ is 2-divisible, but the partition $\pi' = 146/2/35$ obtained by deleting elements $\{7, 8\}$ is not 2-divisible because the cardinalities of $\{1, 4, 6\}$ and $\{2\}$ are not even.

Though the processes we study are not sampling consistent in the ordinary sense, the finite-dimensional processes we generate are exchangeable and, in the Markov chain case, reversible with respect to the k -divisible extension of the Ewens–Pitman distribution [12, 18]. Reversible processes for partition-valued Markov chains have been studied previously; see [3, 4].

Our main discussion focuses on the analog to the Ewens distribution for divisible partitions and Markov chains; however, these conclusions apply more generally to any paintbox measure. We develop these ideas formally in Sections 3.2 and 3.3. We make some concluding remarks in Section 4.

2. Preliminaries: Random partitions

A partition \mathbf{n} of $n = 1, 2, \dots$ is a non-increasing collection of positive integers $\mathbf{n} = (n_1, \dots, n_k)$, called *parts* or *summands*, whose sum is n . Alternatively, \mathbf{n} can be expressed in terms of its multiplicities $(\lambda_1, \dots, \lambda_n)$, $\mathbf{n} = 1^{\lambda_1} 2^{\lambda_2} \dots n^{\lambda_n}$, such that $\sum \lambda_j = k$ is the number of parts and $\sum j\lambda_j = n$. For each $n \in \mathbb{N}$, the Ewens sampling formula with parameter $\theta > 0$, $\text{ESF}(\theta)$, is the probability distribution on integer partitions of n with closed form expression

$$p_n(\mathbf{n}; \theta) = \frac{\theta^{\#\mathbf{n}}}{\theta^{\uparrow n}} \frac{n!}{\prod_{j=1}^n j^{\lambda_j} \lambda_j!}, \quad \mathbf{n} = 1^{\lambda_1} \dots n^{\lambda_n} \in \mathcal{P}_n, \quad (1)$$

where $\#\mathbf{n} := \sum_{j=1}^n \lambda_j$ is the number of parts of \mathbf{n} and $\theta^{\uparrow n} := \theta(\theta+1)\dots(\theta+n-1)$. Ewens [12] first derived (1) while studying the sampling theory of neutral alleles, but the Ewens sampling formula also occurs in purely mathematical contexts, e.g., as the asymptotic distribution of large prime factors [9] and as a special case of the α -permanent of a matrix [6, 7].

Ewens's sampling formula more naturally resides on the space $\mathcal{P}_{[n]}$ of partitions of $[n]$, where it is the $(0, \theta)$ sub-family of the two-parameter Ewens–Pitman (α, θ) family with finite-dimensional marginal distributions

$$\varrho_{\alpha, \theta}^{(n)}(\pi) = \frac{(\theta/\alpha)^{\uparrow \#\pi}}{\theta^{\uparrow n}} \prod_{b \in \pi} -(-\alpha)^{\uparrow \#b}, \quad \pi \in \mathcal{P}_{[n]}, \quad (2)$$

where $\#\pi$ denotes the number of blocks of π , $\#b$ the cardinality of block b , and (α, θ) satisfies either

- $\alpha = -\kappa < 0$ and $\theta = m\kappa$, for $m = 1, 2, \dots$, or
- $0 \leq \alpha \leq 1$ and $\theta > -\alpha$.

For fixed (α, θ) , $\varrho_{\alpha, \theta}^{(n)} := (\varrho_{\alpha, \theta}^{(n)}, n \in \mathbb{N})$ is a consistent collection of exchangeable probability measures on the system $(\mathcal{P}_{[n]}, n \in \mathbb{N})$. In particular, $(\varrho_{\alpha, \theta}^{(n)}, n \in \mathbb{N})$ is

- *exchangeable* if, for every $n \in \mathbb{N}$, $\varrho_{\alpha, \theta}^{(n)}(\pi)$ depends on π only through its block sizes, and
- *consistent under subsampling* if, for all $m \leq n$, the image measure of $\varrho_{\alpha, \theta}^{(n)}$ by restriction to $\mathcal{P}_{[m]}$ is $\varrho_{\alpha, \theta}^{(m)}$; that is, $\varrho_{\alpha, \theta}^{(m)} = \varrho_{\alpha, \theta}^{(n)} \mathbf{R}_{m, n}^{-1}$, where $\mathbf{R}_{m, n} : \mathcal{P}_{[n]} \rightarrow \mathcal{P}_{[m]}$

denotes the restriction map,

$$\mathbf{R}_{m,n}\pi := \{B_1 \cap [m], \dots, B_l \cap [m]\} \setminus \{\emptyset\}, \quad \pi = B_1 / \dots / B_l \in \mathcal{P}_{[n]}.$$

As a result, the finite-dimensional marginals in (2) determine a unique probability measure $\varrho_{\alpha,\theta}$ on $\mathcal{P}_{\mathbb{N}}$, called the Ewens–Pitman(α, θ) distribution. Throughout the paper, we assume that a pair (α, θ) is always within the parameter space of the Ewens–Pitman model, and we call any random partition with finite-dimensional distributions (2) an (α, θ) -partition. The distribution on $\mathcal{P}_{[n]}$ corresponding to $\text{ESF}(\theta)$ in (1) is Ewens–Pitman($0, \theta$).

Sampling consistency of the Ewens–Pitman family is easily observed through its *Chinese restaurant* construction. We construct a sequence $\Pi := (\Pi_1, \Pi_2, \dots)$ of finite exchangeable partitions by putting $\Pi_1 = \{1\}$ and, given $\Pi_n = \pi = B_1 / \dots / B_m \in \mathcal{P}_{[n]}$, we generate Π_{n+1} by inserting the element $n+1$

- in occupied block $b \in \pi$ with probability $(\#b - \alpha)/(n + \theta)$ and
- in its own block of Π_{n+1} with probability $(\theta + m\alpha)/(n + \theta)$.

We write $\text{CRP}(n, \alpha, \theta)$ to denote the conditional probability distribution of Π_{n+1} given Π_n above.

