On the Genealogy of Large Populations

J. F. C. KINGMAN

Abstract

A new Markov chain is introduced which can be used to describe the family relationships among n individuals drawn from a particular generation of a large haploid population. The properties of this process can be studied, simultaneously for all n, by coupling techniques. Recent results in neutral mutation theory are seen as consequences of the genealogy described by the chain.

WRIGHT–FISHER MODEL; NEUTRAL MUTATION; RANDOM EQUIVALENCE RELATIONS; COALESCENT; EWENS SAMPLING FORMULA; COUPLING; ENTRANCE BOUNDARY

1. Introduction

In modelling the way in which some characteristic changes with time in a biological population it is often necessary to take account of correlations between individuals caused by kinship. These can be purely genetical, but they need not be; in a spatially distributed population for instance there may be a tendency for the more closely related members to cluster together. Such correlations are never easy to handle, and have given rise to interesting developments in spatial branching processes (e.g. Fleischmann and Siegmund-Schultze (1978)), in point processes (Kerstan, Matthes and Mecke (1978), Felsenstein (1975), Sawyer (1977)), and in Malécot's elaboration of 'identity by descent' (Malécot (1969)).

Another example, in which the role of kinship is particularly clear, is the Ohta–Kimura (1973) model of neutral mutation. Professor Moran (1975) has himself made a striking contribution to its analysis, and more recent progress is summarised in Kingman (1980), but special reference should be made to a paper of Kesten (1980), in which explicit recognition of the genealogy† leads to very deep and surprising results.

† Note that 'genealogy' means the whole family tree structure, and not just the sizes of related families as in Kendall (1975).

 0021-9002/82/050027-17 $01.95

My purpose here is to suggest that, by bringing the genealogy into the open, the complex problems implied by family relationships even in simple models can be seen in perspective and more readily attacked. Indeed, a simple process will be shown to be a general and robust approximation when the population is large and the characters of interest uncorrelated with fitness or fertility.

One crucial limitation of the present analysis is that it applies only to haploid models: an individual has just one parent. To apply the theory to a diploid population, one must consider only individuals of one sex, or (much more usefully in genetical problems) work with the haploid gametes which carry genetic material from one generation to the next. It would be of great interest to seek a comparable analysis of truly diploid genealogy (cf. Wachter et al. (1978), Chapter 9).

The basic strategy (which has affinities with Kallenberg's (1977) idea of backward trees) is to select n individuals from a particular generation, and to trace back their descent, noting when there are common ancestors. Typically there will come a time (after going back a number of generations comparable with the population size) when all n have a common ancestor. We seek to describe the family tree over this period, and when the population is large this can often be done in terms of a finite Markov chain, the *n-coalescent.* This is a single process, free of parameters except for n, an interesting object of study in its own right, whose properties explain and draw together some of the results of more specialised studies.

2. The Wright–Fisher model and the n-coalescent

To a population geneticist the branching process models so beloved of the probabilist fail to carry conviction because, in biological reality, the total population size is more often determined by external factors like availability of food or living space, or the action of predators, than by summing independent family sizes. His first approximation to reality is therefore a model in which the total population size is a fixed number dictated by external constraints, and the most popular is that associated with the names of Sewall Wright and R. A. Fisher.[†]

This assumes discrete, non-overlapping generations $G_0, G_1, G_2, \cdots$ in which each generation contains a fixed number N of individuals. (In most genetical applications these are the successful gametes in the reproductive process, so that N is twice the 'effective population size', $N = 2N_e$.) Each member of G_{r+1} is the child of exactly one member of G_r, but the number of children born to the jth member of G_r is a random variable ν_j, subject of course to the constraint

[†] The book by Ewens (1979) discusses these issues and their history in a manner accessible to the mathematician.

$$\sum_{j=1}^{N} \nu_j = N. \tag{2.1}$$

In the neutral Wright–Fisher model the ν_j are assumed to have a symmetric multinomial distribution:

$$\mathbb{P}\{\nu_j = n_j \ (j = 1, 2, \cdots, N)\} = N!/n_1!\, n_2! \cdots n_N!\, N^N. \tag{2.2}$$

This process is particularly tractable because it has a simple 'backwards' structure: (2.2) is equivalent to the prescription that each member of G_{r+1} chooses its parent at random, independently and uniformly from the N individuals of G_r.

Consider two particular members of G_r. They have the same parent with probability N^{-1}, and different parents with probability $1 - N^{-1}$. The probability that they have distinct parents but the same grandparent is $(1 - N^{-1})N^{-1}$, and so on; the probability that they have distinct ancestors in G_{r-s} is $(1 - N^{-1})^s$.

A much more difficult problem is to compute the probability $\gamma(N, s)$ that the whole of G_r has a common ancestor in G_{r-s}. Since $1 - \gamma(N, s)$ is the probability of the union, over $i < j$, of the event that the ith and jth members of G_r have distinct ancestors in G_{r-s}, we have the inequalities

$$(1 - N^{-1})^s \leqq 1 - \gamma(N, s) \leqq \tfrac{1}{2}N(N-1)(1 - N^{-1})^s. \tag{2.3}$$

The upper bound here is very crude, and can be reduced to one which differs from the lower bound by a factor of 3:

$$1 - \gamma(N, s) \leqq 3((N-1)/(N+1))(1 - N^{-1})^s \leqq 3(1 - N^{-1})^s. \tag{2.4}$$

The constant 3 is best possible (Kingman (1980)).

