Combinatorics, Probability and Computing (2007) 16, 819–827. © 2007 Cambridge University Press
doi:10.1017/S0963548306008352 First published online 4 January 2007 Printed in the United Kingdom

Poisson Representation of a Ewens Fragmentation Process

ALEXANDER GNEDIN[1] and JIM PITMAN[2†]

[1]Mathematical Institute, Utrecht University, The Netherlands
(e-mail: gnedin@math.uu.nl)
[2]Department of Statistics, University of California, Berkeley, USA
(e-mail: pitman@stat.Berkeley.EDU)

Received 14 August 2006; revised 11 October 2006

A simple explicit construction is provided of a partition-valued fragmentation process whose distribution on partitions of $[n] = \{1, \dots, n\}$ at time $\theta \geqslant 0$ is governed by the Ewens sampling formula with parameter θ. These partition-valued processes are exchangeable and consistent, as n varies. They can be derived by uniform sampling from a corresponding mass fragmentation process defined by cutting a unit interval at the points of a Poisson process with intensity $\theta x^{-1} dx$ on $\mathbb{R}_+$, arranged to be intensifying as θ increases.

1. Introduction

There has been much recent interest in models for random processes of fragmentation and coagulation: see Chapter 5 of [14] and the recent book [3]. Mekjian and others [4, 11, 12, 13] have considered Ewens partitions with parameter θ as a model for fragmentation phenomena, with the intuitive notion that increasing θ corresponds to further fragmentation. But it does not seem obvious how to construct a nice Markovian fragmentation process corresponding to this idea.

It was pointed out in [14] that it is possible to construct a sequence of partition-valued processes $(\Pi_{n,\theta},\ \theta \geqslant 0)$ $(n = 1, 2, \dots)$ with the following properties.

Ewens distribution. $\Pi_{n,\theta}$ is for each $n = 1, 2, \dots$ and $\theta \geqslant 0$ a random partition of the set $[n] := \{1, 2, \dots, n\}$, with distribution determined by the following formula: for each composition $(n_1, \dots, n_k)$ of $n = \sum_{i=1}^{k} n_i$, and each partition π of $[n]$ into k blocks of sizes $n_1, \dots, n_k$,

$$\mathbb{P}(\Pi_{n,\theta} = \pi) = \frac{\theta^{k-1}}{[\theta + 1]_{n-1}} \prod_{i=1}^{k} (n_i - 1)! \tag{1.1}$$

where $[x]_m := x(x+1)\cdots(x+m-1)$ is the Pochhammer factorial.

† Research supported in part by NSF grant DMS-0405779.

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

Fragmentation. $\Pi_{n,\theta}$ is a refinement of $\Pi_{n,\phi}$ if $\theta > \phi$, for all $n \geqslant 1$, that is, each block of $\Pi_{n,\phi}$ is some union of blocks of $\Pi_{n,\theta}$.

Consistency. For each $n < m$, a process with the same distribution as $(\Pi_{n,\theta},\ \theta \geqslant 0)$ is obtained by restriction of $(\Pi_{m,\theta},\ \theta \geqslant 0)$ to $[n]$.

Exchangeability. For each n, the law of $(\Pi_{n,\theta},\ \theta \geqslant 0)$ is invariant under permutations of the set $[n]$.

We call a sequence of partition-valued processes $(\Pi_{n,\theta},\ \theta \geqslant 0)$ $(n = 1, 2, \ldots)$ with these properties a *family of Ewens fragmentations*. One family of Ewens fragmentations associated with Kingman's coalescent was analysed in [2].

Some general theory [5] implies that any such family of processes can be defined in the strong sense consistently (*i.e.*, so that $\Pi_{m,\theta}|_{[n]} = \Pi_{n,\theta}$ for $m > n$) on a single probability space by means of uniform sampling of points engaged in a process of fragmentation of a total mass 1 into a countable collection of submasses with sum 1, with more and more refined splitting of the submasses as the time parameter θ increases. See [3, Chapter 3] and [14, Chapter 5] for further background and references.

The problem was posed in [14] of characterizing the dynamics of a family of Ewens fragmentations, preferably in a Markovian way. For applications, it is desirable to have a model which can easily be simulated for modest values of n. But previous efforts fall short in this respect. In this note we partly solve this problem by constructing a new family of Ewens fragmentations. Our family is not Markovian, but it enjoys the Markov property and follows a very simple transition rule when viewed as a fragmentation process in the extended space of *ordered* partitions. This simplification by passing to an ordered structure extends our previous work on regenerative partitions and their relatives [7, 8, 9].

