Stochastic Processes and their Applications 13 (1982) 235–248
North-Holland Publishing Company

THE COALESCENT

J.F.C. KINGMAN
Mathematical Institute, University of Oxford, Oxford OX1 3LB, England

Received 26 November 1980
Revised 26 September 1981

The n-coalescent is a continuous-time Markov chain on a finite set of states, which describes the family relationships among a sample of n members drawn from a large haploid population. Its transition probabilities can be calculated from a factorization of the chain into two independent components, a pure death process and a discrete-time jump chain. For a deeper study, it is useful to construct a more complicated Markov process in which n-coalescents for all values of n are embedded in a natural way.

Genetical models	haploid genealogy
random equivalence relations	coupling
exchangeability	Markov process
jump chain	

1. The n coalescent

For any natural number n, let $\mathscr{E}_n$ denote the finite set of equivalence relations on $\{1, 2, \ldots, n\}$. For $R \in \mathscr{E}_n$, denote by $|R|$ the number of equivalence classes of R. A continuous-time Markov chain $\{R_t; t \geq 0\}$ with state space $\mathscr{E}_n$ is said to be an *n-coalescent* if R_0 is the identity relation

$$\Delta = \{(i, i); i = 1, 2, \ldots, n\}, \tag{1.1}$$

and the transition rates

$$q_{\xi\eta} = \lim_{h \downarrow 0} h^{-1} \mathbf{P}\{R_{t+h} = \eta \,|\, R_t = \xi\}, \tag{1.2}$$

$\xi, \eta \in \mathscr{E}_n$, $\xi \neq \eta$, are given by

$$q_{\xi\eta} = \begin{cases} 1 & \text{if } \xi < \eta, \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

Here $\xi < \eta$ denotes that η is obtained from ξ by combining two of its equivalence classes, so that

$$\xi < \eta \Leftrightarrow \xi \subset \eta, \quad |\xi| = |\eta| + 1. \tag{1.4}$$

0304-4149/82/0000-0000/$02.75

Because $\mathscr{E}_n$ is finite, such chains exist and all have the same finite-dimensional distributions (the same *name*, in Kendall's terminology [4]). By the usual abuse of language, we talk of *the* n-coalescent when we wish to make generic statements about n-coalescents.

The n-coalescent was introduced in [8] in response to the demands of population genetics. If a sample of n individuals is taken at time t_0 from a large haploid population, and if R_t consists of those pairs (i, j) for which the ith and jth members of the sample have a common ancestor alive at time $t_0 - t$, then (with a proper time scale, and making certain biological assumptions) the process $\{R_t\}$ has the stochastic structure of the n-coalescent. We refer the reader interested in these applications to [8]; the robustness of the n-coalescent as an approximation for large population size in a variety of models will be explored elsewhere.

We are concerned with the properties of the Markov chain itself. Follow [8] in noting that the total transition rate

$$q_\xi = \lim_{h \downarrow 0} h^{-1} \mathbf{P}\{R_{t+h} \neq \xi \mid R_t = \xi\} = \sum_{\eta \neq \xi} q_{\xi\eta} \tag{1.5}$$

out of ξ is given by

$$q_\xi = \tfrac{1}{2}|\xi|(|\xi| - 1), \tag{1.6}$$

so that the sojourn time in any state ξ with $|\xi| = k$ has a probability density

$$d_k \, \mathrm{e}^{-d_k t} \quad (t > 0), \qquad d_k = \tfrac{1}{2}k(k-1), \tag{1.7}$$

depending only on $|\xi|$. Moreover, the transition from ξ must be to a state η with $|\eta| = |\xi| - 1$. Hence the process

$$D_t = |R_t| \tag{1.8}$$

is itself a Markov chain with states $1, 2, \ldots, n$, having transition rates

$$\lim_{h \downarrow 0} h^{-1} \mathbf{P}\{D_{t+h} = l \mid D_t = k\} = \begin{cases} d_k & \text{if } l = k-1, \\ 0 & \text{if } l \neq k, k-1. \end{cases} \tag{1.9}$$

In the usual terminology $\{D_t; t \geq 0\}$ is a pure death process with initial state n and death rates d_k.

The state 1 is absorbing for $\{D_t\}$, corresponding to the absorbing state

$$\Theta = \{(i, j); i, j = 1, 2, \ldots, n\} \tag{1.10}$$

for $\{R_t\}$. The *transit time*

$$T = \inf\{t \geq 0, R_t = \Theta\} = \inf\{t \geq 0; D_t = 1\} \tag{1.11}$$

can be represented as

$$T = \sum_{k=2}^{n} \tau_k, \tag{1.12}$$

where τ_k is the sojourn time of $\{D_t\}$ in state k; the τ_k are independent with respective distributions (1.7). These simple facts are exploited in [8, Section 5].