These bounds show that, when N is large, the number of generations which it is necessary to count back before the whole of G_r has a single common ancestor is of the order of N, and this defines a natural time unit of N generations. Notice that the backwards description allows us to extend the whole process to negative values of r, and we shall therefore suppose that G_r (and the family ties between G_r and G_{r+1}) are defined for both positive and negative integers r.

Now fix r, and select $n(\leqq N)$ particular individuals $\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_n$ from G_r. The family tree of these and their ancestors may be described by means of a sequence of equivalence relations† $\mathcal{R}_s$ $(s = 0, 1, 2, \cdots)$ on the set $\{1, 2, \cdots, n\}$, where $\mathcal{R}_s$ contains a pair (i, j) if $\mathcal{I}_i$ and $\mathcal{I}_j$ have common ancestor in G_{r-s}. Note that

$$\mathcal{R}_0 = \Delta = \{(i, i);\ i = 1, 2, \cdots, n\}, \tag{2.5}$$

† Equivalence relations (apart from the special cases Δ and Θ) will be denoted either by the letter R in some fount, or by Greek lower case ξ or η, with affixes as necessary.

that

$$\mathcal{R}_s \subseteq \mathcal{R}_{s+1}, \tag{2.6}$$

and that $\gamma(N, s)$ is the probability, in the special case $n = N$, that $\mathcal{R}_s$ is the relation

$$\Theta = \{(i, j);\ i, j = 1, 2, \cdots, n\}. \tag{2.7}$$

Each equivalence class of $\mathcal{R}_s$ corresponds to a member of G_{r-s} (but not conversely). Two such members may choose the same parents in G_{r-s-1}, in which case the corresponding equivalence classes of $\mathcal{R}_s$ are combined in $\mathcal{R}_{s+1}$, or different parents, in which case the classes are not so combined. It follows that the sequence $(\mathcal{R}_s)$ forms a (discrete-time) Markov chain, whose state space is the finite set $\mathcal{E}_n$ of equivalence relations on $\{1, 2, \cdots, n\}$, and whose transition probabilities

$$p_{\xi\eta} = \mathbb{P}\{\mathcal{R}_{s+1} = \eta \mid \mathcal{R}_s = \xi\} \tag{2.8}$$

are calculated as follows. By (2.6), $p_{\xi\eta} = 0$ unless $\xi \subseteq \eta$. If $\xi \subseteq \eta$ we can label the equivalence classes of η as C_α $(\alpha = 1, 2, \cdots, a)$ and those of ξ as $C_{\alpha\beta}$ $(\alpha = 1, 2, \cdots, a;\ \beta = 1, 2, \cdots, b_\alpha)$, where

$$C_\alpha = \bigcup_{\beta=1}^{b_\alpha} C_{\alpha\beta}.$$

Then $p_{\xi\eta}$ is the probability that, if balls $C_{\alpha\beta}$ are placed at random (independently and uniformly) into N boxes, then for each α all the $C_{\alpha\beta}$ fall into the same box, the boxes for different α being distinct. This is a combinatorial quantity, depending only on $N, a, b_1, b_2, \cdots, b_a$.

There is no need to compute the $p_{\xi\eta}$ in detail, since for present purposes it is enough to record that $p_{\xi\eta}$ is of order N^{-2} unless either $\xi = \eta$, or η is derived from ξ by amalgamating two of its equivalence classes (in which case we write $\xi \prec \eta$). In fact,

$$p_{\xi\eta} = \delta_{\xi\eta} + q_{\xi\eta} N^{-1} + O(N^{-2}), \tag{2.9}$$

where $\delta_{\xi\eta}$ is the Kronecker delta, and

$$q_{\xi\eta} = \begin{cases} -\frac{1}{2}k(k-1) & \text{if } \xi = \eta \text{ and } k = |\xi| \\ 1 & \text{if } \xi \prec \eta \\ 0 & \text{otherwise,} \end{cases} \tag{2.10}$$

$|\xi|$ denoting the number of equivalence classes of ξ.

In matrix notation, the stochastic matrix

$$P_N = (p_{\xi\eta};\ \xi, \eta \in \mathcal{E}_n)$$

satisfies

$$P_N = I + N^{-1}Q + O(N^{-2}) \tag{2.11}$$

as $N \to \infty$. Stochastic matrices are contraction operators for the norm

$$\|A\| = \max_{\xi} \sum_{\eta} |a_{\xi\eta}|, \tag{2.12}$$

and contraction operators satisfy the inequality (well known, and easily proved by induction on r)

$$\|A_1 A_2 \cdots A_r - B_1 B_2 \cdots B_r\| \leqq \sum_{s=1}^{r} \|A_s - B_s\|. \tag{2.13}$$

Taking $A_s = P_N$, $B_s = \exp(N^{-1}Q)$ (B_s is stochastic because Q is a 'Q-matrix' with positive off-diagonal elements and zero row sums) and r as the integer part of Nt for fixed $t > 0$, we have

$$\|P_N^{[Nt]} - \exp(N^{-1}[Nt]Q)\| \leqq [Nt]\|P_N - \exp(N^{-1}Q)\| = O(N^{-1})$$

by (2.11), so that

$$\lim_{N\to\infty} P_N^{[Nt]} = \exp(tQ). \tag{2.14}$$

In other words, if

$$R_t = \mathscr{R}_{[Nt]} \tag{2.15}$$

denotes the process on $\mathscr{E}_n$ in the natural time scale, it converges in distribution as $N \to \infty$ to the (continuous-time) Markov chain with infinitesimal generator Q. This latter process we call the *n-coalescent*, and its study will concern us from Section 5 onwards. Before embarking on this, we indicate very briefly the way in which it relates to previous work on the neutral mutation problem, and the fact that it arises from other models than the Wright–Fisher.