2. Construction

Recall that a *composition* of n is a sequence of positive integers $(n_1, \ldots, n_k)$ with sum n. We regard a composition of n as a way of distributing n unlabelled balls in an ordered sequence of k non-empty boxes, with n_i balls in the ith box. A composition of n is also conveniently encoded by the binary sequence of 0s and 1s obtained by concatenating subsequences of the form $1, 10, 100, \ldots$, where the ith subsequence in the concatenation has length n_i. So the symbols 1 occur at places $1, n_1 + 1, n_1 + n_2 + 1, \ldots \sum_{i=1}^{k-1} n_i + 1$. Using a particular composition $(3, 4, 1)$ of 8 for illustration, the balls-in-boxes picture is suggested by the notation [000] [0000] [0]. The binary representation is obtained by replacing each [0 by a 1 and deleting each] to obtain the sequence 10010001. Let x and y be two compositions of n, each represented as a binary sequence, say $x = (x_i, 1 \leqslant i \leqslant n)$ and $y = (y_i, 1 \leqslant i \leqslant n)$. Say x is a *refinement* of y if $x_i \geqslant y_i$ for every i. In terms of the balls-in-boxes picture, x is derived from y by splitting boxes into sub-boxes, and in terms of the binary representation x is derived by switching some 0s to 1s. For instance, 11010101 is a refinement of 10010001.

Given a stochastic process $(C_{n,\theta},\ \theta \geqslant 0)$, with values in compositions of n, we define an *associated partition-valued process* $(\Pi_{n,\theta}, \theta \geqslant 0)$ by first assigning each of the n places for a ball in the balls-in-boxes representation a number in $[n]$ according to a uniform random

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

permutation of $[n]$ independent of $(C_{n,\theta},\ \theta \geqslant 0)$, thus obtaining an *ordered partition* of $[n]$, then ignoring the order of the boxes to obtain a partition of $[n]$. If $(C_{n,\theta},\ \theta \geqslant 0)$ is refining as θ increases then the associated partition-valued process $(\Pi_{n,\theta},\ \theta \geqslant 0)$ is a fragmentation process whose law is invariant under permutations of $[n]$. The process $(\Pi_{n,\theta},\ \theta \geqslant 0)$ then describes a process of randomly splitting up a collection of balls labelled by $[n]$ into an unordered collection of boxes.

Theorem 2.1. *Let Θ_j for $j = 1, 2, \ldots$ be a sequence of independent random variables with distributions*

$$\mathbb{P}(\Theta_j \leqslant \theta) = \theta/(\theta + j - 1), \quad \theta \geqslant 0, \tag{2.1}$$

where $0/0 = 1$. Let $C_{n,\theta}$ for $n = 1, 2, \ldots$ be the random composition of n whose binary representation is the sequence of indicator variables $1(\Theta_j \leqslant \theta)$ for $1 \leqslant j \leqslant n$. Then the sequence $(\Pi_{n,\theta},\ \theta \geqslant 0)$ of partition-valued processes associated with $(C_{n,\theta},\ \theta \geqslant 0)$ defines a family of Ewens fragmentations.

That $\Pi_{n,\theta}$ has the Ewens distribution (1.1) can be read from the known result [1, 9, 14] that in the binary representation of the composition of n derived from the block sizes of a Ewens partition of $[n]$ in reversed size-biased order, the digits are independent Bernoulli variables with parameters $\theta/(\theta + j - 1)$ as in (2.1). The device (2.1) with independent variables Θ_j is then the simplest way to make these indicators simultaneously for all j and n to be increasing in θ, which is all that is needed to make $(\Pi_{n,\theta},\ \theta \geqslant 0)$ a Ewens fragmentation. What is much less obvious is the consistency of these processes for various n. To put this another way, if in the process of splitting a set of m balls according to the indicators $1(\Theta_j \leqslant \theta)$ for $1 \leqslant j \leqslant m$ we pass to the balls-in-boxes picture and just observe the splitting process restricted to a uniformly chosen random subset of $n < m$ balls, this sub-process is identical in distribution to the process of splitting of the first n balls using the indicators $1(\Theta_j \leqslant \theta)$ for $1 \leqslant j \leqslant n$. This sampling consistency property of compositions, which is so intuitive in the balls-in-boxes picture, is quite painful to express entirely in the binary encoding. We circumvent that difficulty by deriving consistency from the Poisson representation of the corresponding mass fragmentation model, which is introduced in the next section. See also [9] for a more extensive discussion of such consistency properties of partition structures derived by random sampling from self-similar random sets, like the self-similar Poisson process in the next section or the zero set of a Brownian motion.