More generally, Kingman’s paintbox process [15] describes the law of any exchangeable partition of \mathbb{N} . Let ν be a probability measure on

$$\Delta^\downarrow := \left\{ (s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0, \sum_j s_j \leq 1 \right\}.$$

Given $(S_1, S_2, \dots) \sim \nu$ and $S_0 := 1 - \sum_j S_j$, we generate a sequence $X := (X_1, X_2, \dots)$ of conditionally independent random variables with

$$P\{X_i = j \mid S\} = \begin{cases} S_j, & j \geq 1 \\ S_0, & j = -i \\ 0, & \text{otherwise.} \end{cases}$$

The partition Π of \mathbb{N} defined by

$$i \text{ and } j \text{ are in the same block of } \Pi \quad \text{if and only if} \quad X_i = X_j$$

is exchangeable and obeys (*Kingman's paintbox distribution*, or *paintbox process*, directed by ν , denoted $\Pi \sim \varrho_\nu$. The Ewens–Pitman(α, θ) law corresponds to the paintbox process directed by the Poisson–Dirichlet(α, θ) law on Δ^\downarrow .

In the next section, we give Chinese restaurant-type constructions for divisible partitions. Through this process, we construct a partition structure on exchangeable k -divisible partitions of $[nk]$. We also introduce Markovian partition structures, which are families of Markov chains consistent under a random *Markovian* deletion process. This extends Kingman's partition structures to sequences of exchangeable k -divisible partitions and further refines the family of exchangeable Markov chains studied in [4].

3. Divisible partitions

We call any subset $A \subseteq \mathbb{N}$ *k-divisible*, or just *divisible*, if $\#A$ is divisible by k , denoted $A \mid k$. We call a partition π *k-divisible* if each of its blocks is k -divisible, denoted $\pi \mid k$, and we write $\mathcal{P}_{[nk] \mid k}$ to denote the space of k -divisible partitions of $[nk]$. For example, $\pi = 148/235679$ is k -divisible for $k = 1$ and $k = 3$.

3.1. Chinese restaurant construction for divisible partitions

Let $k \in \mathbb{N}$, (α, θ) be fixed, and \mathbb{N} be a population of individuals. Suppose individuals arrive at a restaurant in groups of size k , regarded as $\{ik + 1, \dots, ik + k\}$ for $i = 1, 2, \dots$. In particular, for every $n \in \mathbb{N}$, we construct a random divisible partition $\Pi_n \in \mathcal{P}_{[nk] \mid k}$ according to the following seating rule. As usual, the tables in the restaurant correspond to the blocks of a random partition.

Divisible random seating rule

- (1) The first k individuals are seated at the same table, $\Pi_1 := \{\{1, \dots, k\}\}$.
- (2) After nk individuals are seated according to Π_n , the next k individuals $nk + 1, \dots, nk + k$ are seated randomly as follows. We initialize by putting $\Pi_n^{(1)} = \Pi_n$.
 - (a) Independently for each $i = 2, \dots, k$, $nk + i$ chooses $u^{(i)}$ uniformly among $[nk + i - 1]$ and immediately displaces $u^{(i)}$ in $\Pi_n^{(i-1)}$ to define $\Pi_n^{(i)}$. If the chosen element $u^{(i)}$ is not in $\Pi_n^{(i-1)}$, then no displacement occurs.
 - (b) After each individual has made its choice of $u^{(i)}$ above, there are k in-

individuals $(nk + 1, w^{(2)}, \dots, w^{(k)})$ waiting to be seated. The group $w^* := \{nk + 1, w^{(2)}, \dots, w^{(k)}\}$ is treated as a single unit and randomly chooses a table $b \in \Pi_n^{(k)} \cup \{\emptyset\}$ according to $\text{CRP}(nk, \alpha, \theta)$, that is,

$$P\{w^* \mapsto b \mid \Pi_n^{(k)} = \pi\} \propto \begin{cases} \#b - \alpha, & b \in \pi \\ \alpha \# \pi + \theta, & b = \emptyset. \end{cases}$$

We define Π_{n+1} as the partition resulting from this seating assignment.

This construction generates a collection (Π_1, Π_2, \dots) of divisible partitions of $[nk]$.

Remark 1. Note the change in notation in (b) above from $u^{(i)}$, denoting displaced individuals, to $w^{(i)}$, denoting those individuals still not seated after the final displacement. This reflects the possibility that a single element can be displaced multiple times, because a particular element can be chosen as $u^{(i)}$ for multiple $i = 2, \dots, k$. The following example illustrates the seating procedure.

Example 3.1. With $n = k = 3$ and fixed (α, θ) , we generate the partition $\Pi_3 = 134689/257$ from the above seating procedure as follows.

- We begin with $\Pi_1 = 123$;
- At time 2, individuals 4, 5, 6 arrive and, following Step (2a), element 5 first chooses $u^{(2)}$ from $\{1, 2, 3, 4\}$, say $u^{(2)} = 4$, and element 6 chooses $u^{(3)}$ from $\{1, 2, 3, 4, 5\}$, say $u^{(3)} = 2$. Then, after Step 2, we have partition 136, with 2, 4, 5 displaced. (Note that the displaced individuals differ from the set $\{4, u^{(2)}, u^{(3)}\} = \{2, 4\}$.) Treating $w^* := \{2, 4, 5\}$ as a single unit, we choose to put $\{2, 4, 5\}$ in the same block as $\{1, 3, 6\}$ with probability $(3 - \alpha)/(3 + \theta)$ to obtain $\Pi_2 = 123456$.
- At time 3, individuals 7, 8, 9 arrive and choose $u^{(2)} = 2$ and $u^{(3)} = 5$ so that the partition after Step (2a) is 134689. We now place $w^* = \{2, 5, 7\}$ in its own block with probability $(\alpha + \theta)/(6 + \theta)$ to get $\Pi_3 = 134689/257$.

Note that there is more than one way to obtain $\Pi_3 = 134689/257$. We derive the distribution of Π_n in Theorem 1 below.

In proving the following theorem, and throughout the paper, we write $i \sim_\pi j$ to denote that elements i and j are in the same block of a partition π . Exchangeability

on $\mathcal{P}_{[nk]|k}$ is defined in the usual way: $\varepsilon_{\alpha,\theta}^{(n)}$ is exchangeable if $\varepsilon_{\alpha,\theta}^{(n)}(\pi)$ depends only on $(\#b, b \in \pi)$, for every $\pi \in \mathcal{P}_{[nk]|k}$.