The assumption that N is constant is not essential. If N varies because of external factors, we simply assign the step from G_r to G_{r+1} a time interval equal to the local value of N^{-1} (cf. Kingman (1978c)). The argument then proceeds, using (2.13) to reach the same conclusion for the $\mathscr{R}$-process observed in this natural, but now non-linear, time scale.

3. Neutral mutation

The neutral Wright–Fisher model has been used by Ohta and Kimura (1973), Moran (1975) and others to assess the amount of genetic variability to be expected at a locus at which selection is not operating, the dominant forces being mutation which increases variability and 'genetic drift' (the randomness of the reproduction mechanism) which tends to reduce it. Almost all the analysis has

been for large populations, and has involved either explicitly or implicitly (by a diffusion approximation) a limit as $N \to \infty$. The results of this limiting operation can be expressed directly in terms of the n-coalescent.

Consider for definiteness the generalized Ohta–Kimura model of Kingman (1976). Here the alleles at a locus are described (doubtless with some loss of discrimination, cf. Singh, Lewontin and Felton (1976)) by values of a single numerical measurement, so that the genetical structure of the rth generation is described by a collection $X_j(r)$ $(j = 1, 2, \cdots, N)$ of real random variables. If the jth member of G_r is the child of the αth member of G_{r-1}, it is assumed that†

$$X_j(r) = X_\alpha(r-1) + Y_{rj}, \tag{3.1}$$

where the effect of mutation is to displace by an amount Y_{rj}, assumed to be independent of other Y, and to have a known characteristic function

$$\tau_N(u) = \mathbb{E}(e^{iuY_{rj}}). \tag{3.2}$$

The mutation probability $\mathbb{P}(Y_{rj} \neq 0)$ is assumed to be of order N^{-1} (so as to achieve a balance between mutation and drift), and the mathematician interprets this as meaning that

$$\phi(u) = \lim_{N\to\infty} N\{1 - \tau_N(u)\} \tag{3.3}$$

exists as a finite limit. The main result, established by direct calculation, is that in the limit as $N \to \infty$ the joint characteristic function

$$\psi_n(u_1, u_2, \cdots, u_n) = \mathbb{E}\left\{\exp\left(i\sum_{j=1}^{n} u_j X_j\right)\right\} \tag{3.4}$$

of n typical individuals in a generation is determined, when

$$\sum_{j=1}^{n} u_j = 0, \tag{3.5}$$

recursively on n, by the equation

$$\begin{aligned}\left\{\tfrac{1}{2}n(n-1) + \sum_{j=1}^{n}\phi(u_j)\right\}\psi_n(u_1, u_2, \cdots, u_n) \\ = \sum_{1\le j<k\le n} \psi_{n-1}(u_1, \cdots, u_{j-1}, u_j + u_k, u_{j+1}, \cdots, u_{k-1}, u_{k+1}, \cdots, u_n).\end{aligned} \tag{3.6}$$

The condition (3.5) is needed because it is only the differences $X_j - X_k$ which have a genuine statistical equilibrium (see Kingman (1980) for an alternative model without this feature). However, (3.6) is enough to determine the joint

† The same model would serve for a population distributed in a linear habitat, in which the individual moves randomly from the position of its parent (Felsenstein (1975)).

distribution of these differences, and so to evaluate quantities of biological significance.

Equation (3.6) is exactly what one would expect from the theory of the n-coalescent. The n selected individuals have a family tree which, measured in units of N generations, converges to the n-coalescent as $N \to \infty$. In any line of descent, mutations occur at an average rate of order N^{-1} per generation, and therefore at unit rate in the natural time scale. In $[Nt]$ generations, the total effect of the mutations in a single line has characteristic function

$$\tau_N(u)^{[Nt]} \to \exp\{-t\phi(u)\}$$

as $N \to \infty$ because of (3.3).

This means that we can formulate a model by using the n-coalescent to describe the genealogy of the n individuals, and the function ϕ to describe the effect of mutation. The latter defines an additive process with independent increments, such that the increment on a time interval t has characteristic function $\exp\{-t\phi(u)\}$. Let $(R_t; t \geqq 0)$ be an n-coalescent and imagine that, at any $t \geqq 0$, each equivalence class of R_t is located at a point of the line. Conditional on R, these points vary with t in a random way; when two classes coalesce they are at the same place, and between transitions of R they perform independent versions of the additive process. It can be checked (by running time backwards) that these conditions can be met, and then the X_j in (3.4) are just the n points of R_0.

If now we analyse R_0 by a backward Kolmogorov equation argument, examining the possible transitions in the time interval $(0, h)$ as $h \to 0$, we arrive at (3.6) as an *exact* consequence of the model just described. The analysis works almost without change when the X_j take values in some space of higher dimension, or when the mutation structure is quite general (Kingman (1980)).

Of course, the algebra implied by (3.6) is very complicated, and it is not easy to derive results in useful form. Ironically the feature of the Ohta–Kimura model which makes it difficult is also that which makes it notably unrealistic. In its original form, in which

$$\mathbb{P}(Y_{rj} = 1) = \mathbb{P}(Y_{rj} = -1) = \tfrac{1}{2}, \tag{3.7}$$

it implies that two successive mutations will with probability $\frac{1}{2}$ cancel each other out and restore the original allele. In fact, such returns to the origin probably occur most infrequently, and a model at the other extreme, ignoring recurrent mutation, is more plausible as well as mathematically simpler.