A typical sample path of $\{R_t\}$ moves through a sequence of equivalence relations

$$\Delta = \mathcal{R}_n < \mathcal{R}_{n-1} < \mathcal{R}_{n-2} < \cdots < \mathcal{R}_2 < \mathcal{R}_1 = \Theta, \tag{1.13}$$

spending time τ_k in $\mathcal{R}_k$. Clearly

$$|\mathcal{R}_k| = k. \tag{1.14}$$

It is a standard fact (see, for instance, [1, Section II.19] or [3, Section 8.3] that the sequence (1.13) forms a Markov chain, the *jump chain* of the n-coalescent.

2. The jump chain

Theorem 1. *In an n-coalescent, the death process $\{D_t; t \geq 0\}$ and the jump chain $\{\mathcal{R}_k; k = n, n-1, n-2, \ldots, 1\}$ are independent, and*

$$R_t = \mathcal{R}_{D_t} \tag{2.1}$$

for all $t \geq 0$. The transition probabilities of the Markov chain $\{\mathcal{R}_k\}$ are given by

$$\mathbf{P}\{\mathcal{R}_{k-1} = \eta \,|\, \mathcal{R}_k = \xi\} = \begin{cases} 2/k(k-1) & \text{if } \xi < \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

whenever $\xi \in \mathcal{E}_n$, $|\xi| = k$, $2 \leq k \leq n$. The absolute probabilities are given by

$$\mathbf{P}\{\mathcal{R}_k = \xi\} = \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \lambda_1! \lambda_2! \cdots \lambda_k!, \tag{2.3}$$

if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the sizes of the equivalence classes of ξ.

Proof. According to the theory of jump chains, the transition probabilities are of the form

$$q_{\xi\eta}/q_\xi \quad (\xi \neq \eta)$$

so long as $q_\xi > 0$ (the chain terminates on reaching a state with $q_\xi = 0$), and conditioned on the jump chain the sojourn times are independent, the sojourn time in a state ξ having probability density

$$q_\xi \, \mathrm{e}^{-q_\xi t} \quad (t > 0).$$

Applying this to the n-coalescent, (2.2) is immedate. If $\mathcal{R}_k = \xi$, then $q_\xi = d_k$, and so the conditional distribution of τ_k, given the jump chain, is the same as its unconditional distribution (1.7). Thus the conditional joint distributions of $\{D_t\}$ given $\{\mathcal{R}_k\}$ are the same as the corresponding unconditional distributions, showing that the two processes are independent. (2.1) follows at once from the definitions of D_t and $\mathcal{R}_k$.

We prove (2.3) by backward induction on k, it being clearly true for $k = n$. By (2.2),

$$p_k(\xi) = \mathbf{P}\{\mathcal{R}_k = \xi\}, \quad \xi \in \mathcal{E}_n, |\xi| = k,$$

satisfies

$$p_{k-1}(\eta) = \sum_{\xi < \eta} \frac{2}{k(k-1)} p_k(\xi).$$

If $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$ are the sizes of the equivalence classes of η, those of ξ are λ_1, $\lambda_2, \ldots, \lambda_{l-1}, \nu, \lambda_l - \nu, \lambda_{l+1}, \ldots, \lambda_{k-1}$ for some l, $1 \leq l \leq k-1$, and some ν, $1 \leq \nu \leq \lambda_l - 1$. If for the purpose of induction we assume that p_k is given by (2.3), we have

$$p_{k-1}(\eta) = \sum_{l=1}^{k-1} \sum_{\nu=1}^{\lambda_l - 1} \frac{2}{k(k-1)} \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \lambda_1! \cdots \lambda_{l-1}! \nu!(\lambda_l - \nu)! \cdots \lambda_{k-1}! \tfrac{1}{2} \binom{\lambda_l}{\nu}$$

$$= \frac{(n-k)!(k-1)!(k-2)!}{n!(n-1)!} \lambda_1! \lambda_2! \cdots \lambda_{k-1}! \sum_{l=1}^{k-1} \sum_{\nu=1}^{\lambda_l - 1} 1,$$

which yields (2.3) with k replaced by $(k-1)$ because

$$\sum_{l=1}^{k-1} \sum_{\nu=1}^{\lambda_l - 1} 1 = \sum_{l=1}^{k-1} (\lambda_l - 1) = n - (k-1).$$

Hence the theorem is proved.

The same induction argument may be used to compute all the joint distributions of $\{\mathcal{R}_k\}$. The reader will readily verify the fact (which is anyway obvious if (2.3) is combined with [8, Section 6] that, for $l < k$, $|\xi| = k$, $|\eta| = l$, $\xi \subset \eta$,

$$\mathbf{P}\{\mathcal{R}_l = \eta \,|\, \mathcal{R}_k = \xi\} = \frac{(k-l)!l!(l-1)!}{k!(k-1)!} \lambda_1! \lambda_2! \cdots \lambda_l!, \tag{2.4}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_l$ are the sizes of the equivalence classes of the relation in $\mathcal{E}_k$ which η induces on the equivalence classes of ξ.

