

THE AGES OF ALLELES AND A COALESCENT

PETER DONNELLY, **University College of Swansea*
SIMON TAVARÉ, ***University of Utah*

Abstract

A new coalescent is introduced to study the genealogy of a sample from the infinite-alleles model of population genetics. This coalescent also records the age ordering of alleles in the sample. The distribution of this process is found explicitly for the Moran model, and is shown to be robust for a wide class of reproductive schemes.

Properties of the ages themselves and the relationship between ages and class sizes then follow readily.

GENEALOGY; INFINITE-ALLELES MODEL; EWENS SAMPLING FORMULA

1. Introduction

The coalescent was introduced by Kingman (1982a, b, c), as a means of representing the genealogy of a sample taken from a population evolving according to one of a large class of models. Watterson (1984) analyzed a related process in the context of reproductive models involving mutation (the so-called infinite-alleles models). In this paper we study a coalescent akin to that of Watterson which also takes account of the age-ordering of alleles present in the sample. When the population evolves according to a Moran model the distribution of this new coalescent is found explicitly. In the spirit of Kingman (1982c) we derive an invariance result in the case where the underlying reproductive mechanism is one of a large class of exchangeable models. This allows us to approximate the genealogy of samples from large populations.

Problems relating to the ages of alleles are basic to the study of much of mathematical population genetics. If the alleles are labelled in order of increasing age, the distribution of class sizes in a sample follows easily from the properties of the new coalescent. This distribution is central, and many of the standard results on ages and allele frequencies follow from it. The structure of the coalescent also allows a characterization of the ages themselves. We give

Received 10 September 1984; revision received 21 December 1984.

* Present address: Department of Statistical Science, University College London, London WC1E 6BT, UK.

** Postal address: Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA.

explicit results in some cases. Finally we exhibit the partition structure induced on an infinite population by the age ordering and relate the properties of this partition structure to some similar structures arising in other areas of mathematical biology.

2. Moran's infinitely-many-neutral-alleles model

Consider a population of fixed size M haploid individuals (or genes) evolving through discrete time (or generations) $r = \dots -2, -1, 0, 1, 2, \dots$. At each time point r , one individual is chosen at random to die, and one individual is chosen at random to have a single offspring. The surviving individuals and the new offspring form the population at the next time, $r+1$. The offspring individual may be of the same allelic type as the parent with probability $1-u$, or, with probability u , may be a novel allelic type that has not been in the population before. This process is a discrete-time version of the infinitely-many-neutral-alleles model of evolution with Moran-type reproduction.

We now take a random sample of n genes without replacement from the population at some time labelled 0, and consider its composition with respect to the ancestral population at time $-m$. As in Watterson (1984) it is convenient to divide the individuals in the sample into two types of disjoint equivalence classes, which we label 'old' and 'new'. First, randomly label the n individuals in the sample $1, 2, \dots, n$. We say that individuals i and j in the sample are in the same *old* equivalence class at time m if i and j have the same ancestor at time $-m$, and no intervening mutation has occurred.

On the other hand, we say that i and j are in the same *new* equivalence class at time m if, for some r satisfying $1 \leq r < m$, individuals i and j have the same ancestor in generation $-r$, this ancestor itself being a mutant, with no intervening mutation between time $-r$ and 0.

With respect to time $-m$, we denote the number of old equivalence classes by D_m , and the number of new equivalence classes by F_m . D_m and F_m satisfy

$$(2.1) \quad D_m = 0, 1, \dots, n, \quad F_m = 0, 1, \dots, n - D_m$$

and, since each individual in the sample is equivalent only to itself with respect to time 0, we have

$$(2.2) \quad D_0 = n, F_0 = 0.$$

The process $\{D_m, m \geq 0\}$ is a Markov chain on the state space $\{n, n-1, \dots, 1, 0\}$ whose properties have been studied *inter alia* by several authors; see Section 3. $\{D_m, m \geq 0\}$ is a (discrete-time) pure death process whose

one-step transition probabilities p_{ij} are determined by

$$(2.3) \quad p_{i,i-1} = 1 - p_{i,i} = \frac{i(i + \theta - 1)}{M(M + \theta)}$$

where

$$(2.4) \quad \theta = Mu/(1 - u).$$

An explicit expression for the transition probabilities of this process is given in Watterson (1984).

The new equivalence classes, F_m in number, arise from ancestors who were mutants at generations $-T_1, -T_2, \dots, -T_{F_m}$, say. Here, we order the $\{T_i\}$ so that $m \cong T_1 > T_2 > \dots > T_{F_m}$. Recall that since there is at most one mutation per generation, no two of the T_i can be equal. Notice that by definition individuals in the same new equivalence class are of the same allelic type, and that the allelic types of individuals in distinct new equivalence classes are different. We may then define unambiguously the ages of the novel allelic types; the youngest allele in the sample is that which corresponds to the (new) equivalence class whose ancestor occurs at time $-T_{F_m}$, and the oldest of the novel alleles in the sample is that which corresponds to the equivalence class whose ancestor is a mutant at time $-T_1$. Note that as the process evolves, the age of a particular allele does not change; for convenience in Section 5, we allow its label to change. Of course with respect to a given time $-m$, there may be no new classes, in which case $F_m = 0$, and the $\{T_i\}$ above are not defined. We shall not be concerned here with the ages of allelic types represented in the old equivalence classes.

The new equivalence classes will be denoted by $\eta_1, \eta_2, \dots, \eta_{F_m}$ (\emptyset denoting the fact that there are no new equivalence classes), and their sizes by

$$(2.5) \quad \mu_j = |\eta_j|, \quad j = 1, \dots, F_m.$$

The η 's above are listed in increasing order of age, so that, for example, η_{F_m} is the equivalence class of individuals of oldest allelic type. The old equivalence classes will be denoted by $\xi_1, \xi_2, \dots, \xi_{D_m}$, with sizes

$$(2.6) \quad \lambda_i = |\xi_i|, \quad i = 1, \dots, D_m;$$

the ordering of the ξ_i is immaterial here.

We now represent the genealogy of the sample of individuals $\{1, \dots, n\}$ by the two-type equivalence classes of the form

$$(2.7) \quad R_m = (\xi_1, \dots, \xi_{D_m}; \eta_1, \dots, \eta_{F_m}).$$

The (backward) evolution of the process $\{R_m, m \cong 0\}$ then describes the genetic history of the sample. The stochastic behavior of R_m is provided in the following basic result.