Thus suppose that mutation always produces a completely novel allele. Two individuals in G_r have the same allele only if no mutation has occurred in the line of descent of either from their common ancestor. If s generations have elapsed since their last common ancestor (an event with probability $N^{-1}(1 - N^{-1})^{s-1}$) the

conditional probability that they are identical is $(1-\beta)^{2s}$, where β is the probability of mutation affecting a particular birth. Hence the unconditional probability is

$$\sum_{s=1}^{\infty} N^{-1}(1-N^{-1})^{s-1}(1-\beta)^{2s} = (1-\beta)^2\{1+2(N-1)\beta\}^{-1},$$

and this converges to $(1+\theta)^{-1}$ as $N\to\infty$, $\beta\to 0$, $2N\beta\to\theta$.

More generally, the results of this limit can be formulated directly using the n-coalescent. Suppose that mutations occur in a Poisson way at a rate $\frac{1}{2}\theta$, independently to the equivalence classes of R_t. We can then define a relation $\mathcal{R}\in\mathscr{E}_n$ which contains (i,j) if, on watching the equivalence classes of R_t containing i and j until the moment at which they coincide, we observe no mutation to either.

A backward Kolmogorov argument like that leading to (3.6) then gives the distribution of $\mathcal{R}$ in the explicit form

$$\mathbb{P}\{\mathcal{R}=\xi\} = \frac{\theta^{k-1}}{(\theta+1)(\theta+2)\cdots(\theta+n-1)}\prod_{\alpha=1}^{k}(\lambda_\alpha-1)! \quad (\xi\in\mathscr{E}_n), \tag{3.8}$$

where $\lambda_1,\lambda_2,\cdots,\lambda_k$ are the sizes of the equivalence classes of ξ.

If (3.8) is multiplied by the number of $\xi\in\mathscr{E}_n$ with the given $\lambda_1,\lambda_2,\cdots,\lambda_k$, we obtain the distribution of the allelic partition of the sample of size n, and this agrees with the celebrated *Ewens sampling formula* (Ewens (1972), Kingman (1978a)). This has been derived many times, but always as an approximation or a limit. It is therefore of some interest that the Ewens formula is an exact consequence of mutation in the n-coalescent.

4. Robustness

Although the n-coalescent arises most easily from the Wright–Fisher formulation, there are many other models which exhibit the same limit for large population size. Suppose for instance that the symmetric multinomial (2.2) is replaced by some other exchangeable joint distribution for the ν_j (the exchangeability implying that we need not label the members of a generation in any particular way), and suppose that the ν_j are independent of the corresponding family sizes in other generations. Then the relations $\mathcal{R}_s$ again form a Markov chain, whose transition probabilities are of the form

$$p_{\xi\eta} = (N)_k^{-1}\mathbb{E}\left\{\sum(\nu_{j_1})_{b_1}(\nu_{j_2})_{b_2}\cdots(\nu_{j_a})_{b_a}\right\}, \tag{4.1}$$

where $a, b_1, b_2, \cdots, b_a$ are determined by ξ and η as in Section 2,

$$k=\sum_{\alpha=1}^{a} b_\alpha, \qquad (N)_k = N(N-1)\cdots(N-k+1),$$

and the summation extends over distinct $j_1, j_2, \cdots, j_a$ in $1 \leqq j \leqq N$.

If for example ξ has just two equivalence classes, and $\eta = \Theta$ only one, then

$$p_{\xi\eta} = N^{-1}(N-1)^{-1}\mathbb{E}\left\{\sum_{j=1}^{N} \nu_j(\nu_j - 1)\right\} = (N-1)^{-1}\mathbb{E}\{\nu_1(\nu_1 - 1)\}$$

$$= (N-1)^{-1}\mathrm{Var}(\nu_1)$$

since, by (2.1), $\mathbb{E}(\nu_1) = 1$. Now assume that, as $N \to \infty$, the variance of a typical family size ν_1 tends to a finite non-zero limit σ^2, and that the moments of ν_1 are bounded:

$$\mathrm{Var}(\nu_1) \to \sigma^2, \qquad \mathbb{E}(\nu_1^m) \leqq M_m \ (m = 1, 2, \cdots). \tag{4.2}$$

Then (4.1) leads, after some algebra and an application of Hölder's inequality, to the conclusion that, for $\xi \neq \eta$

$$p_{\xi\eta} = q_{\xi\eta}\sigma^2 N^{-1} + O(N^{-2}), \tag{4.3}$$

where $q_{\xi\eta}$ is given by (2.10). As in Section 2, this implies that the continuous-time process

$$R_t = \mathcal{R}_{[N\sigma^{-2}t]} \tag{4.4}$$

converges in distribution to the n-coalescent. Notice the way in which the family-size variance affects the natural time scale.

A degenerate example is the Moran process, which has

$$\mathbb{P}(\nu_1 = 2) = N^{-1}, \quad \mathbb{P}(\nu_1 = 0) = N^{-1}, \quad \mathbb{P}(\nu_1 = 1) = 1 - 2N^{-1}, \tag{4.5}$$

so that $\mathrm{Var}(\nu_1) = 2N^{-1}$. For this one must use the different time change $r = \frac{1}{2}N^2 t$ to achieve the n-coalescent in the limit. In the Moran (1958) formulation mutation only operates on the newly-born individual, and this difference from the Wright–Fisher model cancels out the different time change, except for the factor of 2 which always separates the conclusions of the two models.