Theorem 1 determines the finite-dimensional distributions of the n-coalescent itself. For example, if $\xi \in \mathcal{E}_n$ has $|\xi| = k$, then (2.1) shows that

$$\mathbf{P}\{R_t = \xi\} = \mathbf{P}\{D_t = k\} \mathbf{P}\{\mathcal{R}_k = \xi\}. \tag{2.5}$$

The first element in this factorization is given by convolutions of the negative exponential distributions (1.7), since

$$\mathbf{P}\{D_t = k\} = \mathbf{P}\Big\{\sum_{r=k+1}^{n} \tau_r \leq t\Big\} - \mathbf{P}\Big\{\sum_{r=k}^{n} \tau_r \leq t\Big\}, \tag{2.6}$$

while the second element is given by (2.3). It is perhaps rather surprising, in view of the many possible sample paths through the complex set $\mathcal{E}_n$, that (2.3) should take such a simple form. However, though simple it is by no means easy to handle when n is large.

It is suggested in [8] that there could be some advantage in embedding n-coalescents for all values of n in a single random process. Specifically, let $\mathscr{E}$ be the (uncountable) set of all equivalence relations on $\mathbb{N}=\{1, 2, 3, \ldots\}$, and define $\rho_n : \mathscr{E} \to \mathscr{E}_n$ by restriction: for $R \in \mathscr{E}$,

$$\rho_n R = \{(i, j); 1 \leq i, j \leq n, (i, j) \in R\}. \tag{2.7}$$

Then a proof was sketched in [8] of the existence of a random process $\{R_t; t \geq 0\}$ with values in $\mathscr{E}$ such that, for all $n \in \mathbb{N}$, $\{\rho_n R_t; t \geq 0\}$ is an n-coalescent.

We here give a different proof of that result, based on the factorisation of Theorem 1, which gives a more direct construction and explicit formulae for the finite-dimensional distributions of the $\mathscr{E}$-valued process. It was noted in [8] that the pure death process could be defined, as it were, for $n = \infty$, by noting that the series $\sum d_k^{-1}$ converges. Thus a pure death process $\{D_t; t > 0\}$ exists with death rates d_k and

$$\lim_{t \downarrow 0} D_t = \infty; \tag{2.8}$$

this makes a transition from k to $(k-1)$ at time

$$\sum_{r=k}^{\infty} \tau_r \tag{2.9}$$

where the τ_r are as before independent with densities (1.7), and (2.9) has finite expectation

$$\sum_{r=k}^{\infty} d_r^{-1} = \frac{2}{k-1}.$$

We now try to define a discrete-time Markov process $\{\mathscr{R}_k; k \in \mathbb{N}\}$, taking values in $\mathscr{E}$, so that (2.1) is the required continuous-time process. To do this requires an efficient way of handling distributions on the set $\mathscr{E}$. This machinery exists when, as here, the distributions are invariant under permutations, essentially because in Theorem 2 we have a variant of de Finetti's theorem.

3. Exchangeable equivalence relations

An equivalence relation on $\mathbb{N}$ is of course a subset of $\mathbb{N}\times\mathbb{N}$, and so $\mathscr{E}$ can be regarded as a subset of the set $2^{\mathbb{N}\times\mathbb{N}}$. If we give $2^{\mathbb{N}\times\mathbb{N}}$ its product topology, $\mathscr{E}$ is closed. Hence the subspace topology for $\mathscr{E}$ is compact and metrisable. It can also be described as the weakest topology making all the functions $\rho_n : \mathscr{E} \to \mathscr{E}_n$ (the latter with the discrete topology) continuous. Since the ρ_n separate points of $\mathscr{E}$, the Stone–Weierstrass theorem shows that any continuous $f: \mathscr{E} \to \mathbb{R}$ can be approximated by functions $g \circ \rho_n$ ($g: \mathscr{E}_n \to \mathbb{R}$). We shall use this topology, and the induced measurable structure, for $\mathscr{E}$ without further comment.

A probability measure on $\mathscr{E}$ is called *exchangeable* if, for any permutation $\pi: \mathbb{N} \to \mathbb{N}$, it is invariant under the induced bijection $\hat{\pi}: \mathscr{E} \to \mathscr{E}$ defined by

$$\hat{\pi}R = \{(\pi i, \pi j); (i, j) \in R\}. \tag{3.1}$$

A random equivalence relation R is called exchangeable if its distribution is, i.e. if $\hat{\pi}R$ has the same distribution as R for all π.

One way of constructing an exchangeable random equivalence relation is by the paintbox construction of [7]. Let $x_0, x_1, x_2, \ldots$ satisfy

$$x_r \geq 0, \qquad \sum_{r=0}^{\infty} x_r = 1. \tag{3.2}$$

Let $Z_1, Z_2, \ldots$ be independent random variables with the same distribution

$$\mathbf{P}\{Z_j = r\} = x_r, \quad r = 0, 1, 2, \ldots, \tag{3.3}$$

and define

$$R = \{(i, j); i = j \text{ or } Z_i = Z_j \geq 1\}. \tag{3.4}$$

It is clear that the distribution of R is an exchangeable probability measure P^x depending only on the sequence

$$x = (x_0, x_1, x_2, \ldots). \tag{3.5}$$

Notice that P^x is unchanged if some of the x_r for $r \geq 1$ are permuted. For this reason it is sometimes convenient to normalise so that

$$x_1 \geq x_2 \geq x_3 \geq \cdots. \tag{3.6}$$

However, x_0 plays a special role, and P^x is affected if it is interchanged with another x_r.