Theorem 1. *Let $\Pi := (\Pi_1, \Pi_2, \dots)$ be a sequence of random partitions generated by the above seating rule. Then each Π_n is marginally an exchangeable k -divisible partition of $[nk]$ with distribution*

$$\varepsilon_{\alpha,\theta}^{(n)}(\pi) = \frac{n!}{(nk)!} \frac{(\theta/\alpha)^{\uparrow\#\pi}}{(\theta/k)^{\uparrow n}} \prod_{b \in \pi} -(-\alpha/k)^{\uparrow(\#b/k)} \frac{\#b!}{(\#b/k)!}, \quad \pi \in \mathcal{P}_{[nk]|k}. \quad (3)$$

Proof. To establish (3), we fix $k \in \mathbb{N}$ and (α, θ) in the parameter space of the Ewens–Pitman model. We show (3) by induction on n . Clearly, (3) holds for $n = 1$ since $\varepsilon_{\alpha,\theta}^{(1)}$ in (3) is the point mass at the single-block partition $\{1, 2, \dots, k\}$.

Now, assume that (3) holds for $n \in \mathbb{N}$ and consider $\pi^* \in \mathcal{P}_{[(n+1)k]|k}$. Let $A_{\pi^*} := \{\pi \in \mathcal{P}_{[nk]|k} : \pi \mapsto \pi^*\}$ be divisible partitions of $[nk]$ for which there is a positive probability of the event $\{(\Pi_n, \Pi_{n+1}) = (\pi, \pi^*)\}$ in the divisible seating process. By definition, the block sizes of each $\pi \in A_{\pi^*}$ and π^* are identical except for the block to which the displaced group $(nk + 1, w^{(2)}, \dots, w^{(k)})$ is inserted in π during step (2b). Let this block be written as $b^* \in \pi^*$.

Each random displacement in step (2a) has probability $1/[(nk + 1) \cdots (nk + k - 1)]$ and the random table assignment in step (2b) follows $\text{CRP}(nk, \alpha, \theta)$, which assigns probability $(\#b - \alpha)/(nk + \theta)$ to b if $b \neq \emptyset$ and $(\theta + \alpha\#\pi)/(nk + \theta)$ if $b = \emptyset$. By the induction hypothesis, each $\pi \in A_{\pi^*}$ is distributed as in (3), so that the joint probability of the event $(\Pi_n, \Pi_{n+1}) = (\pi, \pi^*)$, for every $\pi \in A_{\pi^*}$, is

$$\frac{n!}{(nk)!} \frac{(\theta/\alpha)^{\uparrow\#\pi}}{(\theta/k)^{\uparrow n}} \prod_{b^* \in \pi^*} -(-\alpha/k)^{\uparrow(\#b^*/k)} \prod_{b \in \pi} \frac{\#b!}{(\#b/k)!},$$

for every $n \geq 1$.

Since this joint probability is the same for all pairs (π, π^*) , we obtain the marginal distribution of Π_{n+1} by multiplying the number of partitions in A_{π^*} . Let $\overleftarrow{\pi}_n^*$ denote the labeled $(nk + 1)$ -shift of π^* . That is, define $\overleftarrow{\pi}_n^*$ as a partition of $\{2, \dots, n\}$ with

$$i \sim_{\overleftarrow{\pi}_n^*} j \quad \text{if and only if} \quad nk + i \sim_{\pi^*} nk + j,$$

and label each $b \in \overleftarrow{\pi}_n^*$ by its smallest element. In step (2a) of the random seating plan, $\overleftarrow{\pi}_n^*$ is obtained by random displacement of the elements of π . Let σ^* be the

permutation $[nk] \rightarrow [nk]$ defined by the product of transpositions

$$\sigma^* := \binom{nk+2}{u^{(2)}} \cdots \binom{nk+k}{u^{(k)}}$$

and let $\varphi_{\pi, \pi^*} : \{2, \dots, n\} \rightarrow [(n+1)k]$ be the operation corresponding to step (2a) of the above divisible seating rule: for $i = 2, \dots, k$, $\varphi_{\pi, \pi^*}(i) = \sigma^*(nk+i)$ is the element occupying position i in $(nk+1, w^{(2)}, \dots, w^{(k)})$. Let $\varphi_{\pi, \pi^*}(\overleftarrow{\pi}_n^*)$ denote the partition obtained by replacing each $i \in \{2, \dots, n\}$ with $\varphi_{\pi, \pi^*}(i)$. Writing $b \in \pi$ to denote the block to which the displaced group is added in step (2b) of the seating process, we have

$$\binom{\#b+k-1}{k-1} = \frac{(\#b+k-1)!}{(k-1)! \#b!}$$

choices of the elements of $\varphi_{\pi, \pi^*}(\overleftarrow{\pi}_n^*)$ for any choice $(u^{(2)}, \dots, u^{(k)})$ of displaced elements; and there are

$$\frac{(k-1)!}{\prod_{b' \in \overleftarrow{\pi}_n^*} \#b'!}$$

ways to arrange these elements into a labeled partition with block sizes corresponding to the block sizes of $\overleftarrow{\pi}_n^*$. Finally, the assignments under φ_{π, π^*} within each $b' \in \overleftarrow{\pi}_n^*$ can be rearranged in $\#b'!$ ways to obtain a total of $(\#b+k-1) \cdots (\#b+1)$ partitions in A_{π^*} . Equation (3) now follows by the induction hypothesis. This completes the proof.

Remark 2. Note the relationship between distributions (2) and (3). These distributions coincide when $k = 1$, because (2a) of the divisible seating rule is nugatory and the divisible random seating rule is equivalent to the usual Chinese restaurant seating rule in this case. Otherwise, these distributions differ as a result of the random shuffling that occurs during step (2a) of the divisible random seating rule.

For $\alpha = 0$, (3) with parameter $(0, \theta)$ becomes

$$\varepsilon_{0, \theta}^{(n)}(\pi) = \frac{n!}{(nk)!} \frac{\theta^{\#\pi} \prod_{b \in \pi} (\#b-1)!}{(\theta/k)^{\uparrow n}}, \quad (4)$$

the marginal distribution of each Π_n in the construction with random seating rule $\text{CRP}(nk, 0, \theta)$ in step (2b). From (4), we obtain the combinatorial identity

$$\frac{1}{(nk)!} \sum_{\pi \in \mathcal{P}_{[nk] | k}} \theta^{\#\pi} \Gamma(\pi) = \left(\frac{\theta}{k}\right)^{\uparrow n} / n!, \quad (5)$$

where $\Gamma(\pi) := \prod_{b \in \pi} (\#b - 1)!$. This identity gives the generating function for k -divisible permutations, that is, permutations of $[nk]$ whose cycle sizes are all divisible by k , and also gives the following special property of the Ewens distribution.