Not all models have the ν_j exchangeable, or independent between generations, but if a reproductive mechanism makes the sequence $(\mathcal{R}_s)$ Markovian (or approximately so to order $o(N^{-1})$) it is worth checking its transition probabilities for large N. If they satisfy (4.3) (perhaps with $O(N^{-2})$ replaced by $o(N^{-1})$), then the n-coalescent will describe the limiting genealogy of n contemporaries, and conclusions like those of Section 3 remain valid.

This whole question of the robustness of the n-coalescent requires further study. A preliminary conclusion is that it represents a good approximation when the population is large and the individuals have no inherited differences of fertility or fitness. Selective advantages and disadvantages (cf. Kingman (1980)) lead to very different behaviour, and the n-coalescent is essentially a 'neutral' genealogy.

5. The transit time

The rest of this paper will be devoted to the properties of the n-coalescent itself. This is a Markov chain in continuous time whose state space is the finite set $\mathscr{E}_n$ of all equivalence relations on the set $\{1, 2, \cdots, n\}$. The chain starts in the state Δ, and has Θ as its only absorbing state, where Δ and Θ are respectively the finest and coarsest relations in $\mathscr{E}_n$. The transition rates $q_{\xi\eta}$ are given by (2.10), and define the chain completely, their form implying that the transitions are of the form $\xi \to \eta$, where $\xi \prec \eta$.

If f is a function from $\mathscr{E}_n$ onto another (finite) set $\mathscr{S}$, the transformed process $f(R_t)$ is not in general Markovian. However, it is a Markov chain (Rosenblatt (1974), §IIId) if, for $\xi \in \mathscr{E}_n$, $v \in \mathscr{S}$, $f(\xi) \neq v$, the sum

$$\sum_{f(\eta)=v} q_{\xi\eta} \tag{5.1}$$

depends on ξ only through $u = f(\xi)$.

For the n-coalescent, this is the case for two interesting functions f. The first, which will not be exploited here, is that which assigns to the relation R the induced partition of n: in the usual notation

$$f(R) = 1^{a_1} 2^{a_2} \cdots n^{a_n} \tag{5.2}$$

where, for each r, a_r is the number of equivalence classes of R having exactly r elements. Thus f destroys the labelling of $\{1, 2, \cdots, n\}$, and the fact that the Markov property is preserved follows by a simple permutation argument.

A more radical lumping of the states of (R_t) is defined by the function

$$f(R) = |R|, \tag{5.3}$$

the number of equivalence classes of R. Since $\xi \prec \eta$ implies that $|\xi| = |\eta| + 1$, the sum (5.1) is 0 unless $u = f(\xi) = v + 1$, when it takes the value $\frac{1}{2}u(u-1)$. Thus $(|R_t|)$ is a very simple Markov chain, a pure death process with initial state n, in which the transition rate from r to $r-1$ is given by

$$d_r = \tfrac{1}{2}r(r-1). \tag{5.4}$$

The *transit time* for the n-coalescent (R_t) is the time T_n at which it first reaches its absorbing state Θ, and this is the time at which the death process $(|R_t|)$ reaches its absorbing state 1. Thus

$$T_n = \sum_{r=2}^{n} \tau_r, \tag{5.5}$$

where the τ_r are independent, with

$$\mathbb{P}\{\tau_r \leqq t\} = 1 - e^{-d_r t}, \quad \mathbb{E}(\tau_r) = d_r^{-1}. \tag{5.6}$$

From (5.4) and (5.5),

$$\mathbb{E}(T_n) = 2 - 2n^{-1}, \tag{5.7}$$

which is bounded as n increases. In fact, if τ_r is defined for all r, then

$$T_n < T = \sum_{r=2}^{\infty} \tau_r, \tag{5.8}$$

where the series converges with probability 1 because T has finite expectation 2.

It is easy to calculate the distribution of T, since for $\theta > 0$

$$\mathbb{E}(e^{-\theta T}) = \prod_{r=2}^{\infty} \left(\frac{d_r}{d_r + \theta} \right) = 2\pi\theta \sec\{\tfrac{1}{2}\pi(1 - 8\theta)^{\frac{1}{2}}\}$$

$$= \sum_{m=2}^{\infty} (-1)^m (2m-1) \frac{d_m}{d_m + \theta},$$

which shows that T has probability density

$$g(t) = \sum_{m=2}^{\infty} (-1)^m \tfrac{1}{2} m(m-1)(2m-1) e^{-\frac{1}{2}m(m-1)t}. \tag{5.9}$$

By (5.8),

$$\mathbb{P}\{R_t \neq \Theta\} = \mathbb{P}\{T_n > t\} \leqq \sum_{m=2}^{\infty} (-1)^m (2m-1) e^{-\frac{1}{2}m(m-1)t}, \tag{5.10}$$

which is a powerful upper bound, uniform in n, for the probability that R has not reached its absorbing state by time t.