Theorem 2.1.

$$(2.8) \quad \mathbb{P}(R_m = (\xi_1, \dots, \xi_{D_m}; \eta_1, \dots, \eta_{F_m}), F_m = l \mid D_m = k) \\ = \frac{(n-k)! k! \theta^l}{n! (k+\theta)_{(n-k)}} \cdot \frac{\lambda_1! \lambda_2! \dots \lambda_k! \mu_1! \dots \mu_l!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_l)},$$

for $k = 0, 1, \dots, n$, $l = 0, 1, \dots, n - k$. Here, $\theta = Mu/(1-u)$ and

$$x_{(n)} = x(x+1) \dots (x+n-1).$$

Proof. The proof is similar in spirit to that of Watterson (1984). Suppose that $D_m = k$, $F_m = l$, $R_m = (\xi, \eta) = (\xi_1, \dots, \xi_k; \eta_1, \dots, \eta_l)$. There are two sorts of change of state that can occur from R_m to R_{m+1} :

(a) Two ancestors at time $-m$ who founded ξ_i and ξ_j (for $1 \leq i < j \leq k$, say) may be descended, without mutation, from a single ancestor at time $-(m+1)$. This results in

$$R_{m+1} = (\xi_1, \dots, \xi_i \cup \xi_j, \dots, \xi_k; \eta_1, \dots, \eta_l),$$

and $D_{m+1} = k - 1$. This specific change has probability

$$(2.9) \quad 2(1-u) \cdot \frac{1}{M} \cdot \frac{1}{M} = \frac{2}{M(M+\theta)}.$$

(b) A particular ancestor at time $-m$ who founded ξ_i , say, may be a mutant offspring of an individual at time $-(m+1)$, in which case ξ_i becomes reclassified as a new class η_{l+1} with respect to time $-(m+1)$. The individuals in ξ_i now correspond to the oldest novel allele. This results in

$$R_{m+1} = (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_k; \eta_1, \dots, \eta_l, \xi_i),$$

and $D_{m+1} = k - 1$, $F_{m+1} = l + 1$. This specific change has probability

$$(2.10) \quad \frac{u}{M} = \frac{\theta}{M(M+\theta)}.$$

The probability of no change of state can be computed from (2.9) and (2.10). Since $D_m = k$, there are $\binom{k}{2}$ changes of the type with probability (2.9), and $\binom{k}{1}$ changes of the type with probability (2.10). Hence

$$(2.11) \quad \mathbb{P}(R_{m+1} = (\xi, \eta) \mid R_m = (\xi, \eta)) = 1 - \frac{k(k+\theta-1)}{M(M+\theta)}.$$

Just as in Watterson's case, the one-step transition probabilities for R_m depend only on D_m . If we now watch the process R_m only at those points at which it changes state, we obtain the embedded Markov chain which we shall

denote by $\{\mathcal{A}_k, k = n, n-1, \dots, 1, 0\}$. Then we have

$$(2.12) \quad R_m = \mathcal{A}_{D_m}.$$

The distribution of R_m is simplified by an independence result analogous to that of Watterson (1984) and Kingman (1982b) to the effect that

$$(2.13) \quad \mathbb{P}(\mathcal{A}_{D_m} = R) = \mathbb{P}(D_m = k)\mathbb{P}(\mathcal{A}_k = R)$$

where R is a two-type partition, with k old equivalence classes. It remains to find the distribution of \mathcal{A}_k .

Suppose that $\mathcal{A}_k = R = (\xi_1, \dots, \xi_k; \eta_1, \dots, \eta_l)$. We want to show that $\mathbb{P}(\mathcal{A}_k = R)$ is given by (2.8):

$$(2.14) \quad \mathbb{P}(\mathcal{A}_k = R) = \frac{(n-k)! k! \theta^l}{n! (k+\theta)_{(n-k)}} \frac{\lambda_1! \lambda_2! \dots \lambda_k! \mu_1! \dots \mu_l!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_l)}$$

Now

$$(2.15) \quad \mathbb{P}(\mathcal{A}_k = R) = \sum_{R'} \mathbb{P}(\mathcal{A}_k = R \mid \mathcal{A}_{k+1} = R') \mathbb{P}(\mathcal{A}_{k+1} = R')$$

where R' is any two-type collection of equivalence classes which may be changed into R by means of the two operations described in (a) and (b) above. These two cases give:

$$(a') \quad R' = (\xi_1, \dots, \xi_{i_1}, \xi_{i_2}, \dots, \xi_k; \eta_1, \dots, \eta_l)$$

where, for some $i \in \{1, \dots, k\}$, $\xi_{i_1} \cup \xi_{i_2} = \xi_i$. Given that a change occurs, the probability of a transition from R' to R is $2/(k+1)(k+\theta)$, from (2.9) and (2.11). Watterson (1984) has shown that the contribution to the right-hand side of (2.15) from transitions of this form is then

$$(2.16) \quad \frac{\theta^l \mu_1! \dots \mu_l! \lambda_1! \dots \lambda_k!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_l)} \cdot \frac{(n-k-1)! k!}{n! (k+\theta)_{(n-k)}} \sum_{i=1}^k (\lambda_i - 1).$$

(b') These changes result in creation of another new class. Here, R' is of the form $R' = (\xi_1, \dots, \xi_k, \eta_l; \eta_1, \dots, \eta_{l-1})$. Given that a change occurs, the probability of a transition to R is, from (2.10) and (2.11), $\theta/(k+1)(k+\theta)$. Hence, once more assuming that $\mathbb{P}(\mathcal{A}_{k+1} = R')$ is of the form specified by (2.14), the contribution to the right-hand side of (2.15) is

$$(2.17) \quad \frac{(n-k-1)! (k+1)! \theta^{l-1}}{n! (\theta+k+1)_{(n-k-1)}} \frac{\lambda_1! \dots \lambda_k! \mu_l! \mu_1! \dots \mu_{l-1}!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_{l-1})} \cdot \frac{\theta}{(k+1)(k+\theta)}$$

$$= \frac{\theta^l \mu_1! \dots \mu_l! \lambda_1! \dots \lambda_k!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_l)} \frac{(n-k-1)! k!}{n! (k+\theta)_{(n-k)}} \left(\sum_{j=1}^l \mu_j \right).$$

Adding (2.16) and (2.17) gives (2.14), since $\sum_{i=1}^k (\lambda_i - 1) + \sum_{j=1}^l \mu_j = n - k$.

Finally, when $k = n$ we have $l = 0$, $\lambda_1 = \cdots = \lambda_n = 1$, so (2.8) gives, correctly, probability 1 to the initial state for \mathcal{A}_n . The proof of the theorem is thus complete.

The following combinatorial result will be needed below. The proof is given in Appendix 1.