The construction can of course be generalised by allowing the sequence x to be random (and interpreting (3.3) and the independence of the Z_j as being conditional on x). This yields the distribution

$$P = \int P^x \mu(\mathrm{d}x), \tag{3.7}$$

where μ is the distribution of x, and the integral extends over all sequences satisfying (3.2). What is much less trivial is that any exchangeable probability measure on $\mathscr{E}$ is of this form for some μ.

Theorem 2. *Let R be an exchangeable random equivalence relation on $\mathbb{N}$. For any $r, n \in \mathbb{N}$, let $\lambda_r(n)$ denote the size of the rth largest equivalence class of $\rho_n R$. Then the limits*

$$X_r = \lim_{n \to \infty} n^{-1} \lambda_r(n) \tag{3.8}$$

exist with probability one, and X_0 may be chosen so that

$$X = (X_0, X_1, X_2, \ldots) \tag{3.9}$$

satisfies (3.2) *and* (3.6). *The conditional distribution of R, given X, is P^X. Hence the distribution of R is given by* (3.7), *where μ is the distribution of X.*

Proof. Let $\mathscr{I}_n$ be the σ-field of events defined in terms of R which are unchanged if R is replaced by $\hat{\pi}R$, for any permutation π for which all $m \geqslant n+1$ are fixed. Note that $\mathscr{I}_n \supseteq \mathscr{I}_{n+1}$, and that $\lambda_r(n)$ is $\mathscr{I}_n$-measurable. The exchangeability of R implies that the conditional distribution of $\rho_n R$, given $\mathscr{I}_n$, is invariant under permutations of $\{1, 2, \ldots, n\}$. There is only one invariant distribution on $\mathscr{E}_n$ with given values of $\lambda_r(n)$ $(r = 1, 2, \ldots)$, and it is given by the following recipe:

Let n balls be coloured with colours $C_1, C_2, \ldots$, so that $\lambda_r(n)$ has colour C_r. Let these be sampled without replacement, and let R_n contain (i, j) if the ith and jth balls have the same colour. Then the distribution of R_n is the conditional distribution of $\rho_n R$, given $\mathscr{I}_n$.

Consider in particular, for $m < n$, the random variable

$$\Lambda_r(m) = \lambda_1(m) + \lambda_2(m) + \cdots + \lambda_r(m).$$

This is not less than the number of the first m balls sampled which are of colours $C_1, C_2, \ldots, C_r$, and this random variable has expectation

$$mn^{-1}\Lambda_r(n).$$

Hence

$$\mathbf{E}\{m^{-1}\Lambda_r(m) \mid \mathscr{I}_n\} \geqslant n^{-1}\Lambda_r(n), \tag{3.10}$$

and a reversed martingale theorem of Doob [2, Theorem VII.4.25] shows that

$$\lim_{n\to\infty} n^{-1}\Lambda_r(n)$$

exists with probability one. This establishes the existence of the limits (3.8), Fatou's lemma shows that

$$\sum_{r=1}^{\infty} X_r \leqslant 1,$$

and the fact that $\lambda_{r+1}(n) \leqslant \lambda_r(n)$ shows that $X_{r+1} \leqslant X_r$. Hence the sequence (3.9), with

$$X_0 = 1 - \sum_{r=1}^{\infty} X_r,$$

satisfies (3.2) and (3.6).

We now compute the conditional distribution of $\rho_m R$, given the limit σ-field

$$\mathscr{I} = \bigcap_{n=1}^{\infty} \mathscr{I}_n. \tag{3.11}$$

For $m < n$, the conditional distribution given $\mathscr{I}_n$ is the distribution of $\rho_{mn}R_n$, where $\rho_{mn} : \mathscr{E}_n \to \mathscr{E}_m$ is the restriction map

$$\rho_{mn}R = \{(i, j); 1 \leq i, j \leq m, (i, j) \in R\}. \tag{3.12}$$

Thus it is the distribution of the "same colour" relation on the first m balls sampled from the n: if $\xi \in \mathscr{E}_m$ has equivalence classes of sizes $\nu_1, \nu_2, \ldots, \nu_k$, then

$$\mathbf{P}\{\rho_m R = \xi \,|\, \mathscr{I}_n\} = \{(n)_m\}^{-1} \sum_{\substack{r_1, r_2, \ldots, r_k \\ \text{distinct}}} (\lambda_{r_1}(n))_{\nu_1} (\lambda_{r_2}(n))_{\nu_2} \cdots (\lambda_{r_k}(n))_{\nu_k}, \tag{3.13}$$

where $(\lambda)_\nu = \lambda(\lambda - 1) \cdots (\lambda - \nu + 1)$.