The right-hand side of (5.10) is asymptotically $3e^{-t}$ for large t, and this constant 3 is the same as that we have already met in (2.4). The proof of (2.4) depended (though this was not admitted in Kingman (1976)) on identifying a supermartingale for the Wright–Fisher process, and this becomes a martingale for the n-coalescent. If (X_t) is a pure death process whose death rates d_r satisfy $\Sigma d_r^{-1} < \infty$, and if

$$\phi_\theta(x) = \prod_{r=x+1}^{\infty} (1 - d_r^{-1}\theta),$$

then $\phi_\theta(X_t)e^{\theta t}$ is a martingale. In the present context the most interesting value of θ is 1, when $\phi_1(x) = (x-1)/(x+1)$. Thus

$$\frac{|R_t| - 1}{|R_t| + 1} e^t \tag{5.11}$$

is a martingale, and in particular

$$\mathbb{E}\left\{ \frac{|R_t| - 1}{|R_t| + 1} \right\} = \frac{n-1}{n+1} e^{-t}. \tag{5.12}$$

Since $(x-1)/(x+1) \geqq \frac{1}{3}$ when $x \geqq 2$, this shows that

$$\mathbb{P}\{R_t \neq \Theta\} = \mathbb{P}\{|R_t| \geqq 2\} \leqq 3\,\frac{n-1}{n+1}\,e^{-t}. \tag{5.13}$$

This complements (5.10); the one is asymptotically sharp as $n \to \infty$, the other as $t \to \infty$.

It will be observed that we have been able to obtain properties of the n-coalescent for all values of n by implicitly thinking of the random variables T_n as being defined on the same probability space. This is an example of the fashionable technique of *coupling*, and one can ask whether it can be exploited systematically by constructing a probability space on which n-coalescents can be defined for all n. There are (at least) two quite different useful ways of doing this. The first is just an elaboration of the arguments used in this section, and is described in Section 6. In Section 7 the second, which fits in more naturally with the genealogical interpretation, is defined and exploited.

6. Temporal coupling

Let $(D_t; t > 0)$ be a pure death process, with death rates (5.4), starting from ∞.† Such a process can be constructed directly from independent random variables τ_r $(r = 2, 3, \cdots)$ satisfying (5.6) by the recipe

$$E_m = T - T_n = \sum_{r=n+1}^{\infty} \tau_r,$$

$$D_n = \begin{cases} n & \text{if } E_n \leqq t < E_{n-1} \quad (n \geqq 2) \\ 1 & \text{if } E_1 \leqq t. \end{cases} \tag{6.1}$$

For the n-coalescent, the Markov chain $(|R_t|; t \geqq 0)$ has the same joint distributions as the chain

$$D(n, t) = D(E_n + t). \tag{6.2}$$

The problem is to enlarge this construction by producing a model, not just for the integer-valued random variables $|R_t|$, but for the $\mathscr{E}_n$-valued variables R_t. This can be done by taking a probability space on which can be defined, independently of the τ_r, a binary fission process (a continuous-time branching process in which, in a time interval $(t, t+\delta t)$, any individual present has a probability $\delta t + O(\delta t)$ of splitting into two new individuals) with just one individual at $t = 0$.

For any $n \geqq 2$, there are exactly n individuals present just before the nth

† This is one of the very few examples of the *application* of the boundary theory of Markov chains, and the only one I know involving an entrance boundary.

splitting. Label these in random order as $\mathscr{I}_1, \mathscr{I}_2, \cdots, \mathscr{I}_n$, and define (for $1 \leqq m \leqq n$) an equivalence relation $\mathscr{R}_m^{(n)}$ on $\{1, 2, \cdots, n\}$ by specifying that $(i, j) \in \mathscr{R}_m^{(n)}$ if $\mathscr{I}_i$ and $\mathscr{I}_j$ have a common ancestor among the m individuals present just before the mth splitting. It is clear that

$$\Delta = \mathscr{R}_n^{(n)} < \mathscr{R}_{n-1}^{(n)} < \cdots < \mathscr{R}_1^{(n)} = \Theta, \quad |\mathscr{R}_m^{(n)}| = m. \tag{6.3}$$

Direct computation of

$$\mathbb{P}\{\mathscr{R}_m^{(n)} = \xi_m \ (m = n, n-1, \cdots, 1)\}$$

for any $\xi_m \in \mathscr{E}_n$ with

$$\Delta = \xi_n < \xi_{n-1} < \cdots < \xi_1 = \Theta \tag{6.4}$$

shows that this probability does not depend on the ξ-sequence, and from this it follows that

$$R_t^{(n)} = \mathscr{R}_{D(n,t)}^{(n)} \tag{6.5}$$

is a (continuous-time) Markov chain with transition rates (2.10). Thus $R^{(n)}$ is an n-coalescent, with

$$|R_t^{(n)}| = D(n, t). \tag{6.6}$$

In other words, the jump chain of the n-coalescent, which describes the sequence of distinct states through which it passes, is independent of the death process which governs the timing of the jumps. A model of this jump chain can be derived from a binary fission process by observing it backwards from the time of the nth splitting.

In this sense the backward genealogy of the binary fission differs from the backward genealogy of the Wright–Fisher model and its relatives in the limit of large population size only in its time scale. This makes precise a percipient remark by Kesten (1980).

7. The natural coupling

The construction of Section 6 represents n-coalescents for all values of n on a single probability space by combining a pure death process with an independent binary fission. It depends on that fact that, for $m < n$, an m-coalescent can be found in an n-coalescent by waiting until the latter has only m equivalence classes and then looking at the coalescence of these m classes, regarded as individuals.

This is not however the most natural way of finding an m-coalescent in an n-coalescent. After all, the n-coalescent was derived by taking n particular members $\mathscr{I}_1, \mathscr{I}_2, \cdots, \mathscr{I}_n$ of a particular generation of a Wright–Fisher population, and examining the ancestry of these individuals as $N \to \infty$. The obvious way

of deriving an m-coalescent is to discard $\mathscr{I}_{m+1}, \cdots, \mathscr{I}_n$ and to look just at the ancestry of $\mathscr{I}_1, \mathscr{I}_2, \cdots, \mathscr{I}_m$.