3. Poisson representation of the mass fragmentation

Let (T_i, V_i) be a listing of the points of a homogeneous Poisson point process on the positive quadrant $\mathbb{R}^2_+$ with rate 1 per unit area. For each fixed $\theta > 0$, the random countable set

$$Z_\theta := \{T_i : 0 < T_i < 1 \text{ and } V_i \leqslant \theta/T_i\}$$

is then the set of points of a Poisson process on $[0, 1]$ with intensity $\theta t^{-1} dt$. To orient the reader, we start by recalling some well-known properties of Z_θ and the induced random composition [1, 6, 9, 10, 14].

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

(i) Let $Y_{1,\theta} > Y_{2,\theta} > \cdots$ be the points of Z_θ in decreasing order. Then

$$Y_{j,\theta} = \prod_{i=1}^{j} W_{i,\theta},$$

where the $W_{i,\theta}$ $(i = 1, 2, \ldots)$ are independent and identically distributed random variables with beta$(\theta, 1)$ distribution.

(ii) If $P_{j,\theta}$ is the length of the jth component interval of $[0,1]\backslash Z_\theta$, working from right to left, that is $P_{j,\theta} = Y_{j-1,\theta} - Y_{j,\theta}$ where $Y_{0,\theta} := 1$, then

$$P_{j,\theta} = (1 - W_{j,\theta}) \prod_{1=1}^{j-1} W_{i,\theta},$$

for $W_{i,\theta}$ i.i.d. beta$(\theta, 1)$ as before. The distribution of this random discrete probability distribution $(P_{j,\theta},\ j \geqslant 1)$ is known as the GEM(θ) distribution.

(iii) The distribution of the decreasing rearrangement of $(P_{j,\theta},\ j \geqslant 1)$ is the *Poisson–Dirichlet distribution* with parameter θ.

(iv) Let $U_1, U_2, \ldots$ be a sequence of independent uniform $[0,1]$ variables, independent of Z_θ, and define a random partition $\Pi_{\infty,\theta}$ of the set $\mathbb{N}$ of positive integers to be the collection of equivalence classes for the random equivalence relation: $i \sim^\theta j$ if and only if either $i = j$ or both U_i and U_j fall in the same component interval of $[0,1]\backslash Z_\theta$. Then $\Pi_{\infty,\theta}$ is an exchangeable random partition of the infinite set $\mathbb{N}$, whose restriction $\Pi_{n,\theta}$ to $[n]$ is a Ewens partition governed by (1.1) for each n.

(v) Let $U_{n,j}$ be the jth-smallest value among $U_1, \ldots, U_n$, and let $X_{n,j}(\theta)$ be the indicator of the event that $U_{n,j}$ is the least value among those of the n values which fall in some component interval of $[0,1]\backslash Z_\theta$. Then, for each fixed n and θ, the $X_{n,j}(\theta)$ for $j = 1, 2, \ldots, n$ are independent, with

$$\mathbb{P}(X_{n,j}(\theta) = 1) = \theta/(\theta + j - 1). \tag{3.1}$$

More precisely, the sequence $(X_{n,j}(\theta), 1 \leqslant j \leqslant n)$ is the binary encoding of a random composition $C_{n,\theta}$ of n which is a particular ordering of the sizes of blocks of $\Pi_{n,\theta}$. If the blocks of $\Pi_{n,\theta}$ are, say, $B_1, \ldots, B_k$, then the sizes of these blocks appear in the composition in increasing order of values of $\min_{i \in B_k} U_i$ [1].

An immediate consequence of the above construction of Z_θ from a Poisson process in the positive quadrant is that the random set Z_θ increases as θ increases. Consequently, the various quantities introduced above describe a process of fragmentation of $[0,1]$ into subintervals. In particular, the partition valued process $(\Pi_{n,\theta},\ \theta \geqslant 0)$ is refining as θ increases, and each of the processes $(X_{n,j}(\theta),\ \theta \geqslant 0)$ is increasing as θ increases. Consistency and exchangeability of the processes $(\Pi_{n,\theta},\ \theta \geqslant 0)$ are obvious from (iv).