As $n \to \infty$ in (3.13), the left-hand side converges to the conditional probability given $\mathscr{I}$. If ξ is such that $\nu_j \geq 2$ for all $j \leq k$, the dominated convergence theorem applies to the right-hand side since $\lambda_r(n) \leq n/r$, and (3.8) shows that

$$\mathbf{P}\{\rho_m R = \xi \,|\, \mathscr{I}\} = \sum_{\substack{r_1, r_2, \ldots, r_k \\ \text{distinct}}} X_{r_1}^{\nu_1} X_{r_2}^{\nu_2} \cdots X_{r_k}^{\nu_k}.$$

Thus we have proved that

$$\mathbf{P}\{\rho_m R = \xi \,|\, \mathscr{I}\} = P^X[\rho_m^{-1}\{\xi\}], \tag{3.14}$$

whenever $m \geq 1$, $\xi \in \mathscr{E}_m$ and ξ has no singletons.

Now extend (3.14) to all ξ by induction on the number of singletons of ξ. Suppose (3.14) is true for all ξ with less than s singletons, and let ξ be a relation in $\mathscr{E}_m$ with s singletons. By exchangeability, we may suppose that one of these is $\{m\}$. Then, if $\rho_{m-1,m}\xi = \eta$,

$$\mathbf{P}\{\rho_m R = \xi \,|\, \mathscr{I}\} = \mathbf{P}\{\rho_{m-1} R = \eta \,|\, \mathscr{I}\} - \sum \mathbf{P}\{\rho_m R = \zeta \,|\, \mathscr{I}\},$$

where the sum extends over $\zeta \neq \xi$ with $\rho_{m-1,m}\zeta = \eta$. Both η and all the ζ have less than s singletons, so that the right-hand side may be evaluated using (3.14). This results in (3.14) for ξ, so that the induction succeeds, and (3.14) is true without restriction.

Since (3.14) is true for all m, the Stone–Weierstrass property establishes that the conditional distribution of R itself, given $\mathscr{I}$, is P^X, and the theorem is proved.

Theorem 2 is a variant of the main result of [7], and the two proofs are closely related. Note that the measure μ constructed in the proof is concentrated on sequences satisfying (3.6), and that it is the only measure μ so concentrated, which satisfies (3.7) for the given P [6]. There is however another way to achieve uniqueness, in the special case when, for some k,

$$\mathbf{P}\{|R| > k\} = 0. \tag{3.15}$$

When this is true, μ can be taken as concentrated on the sequences with

$$x_{k+1} = x_{k+2} = \cdots = 0 \tag{3.16}$$

and to be (finitely) exchangeable with respect to $x_1, x_2, \ldots, x_k$. If this requirement is substituted for (3.6), μ is again unique.

One very important special case occurs when $x_0 = 0$ and μ is proportional to Lebesgue measure on the simplex

$$\Delta_k = \left\{(x_1, x_2, \ldots, x_k); x_r \geq 0, \sum_{r=1}^{k} x_r = 1\right\}. \tag{3.17}$$

The corresponding probability measure on $\mathscr{E}$ will be denoted by $\mathscr{P}_k$, so that

$$\mathscr{P}_k = \int_{\Delta_k} \cdots \int P^{(0, x_1, x_2, \ldots, x_k, 0, \ldots)}(k-1)!\, dx_1\, dx_2 \cdots dx_{k-1}. \tag{3.18}$$

Now suppose that R has distribution $\mathscr{P}_k$. We may compute the distribution of the restriction $\rho_n R$ because, for $\xi \in \mathscr{E}_n$, $|\xi| = l \leq k$,

$$\begin{aligned}
\mathbf{P}\{\rho_n R = \xi\} &= \int_{\Delta_k} \cdots \int \mathbf{P}\{Z_i = Z_j \Leftrightarrow (i,j) \in \xi; 1 \leq i, j \leq n\}(k-1)! \\
&\qquad \times dx_1 \cdots dx_{k-1} \\
&= \int_{\Delta_k} \cdots \int \sum\nolimits^{\xi} x_{r_1} x_{r_2} \cdots x_{r_n} (k-1)!\, dx_1 \cdots dx_{k-1} \\
&= \sum\nolimits^{\xi} \frac{\lambda_1! \lambda_2! \cdots \lambda_l! (k-1)!}{(k-1+\lambda_1+\cdots+\lambda_l)!} \\
&= \frac{k!}{(k-l)!} \frac{\lambda_1! \lambda_2! \cdots \lambda_l! (k-1)!}{(k-1+n)!},
\end{aligned}$$

where the sum extends over all $r_1, r_2, \ldots, r_n$ for which $r_i = r_j$ if and only if $(i,j) \in \xi$, and $\lambda_1, \lambda_2, \ldots, \lambda_l$ are the sizes of the equivalence classes of ξ.