Formally, define a function $\rho_{mn} : \mathscr{E}_n \to \mathscr{E}_m$ $(m < n)$ by restriction: for $\xi \in \mathscr{E}_n$,

$$\rho_{mn}\xi = \{(i,j); 1 \leqq i,j \leqq m, (i,j) \in \xi\}. \tag{7.1}$$

It can then be checked directly, or deduced from Section 2, that if (R_t) is an n-coalescent then $(\rho_{mn}R_t)$ is an m-coalescent.

Thus, on the probability space of an n-coalescent, we can construct m-coalescents for all m in $2 \leqq m \leqq n$. Can we, in a similar way, construct n-coalescents for all n on a single probability space? There is a general 'projective limit' technique which answers such questions, but as in Section 6 it is better to proceed concretely.

Let $\mathscr{E}$ be the (uncountable) set of equivalence relations on the set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of natural numbers. For $\xi \in \mathscr{E}$ and $n \geqq 2$, denote by $\rho_n\xi$ the restriction of ξ to $\{1, 2, \cdots, n\}$ (as in (7.1)), so that ρ_n maps $\mathscr{E}$ into $\mathscr{E}_n$ and

$$\rho_{mn}(\rho_n\xi) = \rho_m\xi \qquad (m < n,\ \xi \in \mathscr{E}). \tag{7.2}$$

We seek to construct a random process $(R_t; t \geqq 0)$ with values in $\mathscr{E}$ such that, for all $n \geqq 2$, $(\rho_n R_t; t \geqq 0)$ is an n-coalescent.

The existence of such a process is most easily proved by the topological Kakutani–Nelson technique.† The set $\mathscr{E}$ can be regarded as a subset of the product space $2^{\mathbb{N}\times\mathbb{N}}$, whose product topology is compact and has $\mathscr{E}$ as a closed set. In the subspace topology $\mathscr{E}$ is therefore compact Hausdorff, and the $\mathscr{E}$-valued process (R_t) exists if its finite-dimensional distributions are specified in a consistent way. Thus we have to specify

$$\mathbb{E}\{\Phi(R_{t_1}, R_{t_2}, \cdots, R_{t_k})\} \tag{7.3}$$

for $0 \leqq t_1 < t_2 < \cdots < t_k$ and bounded continuous functions $\Phi : \mathscr{E}^k \to \mathbb{R}$. The requirement that $(\rho_n R_t)$ be an n-coalescent fixes the value of (7.3) whenever Φ is of the form

$$\Phi(\xi_1, \xi_2, \cdots, \xi_k) = \phi(\rho_n\xi_1, \rho_n\xi_2, \cdots, \rho_n\xi_k), \tag{7.4}$$

and (7.2) establishes consistency between different values of n. The Stone–Weierstrass theorem shows that the functions Φ of the form (7.4) for some n form a uniformly dense subset of the set of all bounded continuous Φ, so that the consistent specification of the finite-dimensional distributions is achieved.

A process (R_t), with values in $\mathscr{E}$, for which each restriction $(\rho_n R_t)$ is an n-coalescent, will be called simply a *coalescent*. The argument just sketched

† In the words of Meyer ((1966), p. 319) '... familière aux lecteurs de Bourbaki, a été popularisée aux Etats-Unis par un article de Nelson (1959)'.

shows that coalescents exist, and that they all have the same joint distributions. A coalescent is a Markov process, whose detailed properties will be explored elsewhere. Briefly, the values of R_t are equivalence relations on the infinite set $\mathbb{N}$ which have only finitely many equivalence classes, $(|R_t|; t > 0)$ is a pure death process starting from ∞, and the jump chain is independent of the death process.

8. The paintbox

Let R be an equivalence relation on $\mathbb{N}$, and let $\pi_n = \pi_n(R)$ be the partition of n induced by the restriction $\rho_n R$ (as in (5.2)). Suppose that we do not know R, but that we do know the partitions π_n for all $n > m$; what can then be said about π_m? Clearly all the relevant information is contained in π_{m+1}, which determines the sizes $(\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_l$, say) of the classes into which R divides $\{1, 2, \cdots, m+1\}$. The partitions π_n $(n \geqq m+2)$ merely add (incomplete) data about the classes into which $m+2, m+3, \cdots$ fall. A knowledge of π_{m+1} does not in general determine π_m, but tells us that the classes in π_m have sizes

$$\lambda_1 - \varepsilon_1, \lambda_2 - \varepsilon_2, \cdots, \lambda_l - \varepsilon_l,$$

where exactly one of the ε_i equals 1, and the rest 0.