The proof of the consistency property claimed in Theorem 2.1 is completed by the following lemma, which shows that in this Poisson set-up the indicator variables

[1] Strictly speaking, this defines $C_{n,\theta}$ in terms of U_js and Z_θ, rather than through $\Pi_{n,\theta}$. Conditionally given $\Pi_{n,\theta} = \{B_1, \ldots, B_k\}$ the composition has the same distribution as the sequence $\#B_j$ arranged by decrease of minimal elements $\min B_j$. Conditionally given the induced partition $\{\#B_j,\ j \leqslant k\}$ of integer n, this arrangement is the inverse size-biased ordering of the block sizes.

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

$(X_{n,j}(\theta), 1 \leqslant j \leqslant n)$ in the binary expansion of the composition $C_{n,\theta}$ associated with the natural ordering of blocks of $\Pi_{n,\theta}$ (as in (v)) can be derived from independent variables $(\Theta_{n,j},\ 1 \leqslant j \leqslant n)$ with the same distribution as $(\Theta_j,\ 1 \leqslant j \leqslant n)$ in Theorem 2.1.

Lemma 3.1. *Let* $\Theta_{n,j} := \inf\{\theta : X_{n,j}(\theta) = 1\}$, *so that* $X_{n,j}(\theta)$ *may be represented as*

$$X_{n,j}(\theta) = 1(\Theta_{n,j} \leqslant \theta).$$

Then for each fixed n *the* $\Theta_{n,j}$ *are independent random variables with*

$$\mathbb{P}(X_{n,j}(\theta) = 1) = \mathbb{P}(\Theta_{n,j} \leqslant \theta) = \theta/(\theta + j - 1), \quad (j = 1, \ldots, n).$$

Proof. Fix $\theta_j > 0$ for $j = 2, \ldots, n$. Observe that the event $(\Theta_{n,j} > \theta_j)$ occurs if and only if there is no point (T_i, V_i) of the Poisson process with $T_i \in [U_{n,j-1}, U_{n,j}]$ and $V_i \leqslant \theta/T_i$. Therefore

$$\mathbb{P}\big(\cap_{j=2}^n(\Theta_{n,j} > \theta_j) \,|\, U_{n,j}, 1 \leqslant j \leqslant n\big) = \exp\left(-\sum_{j=2}^n \int_{U_{n,j-1}}^{U_{n,j}} \frac{\theta_j}{t} dt\right) = \prod_{j=2}^n \left(\frac{U_{n,j-1}}{U_{n,j}}\right)^{\theta_j}$$

and hence

$$\mathbb{P}\big(\cap_{j=2}^n(\Theta_{n,j} > \theta_j)\big) = \mathbb{E}\left[\prod_{j=2}^n \left(\frac{U_{n,j-1}}{U_{n,j}}\right)^{\theta_j}\right] = \prod_{j=2}^n \frac{j-1}{j-1+\theta_j},$$

because the ratios $U_{n,j-1}/U_{n,j}$ are independent with beta$(j-1, 1)$ distributions, $2 \leqslant j \leqslant n$. □

As before, let $(C_{n,\theta},\ \theta \geqslant 0)$ be the process of refining compositions of n, defined either through indicators as in Theorem 2.1, or by means of the Poisson construction as in (v) above. Immediately from the definition, we have the following result.

Corollary 3.2. $(C_{n,\theta}, \theta \geqslant 0)$ *is a Markov process whose inhomogeneous transition rates are determined by the following rule. If at time* θ *the state is the composition of* n *encoded by some binary sequence starting with* 1, *each* 0 *is switching to* 1 *at rate* $1/(\theta + j - 1)$, *where* j *is the place of this* 0 *in the sequence, while all other transition rates are trivial.*

Proof. Indeed,

$$\mathbb{P}(\Theta_j \in [\theta,\ \theta + d\theta] \,|\, \Theta_j > \theta) = \frac{\left(\frac{\theta}{\theta+j-1}\right)' d\theta}{1 - \frac{\theta}{\theta+j-1}} = \frac{d\theta}{\theta + j - 1}.$$ □