Corollary 3.1. *For $\theta > 0$, the distribution $\varepsilon_{0,\theta}^{(n)}$ in (3) is that of a $(0, \theta)$ -partition conditioned to be k -divisible.*

3.2. Divisible partition structures

We specify a random deletion scheme for $\mathcal{P}_{[nk] | k}$ as follows. Given $\pi \in \mathcal{P}_{[nk] | k}$, let $b_* \in \pi$ denote the block of π containing $(n-1)k + 1$.

- (i) Sequentially, for $i = k, k-1, \dots, 2$, an element $u^{(i)}$ is chosen uniformly from the set

$$(b_* \cap [(n-1)k + i]) \setminus \{(n-1)k + 1, u^{(k)}, \dots, u^{(i+1)}\}.$$

Let π^* be the image of π under permutation by the product of transpositions

$$\sigma^* := \begin{pmatrix} (n-1)k + 2 \\ u^{(2)} \end{pmatrix} \cdots \begin{pmatrix} (n-1)k + k \\ u^{(k)} \end{pmatrix}. \quad (6)$$

- (ii) Obtain $\pi' \in \mathcal{P}_{[(n-1)k] | k}$ by deleting $\{(n-1)k + 1, \dots, (n-1)k + k\}$ from π^* .

Definition 1. We call a collection $\varepsilon = (\varepsilon^{(n)}, n \in \mathbb{N})$ of probability distributions a *divisible partition structure* if, for every $n \in \mathbb{N}$, $\varepsilon^{(n)}$ is the distribution of Π' obtained by applying the divisible deletion scheme to $\Pi \sim \varepsilon^{(n+1)}$.

Theorem 2. *For any (α, θ) , $\varepsilon_{\alpha,\theta} := (\varepsilon_{\alpha,\theta}^{(n)}, n \in \mathbb{N})$ is a divisible partition structure under the above deletion scheme.*

Proof. For $\pi' \in \mathcal{P}_{[(n-1)k] | k}$, we define $A_{\pi'} := \{\pi \in \mathcal{P}_{[nk] | k} : \pi' \leftarrow \pi\}$ to be the set of divisible partitions of $[nk]$ for which there is positive probability of obtaining π' from the above deletion scheme. Any $\pi \in A_{\pi'}$ has the same block structure as π' except for the block $b_* \in \pi$ containing $(n-1)k + 1$, which has k more elements than its corresponding block in π' and will be reduced by k during the deletion process. From the proof of Theorem 1, we can express the probability of $\pi \in A_{\pi'}$ as

$$\varepsilon_{\alpha,\theta}^{(n)}(\pi) = \varepsilon_{\alpha,\theta}^{(n-1)}(\pi') \frac{(\#b_* - 1)^{\downarrow(k-1)}}{(nk - 1)^{\downarrow(k-1)}} \left[\frac{\#b_* - k - \alpha}{(n-1)k + \theta} \mathbb{I}_{\{\#b_* > k\}} + \frac{\theta + \alpha \# \pi'}{(n-1)k + \theta} \mathbb{I}_{\{\#b_* = k\}} \right]. \quad (7)$$

For any $b \in \pi'$, the probability that $(n-1)k+i$, $i = 2, \dots, k$, displaces a specific element of b_* is $1/(b_* - k + i - 1)$ and, in total, there are $nk - k + i - 1$ elements which $(n-1)k+i$ has the option of displacing. (Every $\pi \in A_{\pi'}$ corresponds to a choice $b \in \pi'$ to insert the displaced group in step (2b) of the random seating rule. Given $b \in \pi$, the choice $((n-1)k+1, u^{(2)}, \dots, u^{(k)})$ corresponds to a unique k -tuple of transpositions σ^* in (6) to obtain π' from π by the deletion process. There are $(nk - k + 1) \cdots (nk + 1)$ total choices for every such $b \in \pi$.) By the law of cases, we have

$$\begin{aligned}
\mathbb{P}\{\Pi_{n-1} = \pi'\} &= \sum_{\pi \in A_{\pi'}} \mathbb{P}\{\Pi_{n-1} = \pi' \mid \Pi_n = \pi\} \varepsilon_{\alpha, \theta}^{(n)}(\pi) \\
&= \sum_{b_*} \sum_{\sigma^*} \varepsilon_{\alpha, \theta}^{(n)}(\pi) \frac{1}{(\#b_* - 1) \cdots (\#b_* - k + 1)} \\
&= \varepsilon_{\alpha, \theta}^{(n-1)}(\pi') \sum_{b_*} \left[\frac{\#b_* - k - \alpha \mathbb{I}_{\{\#b_* > k\}}}{(n-1)k + \theta} + \frac{\theta + \alpha \# \pi'}{(n-1)k + \theta} \mathbb{I}_{\{\#b_* = k\}} \right] \\
&= \varepsilon_{\alpha, \theta}^{(n-1)}(\pi').
\end{aligned}$$

This completes the proof.

Though we do not pursue it in detail, we conclude this section with the observation that exchangeable divisible partition structures are in correspondence with Kingman's paintbox measures. In particular, the $\varepsilon_{\alpha, \theta}$ -family of measures in (3) is in correspondence with the Poisson–Dirichlet(α, θ) measures, for all $k = 1, 2, \dots$

Theorem 3. *Under the above deletion scheme, exchangeable divisible partition structures are in one-to-one correspondence with Kingman's paintbox measures.*

Proof. We need only sketch the proof since the arguments follow by Theorems 1 and 2 and Kingman's paintbox representation. Given a collection $(\Pi_n, n \geq 1)$ of exchangeable k -divisible partitions that is consistent in distribution under the above divisible deletion scheme, we can obtain a collection $(\Pi_n^*, n \geq 1)$ of exchangeable partitions of $(\mathcal{P}_{[n]}, n \geq 1)$ by “deflating” each block by a factor of k and choosing a representative element $1, \dots, n$ of each group of size k within each block. The result will be an exchangeable partition of $[n]$, which must obey one of Kingman's paintbox distributions. The rest now follows by analogous argument to previous theorems.