Proposition 2.1. Let $\mu_1, \mu_2, \dots, \mu_l$ be any l positive numbers, and let Π be the set of all permutations $\pi = (\pi(1), \dots, \pi(l))$ of $(1, \dots, l)$. Then

$$\sum_{\pi \in \Pi} [\mu_{\pi(1)}(\mu_{\pi(1)} + \mu_{\pi(2)}) \cdots (\mu_{\pi(1)} + \cdots + \mu_{\pi(l)})]^{-1} = [\mu_1 \cdots \mu_l]^{-1}.$$

Recall that in this paper R_m is a two-type coalescent in which the new equivalence classes η_1, η_2, \dots are ordered according to increasing age. It is also of interest to study the coalescent W_m , say, in which the ages of the new classes are immaterial. This process is precisely the one studied by Watterson, and it provided the starting point for our analysis. The probability distribution of W_m is obtained from that of R_m by summing (2.8) over all permutations π of the labels of η_1, \dots, η_l . Using the result of Proposition 2.1, with $R = (\xi; \eta)$, we obtain the following result.

Theorem 2.2 (Watterson (1984)).

$$\mathbb{P}(F_m = l, W_m = R \mid D_m = k) = \frac{(n-k)! k! \theta^l}{n! (k+\theta)_{(n-k)}} \left(\prod_{i=1}^k \lambda_i! \right) \left(\prod_{j=1}^l (\mu_j - 1)! \right).$$

3. An invariance principle

The explicit results provided in Theorem 1 were based on the Moran reproduction scheme in discrete time. In the spirit of Kingman (1982a) and Donnelly (1985), it is natural to ask in what sense the results of Theorem 2.1 are robust against changes in the reproduction mechanism.

We shall consider a class of neutral models with non-overlapping generations introduced by Cannings (1974). Randomly label the M individuals in a particular generation, and let v_i be the number of offspring born to the i th individual, $i = 1, 2, \dots, M$. The random variables $\{v_i\}$ are exchangeable, and constant population size is maintained by requiring $v_1 + v_2 + \cdots + v_M = M$. We shall also assume that the distribution of $\{v_1, \dots, v_M\}$ is invariant over time. The assumption of neutrality implies that the v_j are independent of family sizes in preceding generations. Each offspring individual may be subject to mutation at birth. Mutations occur independently for each individual with probability u , and, as in the infinite-alleles Moran model, result in a novel allelic type.

From the population, we take a random sample of size n individuals at time 0, and as in Section 2, define the corresponding coalescent with ages, which we here denote by $\{\tilde{R}_m, m = 0, 1, \dots\}$. The behavior of $\{\tilde{R}_m, m = 0, 1, \dots\}$ for large population sizes M is described below.

Theorem 3.1. Assume that as $M \rightarrow \infty$:

- (i) $\text{Var}(v_1) \rightarrow \sigma^2 > 0.$
- (ii) $\sup_M \mathbb{E}(v_1^k) < \infty, \quad k = 1, 2, \dots$
- (iii) $u = \frac{\theta\sigma^2}{2M} + o(M^{-1}), \quad \text{for some } \theta > 0.$

Then for fixed n , with $[x]$ denoting the integer part of x :

(a) The finite-dimensional distributions of the process $\{\tilde{R}_{[M\sigma^{-2}t]}, t \geq 0\}$ converge as $M \rightarrow \infty$ to those of a continuous-time two-type coalescent with ages $\{R_t, t \geq 0\}$.

(b) $\{R_t, t \geq 0\}$ is a Markov process whose state space is the collection of two-type equivalence relations of the form (ξ, η) , with transition rates

$$(3.1) \quad \begin{aligned} & q((\xi_1, \dots, \xi_k; \eta_1, \dots, \eta_l) \\ & \quad \rightarrow (\xi_1, \dots, \xi_i \cup \xi_j, \dots, \xi_k; \eta_1, \dots, \eta_l)) = 1 \\ & \quad \rightarrow (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_k; \eta_1, \dots, \eta_l, \xi_i)) = \frac{\theta}{2}, \end{aligned}$$

for $1 \leq i < j \leq k$.

(c) The process R_t may be represented as $R_t = \mathcal{A}_{D_t}$, where $\{\mathcal{A}_k, k = n, \dots, 0\}$ is the jump-chain of $\{R_t\}$, and $\{D_t, t \geq 0\}$ is a pure death process with death rate $d_k = k(k + \theta - 1)/2$ from state k . Further, $\{D_t\}$ is independent of $\{\mathcal{A}_k\}$, and

$$(3.2) \quad \mathbb{P}(R_t = (\xi, \eta)) = \mathbb{P}(D_t = k) \mathbb{P}(\mathcal{A}_k = (\xi, \eta)),$$

where k is the number of (old) equivalence classes in ξ .

$$(d) \quad \mathbb{P}(\mathcal{A}_k = (\xi, \eta)) = \frac{(n-k)! k! \theta^l}{n! (k+\theta)_{(n-k)}} \left(\prod_{i=1}^k \lambda_i! \right) \frac{\mu_1! \cdots \mu_l!}{\mu_1(\mu_1 + \mu_2) \cdots (\mu_1 + \cdots + \mu_l)}$$

where $\lambda_i = |\xi_i|$, $1 \leq i \leq k$ and $\mu_i = |\eta_i|$, $1 \leq i \leq l$.

Proof. Suppose that $R_m = (\xi, \eta)$. On looking back one more generation, we observe a transition to a state of the form $R' = (\xi_1, \dots, \xi_i \cup \xi_j, \dots,$

$\xi_k; \eta_1, \dots, \eta_l$) if the individuals i and j at time $-m$ who are the ancestors of ξ_i and ξ_j are descended from a common parent at time $-(m+1)$, and neither is a mutant. Under assumptions (i) and (ii) above, Kingman (1982c) showed that the probability of i and j sharing a common parent is $\sigma^2/M + o(M^{-1})$, so that, using (iii), the probability $p_{RR'}$ of a transition of this type is

$$(3.3) \quad p_{RR'} = \left(\frac{\sigma^2}{M} + o(M^{-1}) \right) (1-u)^2 = \frac{\sigma^2}{M} + o(M^{-1}).$$

Next, we might have a transition from R to $R' = (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_k; \eta_1, \dots, \eta_b, \xi_i)$ if individual i at time $-m$, who is the ancestor of ξ_b is a mutant offspring of the parent at time $-(m+1)$. This transition has probability u , so that by (iii) once more

$$(3.4) \quad p_{RR'} = \frac{\sigma^2 \theta}{2M} + o(M^{-1}).$$

All other changes that can occur have probability $o(M^{-1})$, either because they involve the coalescing of more than two old equivalence classes (which has probability $o(M^{-1})$, by Kingman (1982c)), mutation of more than one individual, or a coalescence and a mutation. Hence if we denote by $P_M = (p_{RR'})$ the one-step transition matrix of $\{\tilde{R}_m\}$, we can write

$$P_M = I + \frac{\sigma^2}{M} Q + o(M^{-1}),$$

where Q is the infinitesimal generator determined by (3.1). Finally, the argument of Kingman (1982a) shows that

$$\lim_{M \rightarrow \infty} P_M^{[M\sigma^{-2}t]} = \exp(Qt),$$

which is sufficient to establish (a) and (b). To verify (c), note that D_t is the number of old equivalence classes in R_t and that, from (3.1), the jump rate for D , from state k , is

$$(3.5) \quad d_k = \binom{k}{2} \cdot 1 + \binom{k}{1} \frac{\theta}{2} = \frac{k(k+\theta-1)}{2}.$$