If we view $\Pi_{\infty,\theta}$ together with the total ordering of the blocks, as induced by the natural order of intervals (recall (iv)), we obtain an exchangeable ordered partition $\Pi^*_{\infty,\theta}$ of $\mathbb{N}$. Let $\Pi^*_{n,\theta}$ be its restriction to $[n]$. The law of $\Pi^*_{n,\theta}$ is given by the ordered version of the Ewens sampling formula (1.1):

$$\mathbb{P}(\Pi^*_{n,\theta} = \pi^*) = \frac{\theta^{k-1}}{[\theta+1]_{n-1}} \prod_{j=1}^k \frac{n_j!}{n_1 + \cdots + n_j}$$

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

for every π^* ordered partition of $[n]$ with block sizes $(n_1,\dots,n_k)$. The process $(\Pi_{n,\theta},\ \theta\geqslant 0)$ is Markovian for every n. The transition mechanism of $(\Pi^*_{n,\theta},\ \theta\geqslant 0)$ is determined by that of $(C_{n,\theta},\ \theta\geqslant 0)$ and the following allocation rule[2]: each time a block B_j of size a splits in two fragments of sizes ξ and η, all $\binom{a}{\xi}$ possible allocations of the elements of B_j among the offspring fragments are equally likely.

4. Further properties

In principle, the finite-dimensional distributions of $(\Pi_{n,\theta},\ \theta\geqslant 0)$ may be determined by some summations of probabilities determined by the Markov process $(C_{n,\theta},\theta\geqslant 0)$. But such formulas appear to be of limited value. It appears that the process $(\Pi_{n,\theta},\theta\geqslant 0)$ is not Markovian.

Proof that for $n>2$ the fragmentation process is not Markovian. Consider the random time $\widehat{\Theta}_n$ of the first split, that is,

$$\widehat{\Theta}_n=\min_{2\leqslant j\leqslant n}\Theta_{n,j}=\inf\{\theta:\Pi_{n,\theta}\neq\{[n]\}\},$$

where $\Pi_{n,0}=\{[n]\}$ is the initial partition with a single block. To show that the Markov property of the fragmentation process $(\Pi_{n,\theta},\theta\geqslant 0)$ does not hold for every $n>2$ we shall focus on the conditional probability

$$Q(t):=\mathbb{P}(\Pi_{n,\theta}=\lambda\,|\,\Pi_{n,\phi}=\lambda,\ \Pi_{n,t}=\lambda)=\mathbb{P}(\Pi_{n,\theta}=\lambda\,|\,\Pi_{n,\phi}=\lambda,\ \widehat{\Theta}_n<t),$$

where ϕ,θ are considered as parameters, $0<t<\phi<\theta$, and λ is the partition of $[n]$ in two blocks $\{1\}$ and $\{2,\dots,n\}$. To disprove the Markov property it is sufficient to show that $Q(t)$ is not constant as t varies.

Note that $\Pi_{n,\phi}=\lambda$ is only possible when the composition $C_{n,\phi}$ assumes either the value $1100\cdots 0$ or the value $100\cdots 01$, and conditionally given either of these values $\Pi_{n,\phi}$ equals λ with probability $1/n$ (as a consequence of exchangeability). Working out this dichotomy,

$$\begin{aligned}
&\mathbb{P}(\Pi_{n,\phi}=\lambda,\widehat{\Theta}_n<t)\\
&\quad=\mathbb{P}(\Theta_{n,2}<t,\Theta_{n,n}>\phi\ ;\ \Theta_{n,3}>\phi,\dots,\Theta_{n,n-1}>\phi)\,\frac{1}{n}\\
&\qquad+\mathbb{P}(\Theta_{n,2}>\phi,\Theta_{n,n}<t\ ;\ \Theta_{n,3}>\phi,\dots,\Theta_{n,n-1}>\phi)\,\frac{1}{n}\\
&\quad=\frac{t}{t+2-1}\left(1-\frac{\phi}{\phi+n-1}\right)P(\phi)\,\frac{1}{n}+\left(1-\frac{\phi}{\phi+2-1}\right)\frac{t}{t+n-1}\,P(\phi)\,\frac{1}{n}\\
&\quad=\left(\frac{t(n-1)}{(t+1)(\phi+n-1)}+\frac{t}{(\phi+1)(t+n-1)}\right)P(\phi)\frac{1}{n},
\end{aligned}$$

where, for brevity, $P(\phi):=\mathbb{P}(\Theta_{n,3}>\phi,\dots,\Theta_{n,n-1}>\phi)$. Noting the inclusion

$$\{\Pi_{n,\theta}=\lambda,\widehat{\Theta}_n<t\}\subset\{\Pi_{n,\phi}=\lambda,\widehat{\Theta}_n<t\}$$

[2] Which is common for all exchangeable fragmentation processes.