3.3. Divisible Markov structures

For $\alpha > 0$ and $m \in \mathbb{N}$, the Ewens–Pitman($-\alpha, m\alpha$) distribution determines a probability measure on the subspace $\mathcal{P}_{\mathbb{N}}^{(m)}$ of partitions of \mathbb{N} having at most m blocks. Previously, Crane [4] introduced an exchangeable family of Markov chains on $\mathcal{P}_{\mathbb{N}}^{(m)}$ with marginal transition probabilities

$$p_{\alpha, m}^{(n)}(\pi, \pi') = m^{\downarrow \#\pi'} \prod_{b \in \pi} \frac{\prod_{b' \in \pi'} (\alpha/m)^{\uparrow \#(b \cap b')}}{\alpha^{\uparrow \#b}}, \quad \pi, \pi' \in \mathcal{P}_{[n]}^{(m)}, \quad (8)$$

where $m^{\downarrow n} := m(m-1)\cdots(m-n+1)$. More recently, structural properties of exchangeable Feller processes on $\mathcal{P}_{\mathbb{N}}^{(m)}$ have been characterized in full [8]. The transition probabilities in (8) are reversible with respect to $\varrho_{-\alpha, m\alpha}^{(n)}$ for every $n \in \mathbb{N}$. We now extend this family to divisible partitions.

For $\alpha > 0$ and $m \in \mathbb{N}$, the distribution (3) with parameter $(-\alpha k, m\alpha k)$ is

$$\varepsilon_{-\alpha k, m\alpha k}^{(n)}(\pi) = m^{\downarrow \#\pi} \frac{n!}{(nk)!} \frac{\prod_{b \in \pi} \alpha^{\uparrow \#b/k} \frac{\#b!}{(\#b/k)!}}{(m\alpha)^{\uparrow n}}. \quad (9)$$

Let $p_{\alpha, m}^{(n)}$ denote the transition probabilities in (8), and let $\mathcal{P}_{[nk]_k}^{(m)}$ be the subset of k -divisible partitions of $[nk]$ with at most m blocks. We describe a Markovian transition procedure on $\mathcal{P}_{[nk]_k}^{(m)}$ as follows. Fix $\pi \in \mathcal{P}_{[nk]_k}^{(m)}$.

- (i) Independently, for each $b \in \pi$, randomly partition b into $\#b/k$ groups of size k according to the uniform distribution on such partitions of b . Label the groups uniquely in $[n]$ to obtain a collection of groups $\{g_1, \dots, g_n\}$.
- (ii) Given $\{g_1, \dots, g_n\}$ from (i), let π^* denote the partition of $\{g_1, \dots, g_n\}$ obtained by regarding each group as a single element in π^* , and generate $\Pi'' \sim p_{\alpha, m}^{(n)}(\pi^*, \cdot)$, as in (8).
- (iii) Given $\Pi'' = \pi''$, obtain the next state $\pi' \in \mathcal{P}_{[nk]_k}^{(m)}$ by replacing each g_i in π'' with the group of k elements it represents from (i).

Example 3.2. To illustrate the above transition procedure, let $n = 3$, $k = 2$, and $\pi = 1246/35$. We generate the transition $\pi \mapsto \pi'$ as follows.

- (i) We randomly partition the blocks of π into sub-blocks of size k and assign labels 1, 2, 3, e.g., $g_1 = 14$, $g_2 = 26$, and $g_3 = 35$;

- (ii) The above procedure yields a partition $\pi^* = 12/3$, from which we generate Π'' according to $p_{\alpha, m}^{(n)}(\pi^*, \cdot)$, say $\pi'' = 13/2$.
- (iii) We obtain π' by substituting g_i for i in π'' , i.e., $\pi' = g_1 g_3 / g_2 = 1345/26$.

As in Example 3.1, there is more than one way to generate the transition $\pi \mapsto \pi'$. We derive the transition probability in Theorem 4.

In the following theorem, we write $\pi \wedge \pi'$ to denote the usual *meet* of π and π' , that is,

$$\pi \wedge \pi' := \{B_i \cap B'_j : B_i \in \pi, B'_j \in \pi'\} \setminus \{\emptyset\}.$$

Theorem 4. *The finite-dimensional transition probabilities of the transition procedure in (i)-(iii) are*

$$\mathcal{E}_{\alpha, m}^{(n)}(\pi, \pi') = m^{\downarrow \# \pi'} \prod_{b \in \pi} \left[\frac{(\#b/k)!}{\#b!} \frac{1}{\alpha^{\uparrow \#b/k}} \prod_{b' \in \pi'} \frac{\#(b \cap b')!}{\left[\frac{\#(b \cap b')}{k}\right]!} (\alpha/m)^{\uparrow \#(b \cap b')/k} \right], \quad (10)$$

for $\pi, \pi' \in \mathcal{P}_{[nk]|k}^{(m)}$ satisfying $\pi \wedge \pi' \in \mathcal{P}_{[nk]|k}$. Moreover, for each $n \in \mathbb{N}$, $\mathcal{E}_{\alpha, m}^{(n)}$ is reversible with respect to $\varepsilon_{-\alpha k, m \alpha k}^{(n)}$ in (3) and is exchangeable with respect to the symmetric group on $[nk]$.

Proof. Fix $n \in \mathbb{N}$ and let $\pi, \pi' \in \mathcal{P}_{[nk]|k}^{(m)}$ satisfy $\pi \wedge \pi' \in \mathcal{P}_{[nk]|k}$. According to the transition procedure, we first group elements within each $b \in \pi$ together in groups of size k . There are

$$\frac{\#b!}{(k!)^{\#b/k}}$$

ways to group elements and subsequently label them uniquely, each equally likely.