Independence of $\{\mathcal{A}_k\}$ and $\{D_t\}$ follows just as in Kingman (1982b) and Watterson (1984). Finally, from (3.3), (3.4) and (3.5) the one-step transition probabilities of $\{\mathcal{A}_k\}$ are determined by

$$\begin{aligned} \mathbb{P}(\mathcal{A}_k = (\xi_1, \dots, \xi_i \cup \xi_j, \dots, \xi_{k+1}; \eta_1, \dots, \eta_l) \mid \mathcal{A}_{k+1}) \\ = (\xi_1, \dots, \xi_{k+1}; \eta_1, \dots, \eta_l) \\ = 2/(k+1)(k+\theta), \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}(\mathcal{A}_k = (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_{k+1}; \eta_1, \dots, \eta_b, \xi_i) \mid \mathcal{A}_{k+1}) \\ = (\xi_1, \dots, \xi_{k+1}; \eta_1, \dots, \eta_l) \\ = \theta/(k+1)(k+\theta). \end{aligned}$$

Hence the jump-chain $\{\mathcal{A}_k\}$ here has precisely the same stochastic structure as the jump chain for the Moran process, and thus (d) follows from Theorem 2.1.

Remarks.

(i) The well-known Wright–Fisher process is the particular case in which (v_1, \dots, v_M) has a joint multinomial distribution with

$$\mathbb{P}(v_i = y_i, 1 \leq i \leq M) = \frac{M!}{\prod_{j=1}^M y_j!} M^{-M}.$$

Here, $\sigma^2 = 1$.

(ii) The transition density and related properties of the process $\{D_t, t \geq 0\}$ have been studied by several authors; see Donnelly (1984), Tavaré (1984), Saunders *et al.* (1984) and Watterson (1984), Griffiths (1980).

The approximation described by Theorem 3.1 allows us to use the two-type coalescent structure of Theorems 2.1 and 2.2 to approximate the genetic history of samples taken from a large population, with mutation rate $u = \sigma^2\theta/2M$. Alternatively, we may look at exact results either for a sample or the whole population for the Moran process, in which $u = \theta/(M + \theta)$. The jump-chains of the two processes are the same, and the time-scale is determined by a pure death process in discrete time (for Moran’s model) or continuous time for the approximating processes.

We note that the analysis of the continuous-time formulation of the Moran process parallels that of Section 2, but the death process evolves in continuous time. With the appropriate time-scaling, this latter death process converges in distribution as $M \rightarrow \infty$ to that described in Theorem 3.1; see Donnelly (1985).

4. Properties of the coalescent with ages

The death process that provides the time-scale for the two-type coalescent process reaches 0 almost surely, at which time the old equivalence classes have disappeared. The process of ‘new’ equivalence classes then describes the genetic history of a random sample of size n taken from a stationary infinite-alleles model. We shall then denote by F the number of new equivalence classes. From (2.8) with $k=0$, we see that the probability $p(l; \eta_1, \dots, \eta_l)$, say, that the sample of size n from the stationary population

contains l allelic types, and equivalence classes η_1, \dots, η_l ranked in increasing order of age is

$$(4.1) \quad p(l; \eta_1, \dots, \eta_l) = \frac{\theta^l \mu_1! \cdots \mu_l!}{\theta_{(n)} \mu_1(\mu_1 + \mu_2) \cdots (\mu_1 + \cdots + \mu_l)},$$

where $\mu_i = |\eta_i|$, $\mu_i \geq 1$, $\mu_1 + \cdots + \mu_l = n$.

Usually we shall not be interested in which particular individuals belong to which equivalence classes. If we multiply the right-hand side of (4.1) by $n!/\mu_1! \cdots \mu_l!$, the number of ways in which the sample could have arisen, we obtain the probability $p_A(l; \mu_1, \dots, \mu_l)$, say, that a sample of n contains l allelic types, the oldest being represented by μ_l individuals, \dots , the youngest by μ_1 :

$$(4.2) \quad p_A(l; \mu_1, \dots, \mu_l) = \frac{\theta^l n!}{\theta_{(n)} \mu_1(\mu_1 + \mu_2) \cdots (\mu_1 + \cdots + \mu_l)}.$$

It is an elementary result about Stirling numbers $S_n^{(l)}$ of the first kind (see Appendix 2, (A.4)) that

$$(4.3) \quad \sum_{\mu_1, \dots, \mu_l}^* n!/\mu_1(\mu_1 + \mu_2) \cdots (\mu_1 + \cdots + \mu_l) = |S_n^{(l)}|,$$

Σ^* being taken over integers $\mu_i \geq 1$, with $\mu_1 + \cdots + \mu_l = n$. Summing (4.2) over μ_1, \dots, μ_l and using (4.3) gives

$$(4.4) \quad \mathbb{P}(F = l) = \frac{\theta^l |S_n^{(l)}|}{\theta_{(n)}}, \quad l = 1, \dots, n,$$

which is a well-known result of Ewens (1972). Further, if the ages of the types are immaterial, then summing (4.2) over all permutations of μ_1, \dots, μ_l gives the probability $p_E(l; \mu_1, \dots, \mu_l)$ that a sample of size n gives l allelic types, with one allele represented μ_1 times, \dots , an l th represented μ_l times. Using the result of Proposition 2.1 gives

$$(4.5) \quad p_E(l; \mu_1, \dots, \mu_l) = \frac{\theta^l n!}{\theta_{(n)} \mu_1 \mu_2 \cdots \mu_l} \\ = \frac{n!}{\theta_{(n)}} \prod_{j=1}^n \left(\frac{\theta}{j} \right)^{\beta_j},$$

where β_j is the number of alleles represented j times; $1\beta_1 + \cdots + n\beta_n = n$, and $\beta_1 + \cdots + \beta_n = l$. Equation (4.5) is the celebrated Ewens sampling formula (Ewens (1972)), derived in the context of coalescents by Kingman (1982a) and Watterson (1984).

For fixed class sizes μ_1, \dots, μ_b , the distribution (4.2) assigns greatest

probability to the configuration in which $\mu_1 \leq \dots \leq \mu_b$, so one might expect older alleles to be more frequent. Crow (1972) posed the question, ‘Is the most frequent allele the oldest?’ Kelly (1977) and Watterson and Guess (1977) showed that the probability of this is the expected relative frequency of the most frequent allele in the sample. This also follows from (4.2) in the form of the following proposition, the original proof of which (Watterson (1976a), Kelly (1977)) exploited reversibility.