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

and applying the above formula to the event on the left-hand side, we compute

$$Q(t) = \frac{\mathbb{P}(\Pi_{n,\theta} = \lambda,\, \widehat{\Theta}_n < t)}{\mathbb{P}(\Pi_{n,\phi} = \lambda,\, \widehat{\Theta}_n < t)} = \left(\frac{t(n\theta + 2n - 2) + (n-1)^2(\theta+1) + \theta + n - 1}{t(n\phi + 2n - 2) + (n-1)^2(\phi+1) + \phi + n - 1} \right) \frac{P(\phi)}{P(\theta)}.$$

This does not depend on t if and only if

$$\frac{n\theta + 2n - 2}{n\phi + 2n - 2} = \frac{(n-1)^2(\theta+1) + \theta + n - 1}{(n-1)^2(\phi+1) + \phi + n - 1},$$

or, equivalently, if and only if the polynomial

$$\begin{aligned}(n\theta + 2n - 2)&((n-1)^2(\phi+1) + \phi + n - 1) \\ &= n(n^2 - 2n + 2)\theta\phi + n^2(n-1)\theta + 2(n-1)(n^2 - 2n + 2)\phi + 2n(n-1)^2\end{aligned}$$

is symmetric in ϕ and θ. To maintain symmetry we must have

$$n^2(n-1) = 2(n-1)(n^2 - 2n + 2),$$

which forces positive n to be either 1 or 2. Thus for $n > 2$ the partition-valued process is not Markovian (while it is trivially Markovian for $n = 1$ or 2).

Transition rates of the Ewens fragmentation. Given the value π of $\Pi_{n,\theta}$, the composition $C_{n,\theta}$ can be recovered by arranging the sequence of block sizes of π in the reversed size-biased order. This property taken together with Corollary 3.2 allows us to compute the transition rates. To illustrate the method, suppose that at time θ the partition $\Pi_{n,\theta}$ is in state π with block sizes $\{a, b\}$, and let σ be some nontrivial refinement of π with block sizes $\{\xi, \eta, b\}$, so $\xi + \eta = a$. Suppose first that $a \neq b$, $\xi \neq \eta$. Then $C_{n,\theta} = (a, b)$ with probability $b/(a+b)$, and $C_{n,\theta} = (b, a)$ with probability $a/(a+b)$. Inspecting all possibilities we see that the fragmentation process jumps from π to σ at rate

$$\frac{b}{a+b}\left(\frac{1}{\theta+\xi} + \frac{1}{\theta+\eta}\right)\binom{a}{\xi}^{-1} + \frac{a}{a+b}\left(\frac{1}{\theta+\xi+b} + \frac{1}{\theta+\eta+b}\right)\binom{a}{\xi}^{-1},$$

where the binomial coefficients account for the number of ways of allocating the elements of the splitting block among two new fragments. A minute's thought shows that the formula is still valid in the case $a = b$; but if $\xi = \eta$ the above expression should be halved.

In principle, there exists a Markovian family of Ewens fragmentations, with the same transition rates as those of $(\Pi_{n,\theta})$. However, this seems to be of little use, because the formulas for these rates become increasingly complicated when the number of blocks grows.

Comparison with the Ewens fragmentation derived from Kingman's coalescent. Another family of Ewens fragmentations was derived in [2] from Kingman's coalescent tree. These fragmentations are not Markovian in the proper sense, and no extended Markov property for them is known. We show next that the partition-valued process in [2] is different from the process constructed in this paper.