Hence a random equivalence relation R on $\mathbb{N}$ with distribution $\mathscr{P}_k$, has

$$\mathbf{P}\{\rho_n R = \xi\} = \mathscr{P}_k(\rho_n^{-1}\{\xi\}) = \frac{k!(k-1)!}{(k-l)!(n+k-1)!} \lambda_1! \lambda_2! \cdots \lambda_l! \tag{3.19}$$

for $\xi \in \mathscr{E}_n$, $|\xi| = l \leq k$ (a result essentially due to Watterson [9]). A comparison with (2.3) is striking, the normalising constants differing because (3.19) allows ξ with fewer than k equivalence classes. Indeed, if we sum (2.3) and (3.19) over ξ with $|\xi| = k$ and divide the results, we see that

$$\mathbf{P}\{|\rho_n R| = k\} = \frac{n!(n-1)!}{(n+k-1)!(n-k)!} \tag{3.20}$$

under $\mathscr{P}_\kappa$, and that

$$\mathbf{P}\{\mathscr{R}_k = \xi\} = \mathbf{P}\{\rho_n R = \xi \,|\, |\rho_n R| = k\} \tag{3.21}$$

whenever $n \geq k$. Since the right-hand side of (3.20) tends to 1 as $n \to \infty$, it is plausible that $\mathscr{P}_k$ is the correct 'limiting form' for the distribution of $\mathscr{R}_k$.

4. The coalescent

Theorem 3. *There exists a Markov sequence* $\{\mathscr{R}_k; k = 1, 2, \ldots\}$, *where the possible values of* $\mathscr{R}_k$ *are the relations in* $\mathscr{E}$ *with exactly* k *equivalence classes, such that* $\mathscr{R}_k$ *has the distribution* $\mathscr{P}_k$ *and*

$$\mathbf{P}\{\mathscr{R}_{k-1} = \eta \,|\, \mathscr{R}_k = \xi\} = \begin{cases} 2/k(k-1) & \text{if } \xi \prec \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

whenever $\xi \in \mathscr{E}$, $|\xi| = k$. *If* $\{D_t; t > 0\}$ *is a pure death process with death rates* $d_k = \frac{1}{2}k(k-1)$, *satisfying* (2.8) *and independent of* $\{\mathscr{R}_k\}$, *then*

$$R_0 = \Delta, \qquad R_t = \mathscr{R}_{D_t} \quad (t > 0) \tag{4.2}$$

defines a Markov process on $\mathscr{E}$ *for which*

$$\{\rho_n R_t; t \geq 0\} \tag{4.3}$$

is an n*-coalescent for any* $n \in \mathbb{N}$.

Proof. Let $(X_1, X_2, \ldots, X_k)$ be uniformly distributed on the simplex Δ_k. A random point X' in Δ_{k-1} may be constructed by re-arranging the values

$$X_1 + X_2, X_3, \ldots, X_k$$

in random order. A trivial calculation then shows that X' is uniformly distributed over Δ_{k-1}. Apply this result to the non-zero paintbox frequencies X_r (arranged in random order) of a random equivalence relation with distribution $\mathscr{P}_k$. Then the X'_r are the corresponding frequencies for a relation R' obtained from R by combining a randomly chosen pair of equivalence classes, and it follows that R' has distribution $\mathscr{P}_{k-1}$. This consistency property suffices to prove the existence of the Markov process $\{\mathscr{R}_k\}$ with the given properties.

Now define R_t by (4.2) and, for $s > 0$, consider the distributions of $\{R_{s+t}; t \geq 0\}$ conditional on $\{R_u; u \leq s\}$. Since R_{s+t} depends only on D_{s+t} and on $\mathscr{R}_k$ for $k \leq |R_s|$, these conditional distributions depend only on $\{D_{s+t}; t \geq 0\}$ and $\{\mathscr{R}_k; k \leq D_s\}$. By the Markov properties of D and $\mathscr{R}$, these in turn depend only on D_s and $\mathscr{R}_{D_s}$, so that they depend only on R_s. Thus R is a Markov process, and indeed a homogeneous Markov process since there is no dependence on s except through R_s.

For any $i \neq j$, the exchangeability of $\mathscr{R}_k$ and (3.19) imply that

$$\begin{aligned}\mathbf{P}\{(i,j) \in \mathscr{R}_k\} &= \mathbf{P}\{(1,2) \in \mathscr{R}_k\} = \mathbf{P}\{\rho_2 \mathscr{R}_k = \rho_2 \Theta\} \\ &= 2/(k+1).\end{aligned}$$

Hence

$$\mathbf{P}\{(i,j) \in R_t\} = \mathbf{E}\{2/(D_t+1)\} \to 0$$

as $t \downarrow 0$ by (2.8). Since R_t increases with t, this shows that, with probability one, any pair $i \neq j$ satisfies $(i,j) \notin R_t$ for all sufficiently small t, so that

$$\lim_{t \downarrow 0} R_t = \Delta \tag{4.4}$$

in the topology of $\mathscr{E}$.