Proposition 4.1. The probability that an allele, A , represented i times in a sample of size n is the oldest in the sample is i/n .

Proof

$$\begin{aligned} & \mathbb{P}(\text{allele } A \text{ is oldest} \mid F = l; \mu_1, \dots, \mu_l \in \{\psi_1, \dots, \psi_{l-1}, i\}) \\ &= \frac{\sum_{\mu_1, \dots, \mu_{l-1} \in \{\psi_1, \dots, \psi_{l-1}\}} \theta^l \frac{n!}{\theta_{(n)} \mu_1 (\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_{l-1}) n}}{p_E(l; \psi_1, \dots, \psi_{l-1}, i)} \\ &= \left(\frac{(n-1)!}{\psi_1 \psi_2 \dots \psi_{l-1}} \right) / \left(\frac{n!}{\psi_1 \psi_2 \dots \psi_{l-1} i} \right) \quad \text{from Proposition 2.1 and (4.5)} \\ &= \frac{i}{n}. \end{aligned}$$

The result now follows since this probability does not depend on $l, \psi_1, \dots, \psi_{l-1}$.

5. The ages themselves

To date we have considered the relationship between class sizes and the age ordering of the alleles. In this section, via a closer analysis of the death process D_p , we examine properties of the ages T_1, T_2, \dots, T_F of the alleles in the sample at stationarity. We restrict attention to the case when D_t evolves as a death process in continuous time with death rate $k(k + \theta - 1)/2$ from state k to state $k - 1$. Analogous results in the discrete-time setting follow *mutatis mutandis*.

Denote by $\tau_j, j = n, n - 1, \dots, 1$, the amount of time for which $D_t = j$, that is the times between successive jumps of the process $\{\mathcal{A}_k\}$. We know that the random variables $\tau_n, \tau_{n-1}, \dots, \tau_1$ follow independent exponential distributions with parameters $n(n - 1 + \theta)/2, (n - 1)(n - 2 + \theta)/2, \dots, \theta/2$ respectively. With each τ_j associate an indicator random variable χ_j to mark whether the change from \mathcal{A}_j to \mathcal{A}_{j-1} involves the coalescence of two old equivalence classes ($\chi_j = 0$) or the creation of a new equivalence class ($\chi_j = 1$). It follows from Theorem 3.1 that the random variables χ_i are independent with

$\mathbb{P}(\chi_i = 1) = \theta/(i + \theta - 1)$, $i = n, n-1, \dots, 1$, and furthermore that the random variables $(\chi_n, \chi_{n-1}, \dots, \chi_1)$ are independent of the random variables $(\tau_n, \tau_{n-1}, \dots, \tau_1)$. Note that $\chi_1 = 1$ a.s.

We note in passing that, as in Watterson (1984), F , the number of different allelic types in the sample, may be written as

$$(5.1) \quad F = \chi_n + \chi_{n-1} + \dots + \chi_1,$$

from which its moments and probability generating function follows readily. Also

$$\begin{aligned} \mathbb{P}(F = l) &= \mathbb{P}(\text{exactly } l-1 \text{ of } \chi_n, \chi_{n-1}, \dots, \chi_2 \text{ are } 1) \\ &= \frac{\theta^l}{(\theta + n - 1)(\theta + n - 2) \dots \theta} \sum' b_1 \dots b_{n-l} \end{aligned}$$

where the summation \sum' is over integers b_1, \dots, b_{n-l} satisfying $1 \leq b_1 < b_2 < \dots < b_{n-l} \leq n-1$. Thus from Appendix 2 (A.2) we have again that

$$\mathbb{P}(F = l) = \frac{\theta^l}{\theta_{(n)}} |S_n^{(l)}|.$$

We introduce the random variables H_i , $i = 1, 2, \dots, n$, defined by

$$(5.2) \quad H_i = \begin{cases} 1 & \\ k & \text{if } \chi_k = 1 \text{ and exactly } i-2 \text{ of } \chi_{k-1}, \dots, \chi_2 \text{ take the value } 1 \\ \infty & \text{if fewer than } i \text{ of } \chi_n, \dots, \chi_1 \text{ take the value } 1. \end{cases}$$

It is evident that H_i takes the value k if the jump from \mathcal{A}_k to \mathcal{A}_{k-1} involves the creation of a new equivalence class, and exactly $i-1$ more new equivalence classes are created as \mathcal{A}_{k-1} evolves to \mathcal{A}_0 . If there are fewer than i new classes in \mathcal{A}_0 then H_i takes the value ∞ . Thus, when it is finite, H_i labels the jump at which the i th-oldest new equivalence class was created. When it is defined, that is when $H_i < \infty$ or equivalently when $F \geq i$, we may write T_i , the age of the i th-oldest allele, as

$$(5.3) \quad T_i = \tau_n + \tau_{n-1} + \dots + \tau_{H_i}.$$

The age of the oldest allele is given by

$$T_1 = \tau_n + \dots + \tau_1,$$

the time for the death process D_t to reach 0. Exact expressions for the distribution of T_1 are given for example in Tavaré (1984). Notice that it follows from the independence of $\{D_t\}$ and $\{\mathcal{A}_k\}$ and T_1 is independent of the process $\{\mathcal{A}_k\}$, and in particular that T_1 is independent of the class sizes μ_1, \dots, μ_F . This latter result is originally due to Kelly (1977).

The joint distribution of the H_i , $i = 1, 2, \dots, F$ is straightforward. For example

$$\mathbb{P}(H_1 = 1, H_2 = k_2, H_3 = k_3, \dots, H_l = k_l, F = l) = \frac{\theta^l (n-1)!}{\theta_{(n)} k_2 k_3 \dots k_l}$$

$$\text{for } l = 1, 2, \dots, n; \quad 1 < k_2 < k_3 < \dots < k_l \leq n.$$

With the aid of the characterization (5.3) this gives, at least in principle, the joint distribution of the ages T_1, \dots, T_F .

For the marginal distribution, note from (5.2) that

$$\mathbb{P}(H_i = k) = \frac{\theta^i}{\theta_{(k)}} \sum' b_1 \dots b_{k-i} \quad k = i, i+1, \dots, n,$$

where the summation \sum' is over integers b_1, \dots, b_{k-i} satisfying $1 \leq b_1 < \dots < b_{k-i} \leq k-2$. Thus from Appendix 2 (A.2) we have

$$(5.4) \quad \mathbb{P}(H_i = k) = \frac{\theta^i}{\theta_{(k)}} |S_{k-1}^{(i-1)}| \quad k = i, i+1, \dots, n.$$

On the set where $H_i = k$, the distribution of $T_i = \tau_n + \dots + \tau_k$ is known. ($\mathbb{P}(\tau_n + \dots + \tau_k \leq t) = \mathbb{P}(D_i \leq k-1)$; this latter can be computed from Tavaré (1984), Equations 5.2 and 5.3, for example.) Averaging this over the distribution (5.4) of H_i gives the distribution of T_i , the age of the i th-oldest allele, explicitly. Of course the moments and generating functions of T_i may also be found via (5.3) and (5.4). We omit the details.