Given a block $b \in \pi$ and the new partition π' , there are

$$\frac{(\#b/k)!}{(k!)^{\#b/k}} \prod_{b' \in \pi'} \frac{\#(b \cap b')!}{\left[\frac{\#(b \cap b')}{k}\right]!}$$

sets of labeled groups of b for which a transition $\pi \mapsto \pi'$ is permissible. (For each $b' \in \pi'$, there are

$$\frac{\#(b \cap b')!}{(k!)^{\#(b \cap b')/k}}$$

labeled partitions of the elements of $b \cap b'$ into blocks of size k . Dividing this number by $\left[\frac{\#(b \cap b')}{k}\right]!$ gives the number of unlabeled partitions of $b \cap b'$ into $\#(b \cap b')/k$ blocks

of size k . Given a partition of b , there are $(\#b/k)!$ ways to label the blocks.) Each permissible partition of b has probability

$$\frac{(k!)^{\#b/k}}{\#b!}.$$

Multiplication of the number of groupings by their probabilities gives a total factor of

$$\prod_{b \in \pi} \frac{(\#b/k)!}{\#b!} \prod_{b' \in \pi'} \frac{\#(b \cap b')!}{\left[\frac{\#(b \cap b')}{k}\right]!}. \quad (11)$$

Given a partition π^* of n groups of size k , we choose π^{**} from the transition probabilities of (8),

$$m^{\downarrow \# \pi'} \prod_{b \in \pi} \frac{\prod_{b' \in \pi'} (\alpha/m)^{\uparrow(\#(b \cap b')/k)}}{\alpha^{\uparrow \#b/k}}. \quad (12)$$

Multiplying (11) and (12) gives (10). Reversibility is clear by checking the detailed balanced condition, and exchangeability is clear by inspection. This completes the proof.

3.4. Divisible Markovian deletion

In the following deletion scheme, we define $\sigma(\pi)$ as the image of $\pi \in \mathcal{P}_{[n]}$ by a permutation $\sigma : [n] \rightarrow [n]$, that is, i and j are in the same block of $\sigma(\pi)$ if and only if $\sigma^{-1}(i)$ and $\sigma^{-1}(j)$ are in the same block of π .

For $n \in \mathbb{N}$, let $\Pi^{n+1} = (\Pi_1^{n+1}, \Pi_2^{n+1}, \dots)$ be a Markov chain on $\mathcal{P}_{[(n+1)k] | k}^{(m)}$. Given $\Pi^{n+1} = (\pi_j^{n+1}, j \geq 1)$, we obtain the sequence $\Pi^n = (\pi_j^n, j \geq 1)$ in $\mathcal{P}_{[nk] | k}^{(m)}$ as follows.

- (i) Obtain π_1^n from π_1^{n+1} by the divisible deletion scheme in Section 3.2. Let σ_1 be the permutation, called the *displacement*, generated in (6).
- (ii) For $j \geq 1$, given $(\Pi_1^{n+1}, \dots, \Pi_{j+1}^{n+1}) = (\pi_1^{n+1}, \dots, \pi_{j+1}^{n+1})$, $(\Pi_1^n, \dots, \Pi_j^n) = (\pi_1^n, \dots, \pi_j^n)$, and displacements $\sigma_1, \dots, \sigma_j$, we put $\sigma^{(j)} = \sigma_j \circ \dots \circ \sigma_1$, denote $\pi^* = \sigma^{(j)}(\pi_j^{n+1})$ and $\pi' = \sigma^{(j)}(\pi_{j+1}^{n+1})$, and let $b_* \in \pi^*$, $b'_* \in \pi'$ be the blocks containing element $nk + 1$.

- (a) Sequentially, for $i = k, \dots, 2$, choose $u^{(i)}$ uniformly from

$$b_* \cap b'_* \cap ([nk - k + i] \setminus \{nk - k + 1, u^{(k)}, \dots, u^{(i+1)}\})$$

and put

$$\sigma_{j+1} = \binom{nk - k + 2}{u^{(2)}} \cdots \binom{nk}{u^{(k)}}. \quad (13)$$

(b) Let $\pi'' = \sigma_{j+1}(\pi')$ and obtain π_{j+1}^n from π'' by deleting $\{nk+1, \dots, nk+k\}$.

Remark 3. The relabeling in (ii) by composing the displacements ensures that the elements $1, 2, \dots$ are consistently labeled in the restricted sequence. This is needed because of the first step of the transition procedure, whereby individuals are randomly grouped into sub-blocks of size k .

Remark 4. For each $n = 1, 2, \dots$, the Markovian deletion operation $\Pi^{n+1} \rightarrow \Pi^n$ is more than an independent application of the divisible deletion scheme from Section 3.2 at each time. For each $t = 1, 2, \dots$, (Π_t^{n+1}, Π_t^n) is marginally distributed as a k -divisible partition structure, but Step (ii) of the Markovian deletion scheme incorporates dependence among the deletions across time.

Definition 2. We call a collection $\Pi = (\Pi^1, \Pi^2, \dots)$ of Markov chains an $(\varepsilon, \mathcal{E})$ -reversible Markov structure if, for each $n \in \mathbb{N}$, Π^n is an $(\varepsilon^{(n)}, \mathcal{E}^{(n)})$ -Markov chain that is reversible with respect to $\varepsilon^{(n)}$, and Π' obtained by applying the above deletion scheme to Π^{n+1} is an $(\varepsilon^{(n)}, \mathcal{E}^{(n)})$ -Markov chain.

Theorem 5. For each $n \in \mathbb{N}$, let Π^n be an $(\varepsilon_{-\alpha k, m\alpha k}^{(n)}, \mathcal{E}_{\alpha, m}^{(n)})$ -Markov chain on $\mathcal{P}_{[nk]}^{(m)}|_k$. Then $\Pi = (\Pi^n, n \in \mathbb{N})$ is an $(\varepsilon_{-\alpha k, m\alpha k}, \mathcal{E}_{\alpha, m})$ -reversible Markov structure.

Proof. Fix $n \in \mathbb{N}$ and let $\Pi^{n+1} = (\Pi_j^{n+1}, j \geq 1)$ be an $(\varepsilon_{-\alpha k, m\alpha k}^{(n+1)}, \mathcal{E}_{\alpha, m}^{(n+1)})$ -Markov chain. Let $\Pi^n = (\Pi_j^n, j \geq 1)$ be obtained from Π^{n+1} by the above deletion scheme. By Theorem 2 and step (i) of the deletion process, $\Pi_1^n \sim \varepsilon_{-\alpha k, m\alpha k}^{(n)}$. We induct on j to show that $\Pi^n = (\Pi_j^n, j \geq 1)$ is an $(\varepsilon_{-\alpha k, m\alpha k}^{(n)}, \mathcal{E}_{\alpha, m}^{(n)})$ -Markov chain.