For $N \in \mathbb{N}$, let the equivalence classes of $\mathscr{R}_N$ be $C_1, C_2, \ldots, C_N$, where the labelling is accomplished in such a way that the smallest element c_r of C_r satisfies $c_1 \leqslant c_2 \leqslant \cdots \leqslant c_N$. This convention ensures that, because of (4.4), for any $n \in \mathbb{N}$,

$$r \in C_r \quad (r = 1, 2, \ldots, n) \tag{4.5}$$

for all sufficiently large N. Define $R_t^{(N)} \in \mathscr{E}_N$ by declaring that $(i,j) \in R_t^{(N)}$ if and only if C_i and C_j lie in the same equivalence class of $R_{T(N)+t}$, where

$$T(N) = \sum_{r=N+1}^{\infty} \tau_r \tag{4.6}$$

is the instant at which D_t enters N. Note that

$$|R_t^{(N)}| = D_{t-T(N)}, \tag{4.7}$$

which is a pure death process starting at N, and that the successive values of $\{R_t^{(N)}; t \geqslant 0\}$ are $\mathscr{R}_k^{(N)}$ $(k = N, N-1, \ldots, 1)$, where $\mathscr{R}_k^{(N)}$ is the relation which $\mathscr{R}_k$ induces on the C_i. From this and (4.1) it follows that $\{\mathscr{R}_k^{(N)}\}$ is a Markov chain, independent of the death process (4.7) and having transition probabilities of the form (4.1). Hence $\{R_t^{(N)}; t \geqslant 0\}$ is an N-coalescent.

Now recall from [8, Section 7] that ρ_{nN} maps N-coalescents into n-coalescents, so that $\{\rho_{nN} R_t^{(N)}\}$ is an n-coalescent. However, (4.5) shows that, for fixed n,

$$\rho_{nN} R_t^{(N)} = \rho_n R_t$$

for all $t > 0$ and all sufficiently large N, so that the joint distributions of $\{\rho_{nN} R_t^{(N)}\}$ converge to those of $(\rho_n R_t)$ as $N \to \infty$. Thus $\{\rho_n R_t\}$ is an n-coalescent, and the theorem is proved.

An $\mathscr{E}$-valued process $\{R_t; t \geqslant 0\}$ for which $\{\rho_n R_t\}$ is an n-coalescent for all n is called a *coalescent*, so that Theorem 3 gives one way of constructing coalescents. There are of course other ways, but the Stone–Weierstrass property shows that they all have the same finite-dimensional distributions [8]. Hence it is legitimate

to talk of *the* coalescent. Any property of (4.2) which is determined by finite-dimensional distributions is true of all coalescents. Actually, most interesting properties also require separability, and are then true of all separable coalescents if they are true of that constructed in Theorem 3; some typical examples are given in the next theorem.

Theorem 4. *Let $\{R_t; t \geq 0\}$ be a separable coalescent. Then $\{R_t\}$ is a Markov process, with*

$$\mathbf{P}\{R_t \in \mathscr{E}_*\} = 1 \tag{4.8}$$

for all $t > 0$, where $\mathscr{E}_*$ consists of all equivalence relations on $\mathbb{N}$ with a finite number of equivalence classes, each of which is infinite. The process

$$D_t = |R_t|$$

is a pure death process, with death rates $\frac{1}{2}k(k-1)$, which satisfies (2.8). *Each sample path of (R_t) runs through a sequence*

$$\cdots \prec \mathscr{R}_k \prec \mathscr{R}_{k-1} \prec \cdots \prec \mathscr{R}_2 \prec \mathscr{R}_1,$$

where $|\mathscr{R}_k| = k$. The sequence $\{\mathscr{R}_k\}$ is independent of $\{D_t\}$, is Markovian and such that $\mathscr{R}_k$ has distribution $\mathscr{P}_k$ and (4.1) *holds. In particular, for $E \subseteq \mathscr{E}$,*

$$\mathbf{P}\{R_t \in E\} = \sum_{k=1}^{\infty} \mathbf{P}\{D_t = k\}\mathscr{P}_k(E). \tag{4.9}$$

Proof. All the statements are true for the particular coalescent constructed in Theorem 3, the only non-trivial one being the infinite character of the equivalence classes, which follows from (3.8). All concern probabilities which can be calculated from the finite-dimensional distributions of $\{R_t\}$, with the aid of separability. Hence they all hold for any separable coalescent.

Note that any equivalence relation in $\mathscr{E}_*$ can be transformed into any other by permuting the elements of $\mathbb{N}$ (so that the group of all $\hat{\pi}$ acts transitively on $\mathscr{E}_*$). Hence (4.8) cannot be strengthened by replacing $\mathscr{E}_*$ by any smaller set which is permutation-invariant. In this sense $\mathscr{E}_*$ is the natural support of the process $(R_t; t > 0)$, though it does not of course contain R_0.

Theorem 4 gives a lot of information about $\{R_t\}$, but does not actually set out its transition function. However, the construction of $\{R_t^{(N)}\}$ in the proof of Theorem 3, which is the 'temporal coupling' of [8], yields this as well. Suppose we require the stochastic structure of $\{R_{s+t}; t \geq 0\}$, given that $R_s = \xi \in \mathscr{E}_*$. The possible values of R_{s+t} are the (finite number of) equivalence relations η with $\xi \subseteq \eta$. Any such η can be described by the relation η/ξ which it induces on the equivalence classes of ξ. Thus the post-s process is described by the values of

$$R_t^{(\geq s)} = R_{s+t}/R_s, \tag{4.10}$$

and the previous argument shows that, if $|R_s| = n$, this is an n-coalescent. Hence the transition function is given by combining (2.3) and (2.5) to evaluate

$$\mathbf{P}\{R_t^{(\geq s)} = \zeta\}$$

for all $\zeta \in \mathscr{E}_n$.