6. Partition by ages

We return now to the distribution of the age-class sizes themselves. We shall be concerned with the distribution (4.2) of age-class sizes from a stationary infinite-alleles sample of size n . To emphasize its dependence on n , the number of alleles in the sample will be denoted in this section by $F^{(n)}$. Let $X_{i,n}$, $i = 1, 2, \dots, F^{(n)}$ denote the frequency of the oldest, \dots , youngest alleles in the sample.

Proposition 6.1. For fixed r , with $x_i \in \{1, \dots, n\}$, $i = 1, \dots, r$ satisfying $n-l+r \geq x_1 + \dots + x_r \geq r$, if $l \geq r$ then

$$(6.1) \quad \mathbb{P}(X_{i,n} = x_i, 1 \leq i \leq r \mid F^{(n)} = l)$$

$$= \frac{n! |S_{n-x_1-\dots-x_r}^{(l-r)}|}{|S_n^{(l)}| (n-x_1-\dots-x_r)! n(n-x_1) \dots (n-x_1-\dots-x_{r-1})},$$

and

$$(6.2) \quad \mathbb{P}(X_{i,n} = x_i, 1 \leq i \leq r; F^{(n)} \geq r) \\ = \frac{\theta^r}{n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})} \frac{n!}{\theta_{(n)}} \frac{\theta_{(n-x_1-\cdots-x_r)}}{(n-x_1-\cdots-x_r)!}.$$

Proof

$$\mathbb{P}(X_{i,n} = x_i, 1 \leq i \leq r; F^{(n)} = l) = \sum' p_A(l; \mu_1, \dots, \mu_{l-r}, x_r, x_{r-1}, \dots, x_1)$$

where \sum' is over all integers μ_1, \dots, μ_{l-r} with $\mu_j \geq 1$, and $\mu_1 + \dots + \mu_{l-r} = n - x_1 - \dots - x_r$

$$= \frac{\theta^l n!}{\theta_{(n)} (n-x_1-\cdots-x_r)! n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})} \\ \times \sum' \frac{(n-x_1-\cdots-x_r)!}{\mu_1(\mu_1+\mu_2) \cdots (\mu_1+\cdots+\mu_{l-r})} \\ = \frac{\theta^l n! |S_{n-x_1-\cdots-x_r}^{(l-r)}|}{\theta_{(n)} (n-x_1-\cdots-x_r)! n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})},$$

using Appendix 2 (A.4).

Now (6.1) follows on dividing this last quantity by $\mathbb{P}(F^{(n)} = l)$ given in (4.4).

To establish (6.2), notice that if $s = n + r - x_1 - \dots - x_r$

$$\mathbb{P}(X_{i,n} = x_i, 1 \leq i \leq r; F^{(n)} \geq r) \\ = \sum_{l=r}^s \mathbb{P}(X_{i,n} = x_i, 1 \leq i \leq r; F^{(n)} = l) \\ = \sum_{l=r}^s \frac{\theta^l n!}{\theta_{(n)} (n-x_1-\cdots-x_r)!} \frac{|S_{n-x_1-\cdots-x_r}^{(l-r)}|}{n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})} \frac{1}{\theta^r n!} \\ = \frac{\theta^r n!}{\theta_{(n)} (n-x_1-\cdots-x_r)! n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})} \sum_{l=0}^{s-r} \theta^l |S_{s-r}^{(l)}| \\ = \frac{\theta^r}{n(n-x_1) \cdots (n-x_1-\cdots-x_{r-1})} \frac{n!}{\theta_{(n)}} \frac{\theta_{(n-x_1-\cdots-x_r)}}{(n-x_1-\cdots-x_r)!}.$$

Remark. When $r = 1$, (6.2) can be written in the form

$$\mathbb{P}(X_{1,n} = j) = \frac{\theta}{n} \frac{n!}{\theta_{(n)}} \cdot \frac{\theta_{(n-j)}}{(n-j)!} \\ = \frac{\theta}{n} \binom{n}{j} / \binom{\theta+n-1}{j}.$$

This last result is due to Kelly (1977) in the context of the Moran model.

Now suppose that the process is evolving in continuous time, as in Section 3. Having taken a limit as $M \rightarrow \infty$, $F^{(n)}$ may be interpreted as the number of alleles in a sample of size n from a stationary infinite alleles diffusion model.

Theorem 6.1. For each fixed integer $r \geq 1$,

$$n^{-1}(X_{1,n}, \dots, X_{r,n}) \xrightarrow{\mathcal{D}} (X_1, X_2, \dots, X_r), \text{ as } n \rightarrow \infty$$

where

$$(6.3) \quad X_1 = Y_1, \quad X_2 = (1 - Y_1)Y_2, \dots, X_r = (1 - Y_1)(1 - Y_2) \dots (1 - Y_{r-1})Y_r,$$

with Y_1, Y_2, \dots independent and identically distributed random variables with probability density function $f(x) = \theta(1 - x)^{\theta-1}$, $0 \leq x \leq 1$.

Proof. First note that from (5.1) and the Borel–Cantelli lemma,

$$(6.4) \quad F^{(n)} \rightarrow \infty \text{ a.s. as } n \rightarrow \infty.$$

Let y_1, y_2, \dots, y_r be positive numbers satisfying $0 < y_i < 1$, $1 \leq i \leq r$, and $0 < \sum_{i=1}^r y_i < 1$. From the asymptotics of the gamma function (e.g. Abramowitz and Stegun (1972), p. 257), it follows that for any sequence of integers $m_n \uparrow \infty$ as $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} m_n^{b-a} \frac{\Gamma(m_n + a)}{\Gamma(m_n + b)} = 1.$$

Hence from (6.2) and (6.4), we have

$$(6.5) \quad \lim_{n \rightarrow \infty} n^r \mathbb{P}(X_{i,n} = [ny_i], 1 \leq i \leq r; F^{(n)} \geq r) = \frac{\theta^r (1 - y_1 - \dots - y_r)^{\theta-1}}{(1 - y_1)(1 - y_1 - y_2) \dots (1 - y_1 - \dots - y_{r-1})}.$$

It is readily verified that if $\{Y_i\}$ and $\{X_i\}$ are defined as in (6.3), then the right-hand side of (6.5) is the joint probability density function of (X_1, \dots, X_r) . The theorem follows from (6.5) and an application of Scheffé’s theorem (cf. Serfling (1980), p. 17).