For $j \geq 1$, assume the marginal law of Π_j^n is $\varepsilon_{-\alpha k, m\alpha k}^{(n)}$ and let $\sigma_1, \dots, \sigma_j$ be the displacement permutations used in the Markovian deletion scheme up to step j . By Theorem 1, $\sigma^{(j)}(\Pi_j^{n+1}) \sim \varepsilon_{-\alpha k, m\alpha k}^{(n)}$ and, by Theorem 4, the conditional distribution of $\sigma^{(j)}(\Pi_{j+1}^{n+1})$ given $\sigma^{(j)}(\Pi_j^{n+1})$ is $\mathcal{E}_{\alpha, m}^{(n+1)}(\sigma^{(j)}(\Pi_j^{n+1}), \cdot)$. Given $(\sigma^{(j)}(\Pi_j^{n+1}), \sigma^{(j)}(\Pi_{j+1}^{n+1})) = (\pi, \pi')$ and $\Pi_j^n = \pi^*$, let $\pi'' \in \mathcal{P}_{[nk]}^{(m)}|_k$ be such that $\mathcal{E}_{\alpha, m}^{(n)}(\pi^*, \pi'') > 0$ and there is a positive probability of obtaining π'' from π' in the divisible Markovian deletion procedure. Let $b_* \in \pi$, $b'_* \in \pi'$ be the blocks containing $nk+1$. We have

$$\frac{\mathcal{E}_{\alpha,m}^{(n+1)}(\pi, \pi')}{\mathcal{E}_{\alpha,m}^{(n)}(\pi^*, \pi'')} = \frac{\#b_*/k}{(\#b_*)^{\downarrow k}} \frac{1}{\alpha + \#b_*/k - 1} \left[\left(\frac{\alpha}{m} + \frac{\#(b_* \cap b'_*)}{k} - 1 \right) \frac{(\#(b_* \cap b'_*))^{\downarrow k}}{\left[\frac{\#(b_* \cap b'_*)}{k} \right]} \mathbb{I}_{\{\#b'_* > k\}} + k! \frac{(m - \#\pi')\alpha}{m} \mathbb{I}_{\{\#b'_* = k\}} \right] \quad (4)$$

In step (ii)(a), σ_{j+1} in (13) has probability $(\#(b_* \cap b'_*)) / ((\#(b_* \cap b'_*))^{\downarrow k})$.

Now, for every π' with $\mathcal{E}_{\alpha,m}^{(n)}(\pi, \pi') > 0$, the block sizes of π' are the same as π'' except for the block $b'_* \in \pi'$, which contains k more elements than its corresponding block in π'' . Given b'_* , there are $(\#b_*)^{\downarrow k} / \#b_*$ partitions π'' from which π' can be obtained by the even Markovian deletion process. Each π'' corresponds to a unique displacement σ_{j+1} in (13). By the law of cases, we have

$$\begin{aligned} \mathbb{P}\{\Pi_{j+1}^n = \pi'' \mid \Pi_j^n = \pi^*, \Pi_j^{n+1} = \pi\} &= \\ &= \sum_{\pi'} \mathbb{P}\{\Pi_{j+1}^n = \pi'' \mid \Pi_j^n = \pi^*, (\Pi_j^{n+1}, \Pi_{j+1}^{n+1}) = (\pi, \pi')\} \mathcal{E}_{\alpha,m}^{(n+1)}(\pi, \pi') \\ &= \mathcal{E}_{\alpha,m}^{(n)}(\pi^*, \pi'') \sum_{b'_*} \sum_{\sigma_{j+1}} \frac{\#b_*}{(\#b_*)^{\downarrow k}} \frac{1}{\alpha + \#b_*/k - 1} \left[\left(\frac{\alpha}{m} + \frac{\#(b_* \cap b'_*)}{k} - 1 \right) \mathbb{I}_{\{\#b'_* > k\}} + \frac{(m - \#\pi')\alpha}{m} \mathbb{I}_{\{\#b'_* = k\}} \right] \\ &= \mathcal{E}_{\alpha,m}^{(n)}(\pi^*, \pi'') \sum_{b'_*} \frac{1}{\alpha + \#b_*/k - 1} \left[\left(\frac{\alpha}{m} + \frac{\#(b_* \cap b'_*)}{k} - 1 \right) \mathbb{I}_{\{\#b'_* > k\}} + \frac{(m - \#\pi')\alpha}{m} \mathbb{I}_{\{\#b'_* = k\}} \right] \\ &= \mathcal{E}_{\alpha,m}^{(n)}(\pi^*, \pi''). \end{aligned}$$

By the induction hypothesis and reversibility (Theorem 4), the unconditional law of Π_{j+1}^n is $\varepsilon_{-\alpha k, m \alpha k}^{(n)}$ and Π^n is an $(\varepsilon_{-\alpha k, m \alpha k}^{(n)}, \mathcal{E}_{\alpha,m}^{(n)})$ -Markov chain. This completes the proof.

4. Concluding remarks

4.1. Divisible paintbox partitions

We have treated the special case of Kingman's paintbox process directed by $\nu = \text{Poisson-Dirichlet}(\alpha, \theta)$; however, we can apply the above descriptions more generally to generate both divisible partition structures and divisible Markov structures associated to any probability measure on the ranked-simplex. In this case, we replace the Ewens-Pitman measure by a paintbox measure ϱ_ν in the above statements, but reversibility

of the resulting family of Markov structures is not guaranteed. The Markov partition structures for the $k = 1$ case have been studied in [4]; the $k > 1$ case follows by modifying the work in [4] according to the program in Section 3.3.

4.2. Divisible random permutations

There is a well-known connection between exchangeable random partitions and random permutations whose distribution depends only on their cycle sizes. The combinatorial identity (5) relates divisible permutations and divisible partitions in the usual way: each partition of $[nk]$ with blocks of size (n_1, \dots, n_m) corresponds to $\prod_{j=1}^m (n_j - 1)!$ permutations of $[nk]$ with the corresponding cycle sizes. The usual approach of sampling uniformly from the subset of permutations corresponding to a given random partition yields a distribution on what we call *k-divisible permutations*. Divisible permutations appear in combinatorics, e.g., Wilf [21], and the generating function (5) for *k-divisible permutations* generates known integer sequences, e.g., [20]:A001818 and [20]:A178575.