As before, consider the time of the first split $\widehat{\Theta}_n$. Let $I_n + 1$ be the almost surely unique index j which makes $\widehat{\Theta}_n = \Theta_{n,j}$. Then the split at time $\widehat{\Theta}_n$ creates a partition $\Pi_{n,\widehat{\Theta}_n}$ with

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

two blocks of sizes I_n and $n-I_n$. Conditioning on $\widehat{\Theta}_n$ gives

$$\mathbb{P}(I_n=i)=\int_0^\infty \frac{\mathbb{P}(\Theta_{n,i+1}\in d\theta)}{\mathbb{P}(\Theta_{n,i+1}>\theta)}\prod_{2\leqslant j\leqslant n}\mathbb{P}(\Theta_{n,j}>\theta), \tag{4.1}$$

which simplifies to

$$\mathbb{P}(I_n=i)=(n-1)!\int_0^\infty \frac{d\theta}{(\theta+i)[\theta+1]_{n-1}}. \tag{4.2}$$

In particular, for $n=4$ this gives

$$\begin{aligned}\mathbb{P}(I_4=1)&=6\left(-\log 2+\frac{1}{4}\log 3+\frac{1}{2}\right),\\ \mathbb{P}(I_4=2)&=6\left(\frac{1}{2}\log 3-\frac{1}{2}\right),\\ \mathbb{P}(I_4=3)&=6\left(\log 2-\frac{3}{4}\log 3-\frac{1}{6}\right).\end{aligned}$$

Compare the second of these evaluations with the corresponding formula in [2, Section 5.1] to see that this Ewens fragmentation process evolves differently to the Ewens fragmentation derived from Kingman's coalescent, which has a different distribution on partitions of [4] with two blocks at the time of the first split. Neither of these distributions is that of $\Pi_{4,\theta}$ given that this partition has exactly two blocks, even though this conditional distribution does not depend on θ. This common conditional distribution is the Gibbs distribution on partitions of [4] into 2 blocks, which assigns probability proportional to $(n_1-1)!(n_2-1)!$ to each partition of [4] into two blocks of sizes n_1 and n_2.

It is still an open question for which n there exists a discrete time fragmentation process, on partitions of $[n]$ whose distribution at time k is the distribution on partitions of $[n]$ into k blocks, which assigns each such partition into blocks of sizes $\{n_1,\dots,n_k\}$ a probability proportional to $\prod_{i=1}^k(n_i-1)!$.

References

[1] Arratia, R., Barbour, A. D. and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach*, Vol. 1 of *EMS Monographs in Mathematics*, European Mathematical Society Publishing House, Zürich.
[2] Berestycki, N. and Pitman, J. (2006) Gibbs distributions for random partitions generated by a fragmentation process. *J. Stat. Phys.*
[3] Bertoin, J. (2006) *Random Fragmentation and Coagulation Processes*, Cambridge University Press.
[4] Chase, K. C. and Mekjian, A. Z. (1994) Nuclear fragmentation and its parallels. *Phys. Rev. C* **49** 2164–2176.
[5] Evans, S. N. and Pitman, J. (1998) Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré* **34** 339–383.
[6] Gnedin, A. (2004) Three sampling formulas. *Combin. Probab. Comput.* **13** 185–193.
[7] Gnedin, A. and Pitman, J. (2004) Regenerative partition structures. *Electron. J. Combin.* **11** # 12.

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press

[8] Gnedin, A. and Pitman, J. (2005) Regenerative composition structures. *Ann. Probab.* **33** 445–479.
[9] Gnedin, A. and Pitman, J. (2005) Self-similar and Markov composition structures. In *Representation Theory, Dynamical Systems, Combinatorial and Algorithmic Methods, Part 13* (A. A. Lodkin, ed.), Vol. 326 of *Zapiski Nauchnyh Seminarov POMI*, pp. 59–84.
[10] Ignatov, T. (1982) A constant arising in the asymptotic theory of symmetric groups, and Poisson–Dirichlet measures. *Theor. Probab. Appl.* **27** 136–147.
[11] Lee, S. J. and Mekjian, A. Z. (1992) Canonical studies of the cluster distribution, dynamical evolution, and critical temperature in nuclear multifragmentation processes. *Phys. Rev. C* **45** 1284–1310.
[12] Mekjian, A. Z. (1991) Cluster distributions in physics and genetic diversity. *Phys. Rev. A* **44** 8361–8374.
[13] 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.
[14] Pitman, J. (2006) *Combinatorial Stochastic Processes*, Vol. 1875 of *Lecture Notes in Mathematics*, Springer.

https://doi.org/10.1017/S0963548306008352 Published online by Cambridge University Press