It should be observed that, although the 'minimal' state space $\mathscr{E}_*$ of $\{R_t\}$ is uncountable, the trajectories have rather limited freedom of choice. Once R_t is observed for some positive t, however small, there are only finitely many possible states through which it may subsequently pass. For this reason it is probably inappropriate to subject the coalescent to the 'powerful machinery' [1, p. x] of the theory of continuous-time processes on a general state space. Questions which might be asked of the general theory can be answered more directly using the factorisation of Theorem 4 and the more straightforward theories of the countable-state process $\{D_t\}$ and the discrete-time process $\{\mathscr{R}_k\}$.

5. Another picture of the jump chain

The essential conceptual and analytical difficulties of the coalescent reside in the jump chain $\{\mathscr{R}_k\}$, and it may therefore be helpful to have an alternative picture or model of this process. Let $U_1, U_2, \ldots, V_1, V_2, \ldots$ be independent random variables, each uniformly distributed on the interval $(0, 1)$. With probability one, all their values will be distinct. For $k \in \mathbb{N}$, define a relation R_k on $\mathbb{N}$ to consist of those pairs (i, j) for which either $i = j$ or there is no point V_l $(l \leq k-1)$ in the interval with endpoints U_i and U_j.

It is clear that, with probability one, R_k is an equivalence relation on $\mathbb{N}$ and that $R_k \subseteq R_{k-1}$. Indeed, since every interval of positive length contains some U_i, we have

$$|R_k| = k, \qquad \cdots < R_{k+1} < R_k < R_{k-1} < \cdots < R_1 = \Theta. \tag{5.1}$$

Because the sequence $\{U_i\}$ is exchangeable, R_k is an exchangeable random element of $\mathscr{E}$, and the limits (3.8) are readily identified by the law of large numbers of (U_i): $X_r (1 \leq r \leq k)$ is the rth largest of the k subintervals into which the points $V_1, V_2, \ldots, V_{k-1}$ divide $(0, 1)$, and $X_0 = X_{k+1} = \cdots = 0$.

It is well knowr (cf. [5, Section 2.8]) that $(X_1, X_2, \ldots, X_k)$ is uniformly distributed over the simplex

$$x_1 \geq x_2 \geq \cdots \geq x_k \geq 0, \quad \sum_{r=1}^{k} x_r = 1,$$

so that if rearranged in random order they define a point uniformly distributed over Δ_k. The calculations of Section 3 therefore show that R_k has distribution $\mathscr{P}_k$.

Now consider the conditional distribution of R_{k-1} given $R_k, R_{k+1}, \ldots$. A knowledge of R_k determines the lengths of the subintervals $X_1, X_2, \ldots, X_k$ but not their order; to this the $R_l (l \geq k+1)$ add only information about the way $V_k, V_{k+1}, \ldots$

fall, which is of no predictive value for R_{k-1}. Hence the conditional distribution depends only on R_k, and symmetry considerations show that any one of the $\frac{1}{2}k(k-1)$ relations $\xi \in \mathscr{E}$ with $\xi \succ R_k$ is as probable as any other.

Thus $\{R_k\}$ is a Markov sequence with the same absolute distributions $\mathscr{P}_k$ and the same one-step-backward transition probability as $\{\mathscr{R}_k\}$, and therefore the two sequences have the same joint distributions.

One may think of this construction in terms of a rectangular paintbox with base (0, 1) and vertical partitions rising from the points V_k. These partitions are removed one by one in descending order of k, and each removal allows the two colours hitherto separated to mix to form a new colour. The U_i are the points at which a brush is dipped into the paintbox in order to paint an infinite collection of balls, and $\mathscr{R}_k$ is the equivalence relation so induced from the colours left after V_k has been removed.

References

[1] K.L. Chung, Markov Chains with Stationary Transition Probabilities (Springer, Berlin, 2nd ed. 1967).
[2] J.L. Doob, Stochastic Processes (Wiley, New York, 1953).
[3] M. Iosifescu, Finite Markov Processes and their Applications (Wiley, New York, 1980).
[4] D.G. Kendall, An introduction to stochastic analysis, D.G. Kendall and E.F. Harding, eds., Stochastic Analysis (Wiley, New York, 1973).
[5] M.G. Kendall and P.A.P. Moran, Geometrical Probability (Griffin, 1963).
[6] J.F.C. Kingman, The population structure associated with the Ewens sampling formula, Theory Pop. Biol. 11 (1977) 274–283.
[7] J.F.C. Kingman, The representation of partition structures, J. London Math. Soc. 18 (1978) 374–380.
[8] J.F.C. Kingman, On the genealogy of large populations, to appear.
[9] G.A. Watterson, The stationary distribution of the infinitely-many neutral alleles diffusion model, J. Appl. Probab. 13 (1976) 639–651.