4.3. Neutral partition structures

We can also consider partitions of a marked population $[n]^*$ consisting of nk individuals, with n individuals of each type $j = 1, \dots, k$. We write $1^{(j)}, \dots, n^{(j)}$ to denote the individuals of type $j = 1, \dots, k$. We call a subset of $[n]^*$ *neutral* if it contains an equal number of elements of each type. A partition of $[n]^*$ is neutral if each of its blocks is. Necessarily, any neutral partition is *k-divisible*.

As in the *k-divisible* case, $k = 1$ corresponds to ordinary set partitions; however, by adding more structure to the partitions, the combinatorial arguments for neutral partitions are more straightforward. The only difference from the *k-divisible* case is that our seating rule is specified to preserve neutrality, not just divisibility. For example, the Chinese restaurant process in the neutral case corresponds to the following seating rule.

Neutral random seating rule

- (1) The first k individuals are seated at the same table, $\Pi_1 := \{1^{(1)}, \dots, 1^{(k)}\}$.
- (2) After nk arrivals are seated according to Π_n , the next k individuals $(n+1)^{(1)}, \dots, (n+$

$1)^{(k)}$ are seated randomly as follows:

- (a.) independently for each $i = 2, \dots, k$, $(n+1)^{(i)}$ chooses $u^{(i)}$ uniformly among $1^{(i)}, \dots, (n+1)^{(i)}$ and displaces $u^{(i)}$ in Π_n ;
- (b.) the displaced group $((n+1)^{(1)}, u^{(2)}, \dots, u^{(k)})$ is treated as a single unit and randomly sits at table $b \in \Pi_n \cup \{\emptyset\}$ according to $\text{CRP}(nk, \alpha, \theta)$.

The finite-dimensional distributions on neutral partitions of $[n]^*$ generated by the above seating rule are

$$\nu_{\alpha, \theta}^{(n)}(\pi) = \frac{(\theta/\alpha)^{\uparrow \# \pi} \prod_{b \in \pi} -(-\alpha/k)^{\uparrow \# b/k} [(\# b/k)!]^{k-1}}{(\theta/k)^{\uparrow n} (n!)^{k-1}}. \quad (15)$$

The distribution in (15) is exchangeable with respect to permutations of $[n]^*$ that permute only elements with the same type. We obtain the deletion rule by reversing the Chinese restaurant seating rule, as we have for the k -divisible deletion rule.

Neutral Markov structures are collections of pairs $\{(\nu^{(n)}, \mathcal{N}^{(n)})\}_{n \geq 1}$ so that $\{\nu^{(n)}\}_{n \geq 1}$ is an exchangeable neutral partition structure and $\{\mathcal{N}^{(n)}\}_{n \geq 1}$ is a collection of exchangeable Markovian transition probabilities consistent under an operation of neutral Markovian deletion, which is analogous to the divisible Markovian deletion scheme. In the neutral setting, we obtain finite-dimensional transition probabilities corresponding to the Ewens–Pitman Markov chain:

$$\mathcal{N}_{\alpha, m}^{(n)}(\pi, \pi') = m^{\downarrow \# \pi'} \prod_{b \in \pi} \left[\frac{\prod_{b' \in \pi'} (\alpha/m)^{\uparrow (\#(b \cap b')/k)} [(\#(b \cap b')/k)!]^{k-1}}{[(\# b/k)!]^{k-1} \alpha^{\uparrow \# b/k}} \right], \quad (16)$$

provided $\pi \wedge \pi'$ is a neutral partition. The transition probability in (16) is reversible with respect to $\nu_{-\alpha k, m \alpha k}^{(n)}(\cdot)$ and is exchangeable with respect to permutations of $[n]^*$ that preserve neutrality. Connections between exchangeable neutral partition structures and Kingman's paintbox measures are analogous to those for divisible partition structures and follow by similar arguments.

Acknowledgements

Harry Crane is partially supported by U.S. NSF DMS-1308899 and U.S. NSA H98230-13-1-0299 grants.

References

- [1] R. A. Bailey. *Association Schemes: Designed Experiments, Algebra and Combinatorics*, volume 84 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004.
- [2] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [3] J. Bertoin. Two-parameter Poisson-Dirichlet measures and reversible exchangeable fragmentation-coalescence processes. *Combin. Probab. Comput.*, 17(3):329–337, 2008.
- [4] H. Crane. A consistent Markov partition process generated from the paintbox process. *J. Appl. Probab.*, 43(3):778–791, 2011.
- [5] H. Crane. Clustering from categorical data sequences. *Journal of the American Statistical Association*, in press, 2014.
- [6] H. Crane. Permanent Partition Models and Markovian Gibbs Structures. *Journal of Statistical Physics*, 153(4):698–726, 2013.
- [7] H. Crane. Some algebraic identities for α -permanent. *Linear Algebra and Its Applications*, 439(11):3445–3459, 2013.
- [8] H. Crane. The cut-and-paste process. *Annals of Probability*, 42(5):1952–1979, 2014.
- [9] P. Donnelly and G. Grimmett. On the asymptotic distribution of large prime factors. *J. London Math. Soc.*, 47:395–404, 1993.
- [10] B. Efron and R. Thisted. Estimating the number of unseen species: How many words did shakespeare know? *Biometrika*, 63:435–447, 1976.
- [11] B. Efron and R. Thisted. Did shakespeare write a newly discovered poem? *Biometrika*, 74:445–455, 1987.

- [12] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3:87–112, 1972.
- [13] R. Fisher, A. Corbet, and C. Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Ecology*, 12:42–58, 1943.
- [14] A. Gnedin, C. Haulk, and J. Pitman. Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In *Probability and mathematical genetics*, volume 378 of *London Math. Soc. Lecture Note Ser.*, pages 264–298. Cambridge Univ. Press, Cambridge, 2010.
- [15] J. F. C. Kingman. Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A*, 361(1704):1–20, 1978.
- [16] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380, 1978.
- [17] P. McCullagh and J. Yang. How many clusters? *Bayesian Anal.*, 3(1):101–120, 2008.
- [18] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [19] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.
- [20] N. Sloane. Online Encyclopedia of Integer Sequences. *Published electronically at <http://www.oeis.org/>*.
- [21] H. S. Wilf. *generatingfunctionology, Third Edition*. AK Peters/CRC Press, 2005.