Remark 1. It was established by Kingman (1975) that if $Z_{(1)} \geq Z_{(2)} \geq \dots$ denote the ordered allele frequencies in the stationary infinite-alleles diffusion model (e.g. Watterson (1976b)), then the joint distribution of $(Z_{(1)}, Z_{(2)}, \dots)$ is a Poisson–Dirichlet with parameter θ . Patil and Taillie (1977) showed that if (X_1, X_2, \dots) are defined as in (6.3), then the joint distribution of the descending order statistics $X_{(1)} \geq \dots \geq X_{(r)} \geq \dots$ is also Poisson–Dirichlet with parameter θ .

Remark 2. An interpretation of the representation of (6.3) as the age-ordered partition of the infinite alleles diffusion model has been obtained independently by R. C. Griffiths (personal communication).

7. Partition structures

There has been considerable recent interest in partition structures on populations. See for example Patil and Taillie (1977), Kingman (1975), (1980). The last section developed a partition of an infinite population based on the ages of the alleles present. Earlier sections studied the partitions induced on samples by ages. We now examine further properties of these partitions and relate them to some partition structures arising elsewhere in mathematical biology.

Consider first partitions of a sample. Suppose we take a sample of size n from a population at equilibrium, and that the distribution of the allelic partition $(l; \mu_1, \dots, \mu_l)$ is given by the Ewens sampling formula (4.5). Recall that the labels of the alleles are now arbitrary, and not based on their ages. Given the sample partition $(l; \mu_1, \dots, \mu_l)$, construct the size-biased partition as follows: choose an individual at random from the n and record his allelic type i_1 , say. After removing from the sample all individuals of this type, choose an individual at random from the remaining individuals, record his allelic type (i_2 , say), and then remove all the individuals of this type. Continuing in this way produces a size-biased permutation $\Pi^* = (i_1, \dots, i_l)$ of $(1, \dots, l)$ and, conditional on $(l; \mu_1, \dots, \mu_l)$

$$(7.1) \quad \begin{aligned} \mathbb{P}(\Pi^* = (i_1, \dots, i_l)) &= \frac{\mu_{i_1}}{n} \cdot \frac{\mu_{i_2}}{n - \mu_{i_1}} \dots \frac{\mu_{i_l}}{\mu_{i_l}} \\ &= \frac{\mu_{i_1} \mu_{i_2} \dots \mu_{i_l}}{\mu_{i_1} (\mu_{i_1} + \mu_{i_2}) \dots (\mu_{i_1} + \dots + \mu_{i_l})}. \end{aligned}$$

Denote by $(l; \mu_1^*, \dots, \mu_l^*)$ the size-biased partition obtained in this way (so that there are μ_1^* genes of the type first chosen, μ_2^* genes of the second type chosen, and so on). Averaging (7.1) over the Ewens distribution (4.5) now gives the distribution of the size-biased partition $(l; \mu_1^*, \dots, \mu_l^*)$ as

$$(7.2) \quad \frac{\theta^l}{\theta_{(n)}} \frac{n!}{\mu_1^* (\mu_1^* + \mu_2^*) \dots (\mu_1^* + \dots + \mu_l^*)}.$$

Thus, identifying the genes labeled j in the size-biased partition with the j th-oldest allele in the sample, the distribution (7.2) of the size-biased partition is the same as that of the age partition (4.2). This equivalence is not surprising. By reversibility (see for example Kelly (1979), exercise 7.2.5) the age partition

is the same as the partition of alleles in terms of length of survival, and by symmetry this latter partition should be equivalent to the size-biased partition. It does provide one interpretation of the way in which novel alleles emerge, and confirms the intuitive observation that the oldest allele is most likely to be the most frequent. Incidentally, the equivalence of (4.2) and (7.2) provides a further proof of Proposition 4.1. Furthermore, it follows from Theorem 2.1 of Patil and Taillie (1977) that if $X_{i,n}$ is the frequency of the i th-oldest allele in a sample of size n , then $\mathbb{E}(X_{i,n})$ is decreasing in i .

The partition of a population given by Theorem 6.1 arises in the context of population diversity as a particular residual allocation model (Engen (1975)). (A residual allocation model is one in which $X_1 = Y_1$, $X_2 = (1 - Y_1)Y_2$, \dots for some sequence of independent random variables Y_i called the residual fractions. In Engen's model the Y_i have probability density function $\theta(1 - x)^{\theta-1}$.) It was noted in the last section that the order statistics of this partition give the partition induced by the Poisson–Dirichlet distribution of Kingman (1975). Patil and Taillie (1977) note that the Engen partition (i.e. that of Theorem 6.1) is the size-biased version of Kingman's partition, and moreover that it is characterized by the fact that it is the only infinite residual allocation model whose residual fractions are independent and identically distributed and which is invariant under size-biased permutation.

It is well known that the relative frequency in the population of the allelic type of an individual chosen at random from an infinite-alleles model at stationarity has probability density function $\theta(1 - x)^{\theta-1}$ (see for example Sawyer (1977)). The density of the allelic type of a randomly chosen individual is exactly the density of the allele labelled 1 in the size-biased permutation of the underlying partition. The result follows by the above argument since the size-biased partition is identical to that of Engen (or equivalently that of Theorem 6.1) in which the probability density function of X_1 is $\theta(1 - x)^{\theta-1}$.

Sawyer and Hartl (1984) obtain the partition of a population of (6.3) in the context of a model for local selection. They also discuss a mechanism which would give rise to (6.3) as an age-ordered partition, but in their case the age ordering is opposite to that of (6.3). That the partition structure they obtain from local selection is equivalent to the partition by ages of an infinite-alleles population, at least as far as taking finite samples in which alleles are labelled arbitrarily, sheds light on the fact that they obtain the Ewens sampling formula as the description of samples from their model.

Acknowledgement

It is a pleasure to acknowledge several helpful discussions with Dr R. C. Griffiths and Dr G. A. Watterson, who also drew our attention to the results of Patil and Taillie.

Appendix 1. Proof of Proposition 2.1

We need to prove that if $\mu_1, \mu_2, \dots, \mu_l$ are positive numbers, and Π denotes the collection of permutations $\pi = (\pi(1), \dots, \pi(l))$ of the set $\{1, 2, \dots, l\}$, then

$$(A.1) \quad \sum_{\pi \in \Pi} \frac{1}{\mu_{\pi(1)}(\mu_{\pi(1)} + \mu_{\pi(2)}) \cdots (\mu_{\pi(1)} + \cdots + \mu_{\pi(l)})} = \frac{1}{\mu_1 \mu_2 \cdots \mu_l}.$$

The proof proceeds inductively. Let Π_j denote the collection of permutations $\pi \in \Pi$ with $\pi(l) = j$. The left-hand side of (A.1) may be written

$$\begin{aligned} & \frac{1}{\mu_1 + \mu_2 + \cdots + \mu_l} \sum_{j=1}^l \sum_{\pi \in \Pi_j} \frac{1}{\mu_{\pi(1)}(\mu_{\pi(1)} + \mu_{\pi(2)}) \cdots (\mu_{\pi(1)} + \cdots + \mu_{\pi(l-1)})} \\ &= \frac{1}{\mu_1 + \cdots + \mu_l} \sum_{j=1}^l \frac{\mu_j}{\mu_1 \mu_2 \cdots \mu_l} \quad (\text{by inductive hypothesis}) \\ &= 1/\mu_1 \cdots \mu_l. \end{aligned}$$

When $l = 1$, the identity in (A.1) is trivial. Hence the proof is complete.