Suppose now that R is a random member of $\mathscr{E}$, which is exchangeable in the sense that its distribution is unchanged by any finite permutation of $\mathbb{N}$. Then, given $\pi_n (n \geqq m+1)$, the conditional probability that $\varepsilon_i = 1$ is proportional to the size of the ith class of π_{m+1}, and is thus equal to $\lambda_i/(m+1)$. In other words, $(\pi_n; n = 2, 3, \cdots)$ is a Markov sequence of partitions whose backward transition probabilities are combinatorially determined in this way. Such sequences have been examined in Kingman (1978b), where it is proved that, if $\lambda_k(n)$ denotes the size of the kth largest class of λ_n, then the limit

$$x_k = \lim_{n \to \infty} \lambda_k(n)/n \tag{8.1}$$

exists for all $k \geqq 1$, with probability 1, and satisfies

$$x_1 \geqq x_2 \geqq x_3 \geqq \cdots \geqq 0, \quad \sum_{k=1}^{\infty} x_k \leqq 1. \tag{8.2}$$

Conditional on the x_k, the joint distributions of the π_n are determined by the 'paintbox construction', and this extends to give the distribution of R itself. Imagine a paintbox with colours $\mathscr{C}_0, \mathscr{C}_1, \mathscr{C}_2, \cdots$ present in the proportions $x_0, x_1, x_2, \cdots$, where

$$x_0 = 1 - \sum_{k=1}^{\infty} x_k. \tag{8.3}$$

Balls $\mathscr{B}_1, \mathscr{B}_2, \cdots$ are painted with colours chosen independently at random from

the paintbox, but $\mathscr{C}_0$ has the special property that different balls painted with it appear of different colours. The relation of identically coloured balls,

$$\{(i,j); i = j \text{ or } \mathscr{B}_i \text{ and } \mathscr{B}_j \text{ are painted the same colour } \mathscr{C}_k\ (k \geqq 1)\} \tag{8.4}$$

has a distribution depending on $x_1, x_2, \cdots$ which is the same as the conditional distribution of R given $x_1, x_2, \cdots$.

Now suppose that (R_t) is a coalescent. For each $t \geqq 0$, R_t is an exchangeable random element of $\mathscr{E}$, for which the limits (8.1) therefore exist. Denoting them by $x_k(t)$, and writing

$$x(t) = (x_1(t), x_2(t), \cdots), \tag{8.5}$$

we have a random process $(x(t); t \geqq 0)$ taking values in the space of sequences satisfying (8.2). If we know the distribution of $x(t)$ for a particular value of t, that of R_t can be deduced.

It would be interesting to study this process, not only because it is there, but because it relates to questions of practical interest. For example, the 'frequency spectrum' in a mutation model can be expressed in terms of the paintbox proportions $x_k(t)$. Kimura and Ohta (1978) have raised the question whether the spectrum in the mutation model (3.7) has a finite integral. Kesten's work suggests a negative answer, but the question is still unresolved, and an attack along the present lines might succeed.

References

EWENS, W. J. (1972) The sampling theory of selectively neutral alleles. *Theoret. Popn Biol.* **3**, 87–112 and 376.

EWENS, W. J. (1979) *Mathematical Population Genetics.* Springer-Verlag, Berlin.

FELSENSTEIN, J. (1975) A pain in the torus: some difficulties with models of isolation by distance. *Amer. Naturalist* **109**, 359–368.

FLEISCHMANN, K. AND SIEGMUND-SCHULTZE, R. (1978) An invariance principle for reduced family trees of critical spatially homogeneous branching processes. *Serdica* **4**, 111–134.

KALLENBERG, O. (1977) Stability of critical cluster fields. *Math. Nachr.* **77**, 7–43.

KENDALL, D. G. (1975) Some problems in mathematical genealogy. *In Perspectives in Probability and Statistics*, ed. J. Gani, Distributed by Academic Press, London for the Applied Probability Trust, Sheffield, 325–345.

KERSTAN, J., MATTHES, K. AND MECKE, J. (1978) *Infinitely Divisible Point Processes.* Wiley, Chichester.

KESTEN, H. (1980) The number of distinguishable alleles according to the Ohta–Kimura model of neutral mutation. *J. Math. Biol.* **10**, 167–187.

KIMURA, M. AND OHTA, T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Nat. Acad. Sci.* **75**, 2868–2872.

KINGMAN, J. F. C. (1976) Coherent random walks arising in some genetical models. *Proc. R. Soc. London* A **351**, 19–31.

KINGMAN, J. F. C. (1978a) Random partitions in population genetics. *Proc. R. Soc. London* A **361**, 1–20.

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

KINGMAN, J. F. C. (1978c) The dynamics of neutral mutation. *Proc. R. Soc. London* A **363**, 135–146.

KINGMAN, J. F. C. (1980) *Mathematics of Genetic Diversity*. Society for Industrial and Applied Mathematics, Washington.

MALÉCOT, G. (1969) *The Mathematics of Heredity*. Freeman, San Francisco.

MEYER, P. A. (1966) *Probabilités et Potentiel*. Hermann, Paris.

MORAN, P. A. P. (1958) Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–72.

MORAN, P. A. P. (1975) Wandering distributions and the electrophoretic profile. *Theoret. Popn Biol.* **8**, 318–330.

NELSON, E. (1959) Regular probability measures on function space. *Ann. Math.* **69**, 630–643.

OHTA, T. AND KIMURA, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* **22**, 201–204.

ROSENBLATT, M. (1974) *Random Processes*. Springer-Verlag, New York.

SAWYER, S. (1977) Asymptotic properties of the equilibrium probability of identity in a geographically structured population. *Adv. Appl. Prob.* **9**, 268–282.

SINGH, K. S., LEWONTIN, R. C. AND FELTON, A. A. (1976) Genetic heterogeneity within electrophoretic 'alleles' of xanthine dehydrogenase in *Drosophila pseudoobscura*. *Genetics* **84**, 609–629.

WACHTER, K. W., HAMMEL, E. A. AND LASLETT, P. (1978) *Statistical Studies of Historical Social Structure*. Academic Press, New York.