Appendix 2

Let $S_n^{(l)}$ be a Stirling number of the first kind. That is, $|S_n^{(l)}|$ is the coefficient of x^l in the expansion of $x_{(n)} = x(x+1) \cdots (x+n-1)$; cf. Abramowitz and Stegun (1972), p. 824. By comparing coefficients, it follows that

$$(A.2) \quad |S_n^{(l)}| = \sum' b_1 b_2 \cdots b_{n-l}$$

where \sum' is over integers b_1, \dots, b_{n-l} satisfying $1 \leq b_1 < b_2 < \cdots < b_{n-l} \leq n-1$. Hence

$$(A.3) \quad |S_n^{(l)}| = \sum'' \frac{(n-1)!}{a_1 a_2 \cdots a_{l-1}},$$

where \sum'' is over integers a_1, \dots, a_{l-1} satisfying $1 \leq a_1 < \cdots < a_{l-1} \leq n-1$. Finally, it follows from (A.3) that

$$(A.4) \quad |S_n^{(l)}| = \sum''' n! / [\mu_1(\mu_1 + \mu_2) \cdots (\mu_1 + \cdots + \mu_l)],$$

where \sum''' is over integers $\mu_i \in \{1, \dots, n\}$, satisfying $\sum_1^l \mu_i = n$.

References

ABRAMOWITZ, M. AND STEGUN, L. A. (1972) *Handbook of Mathematical Functions*. Dover, New York.

- CANNINGS, C. (1974) The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. Appl. Prob.* **6**, 260–290.
- CROW, J. F. (1972) The dilemma of nearly neutral mutations; how important are they for evolution and human welfare. *J. Hered.* **63**, 306–316.
- DONNELLY, P. J. (1984) The transient behavior of the Moran model in population genetics. *Math. Proc. Camb. Phil. Soc.* **95**, 349–358.
- DONNELLY, P. J. (1985) Dual processes and an invariance result for exchangeable models in population genetics. *J. Math. Biol.*
- ENGEN, S. (1975) A note on the geometric series as a species frequency model. *Biometrika* **62**, 694–699.
- EWENS, W. J. (1972) The sampling theory of selectively neutral alleles. *Theoret. Popn Biol.* **3**, 87–112.
- GRIFFITHS, R. C. (1980) Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoret. Popn Biol.* **17**, 37–50.
- KELLY, F. P. (1977) Exact results for the Moran neutral allele model. *Adv. Appl. Prob.* **9**, 197–201.
- KELLY, F. P. (1979) *Reversibility and Stochastic Networks*. Wiley, New York.
- KINGMAN, J. F. C. (1975) Random discrete distributions. *J. R. Statist. Soc.* **B37**, 1–22.
- KINGMAN, J. F. C. (1980) *Mathematics of Genetic Diversity*. CBMS-NSF Regional Conference in Applied Mathematics **34**, Society for Industrial and Applied Mathematics, Philadelphia.
- KINGMAN, J. F. C. (1982a) On the genealogy of large populations. *J. Appl. Prob.* **19A**, 27–43.
- KINGMAN, J. F. C. (1982b) The coalescent. *Stoch. Proc. Appl.* **13**, 235–248.
- KINGMAN, J. F. C. (1982c) Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, ed. G. Koch and F. Spizzichino, North-Holland, Amsterdam, 97–112.
- PATIL, G. P. AND TAILLIE, C. (1977) Diversity as a concept and its implications for random communities. *Bull. Internat. Stat. Inst.* **47**, 497–515.
- SAUNDERS, I. W., TAVARÉ, S., AND WATTERSON, G. A. (1984) On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob.* **16**, 471–491.
- SAWYER, S. (1977) On the past history of an allele now known to have frequency p . *J. Appl. Prob.* **14**, 439–450.
- SAWYER, S. AND HARTL, D. (1984) A sampling theory for local selection. *J. Genet.*
- SERFLING, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- TAVARÉ, S. (1984) Line-of-descent and genealogical processes and their applications in population genetics models. *Theoret. Popn Biol.* **26**, 119–164.
- WATTERSON, G. A. (1976a) Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. *Theoret. Popn Biol.* **10**, 239–253.
- WATTERSON, G. A. (1976b) The stationary distribution of the infinitely many neutral alleles diffusion model. *J. Appl. Prob.* **13**, 639–651.
- WATTERSON, G. A. (1984) Lines of descent and the coalescent. *Theoret. Popn Biol.* **26**, 77–92.
- WATTERSON, G. A. AND GUESS, H. A. (1977) Is the most frequent allele the oldest? *Theoret. Popn Biol.* **11**, 141–160.

CORRECTION

DONNELLY, P. AND TAVARÉ, S. (1986) The ages of alleles and a coalescent. *Adv. Appl. Prob.* **18**, 1-19.

There is an omission in the formula (4.5). The stated formula is not correct, nor is it the Ewens sampling formula. It should read

$$p_E(l; \mu_1, \mu_2, \dots, \mu_l) = \frac{\theta^l}{\theta_{(n)}} \frac{n!}{\mu_1 \mu_2 \dots \mu_l \beta_1! \beta_2! \dots \beta_n!}.$$

This is the Ewens sampling formula. The factor $(\beta_1! \beta_2! \dots \beta_n!)^{-1}$ arises because the summation of (4.2) over all permutations of μ_1, \dots, μ_l counts an unordered collection $\{\mu_1, \dots, \mu_l\}$ of class sizes exactly $(\beta_1! \beta_2! \dots \beta_n!)$ times.

The proof of Proposition 4.1 should then be amended to read:

$\mathbb{P}(\text{allele } A \text{ is oldest} \mid A \text{ has } i \text{ representatives})$

$$= E(\mathbb{P}(\text{allele } A \text{ is oldest} \mid F = l, A \text{ has } i \text{ representatives}, A_1 \text{ has } \psi_1 \text{ representatives}, \dots, A_{l-1} \text{ has } \psi_{l-1} \text{ representatives}))$$

$$= E \left(\frac{\sum_{\mu_1, \dots, \mu_{l-1} \in (\psi_1, \dots, \psi_{l-1})} \frac{\theta^l}{\theta_{(n)}} \frac{n!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_{l-1})^n}}{\sum_{\mu_1, \dots, \mu_l \in (\psi_1, \dots, \psi_{l-1}, i)} \frac{\theta^l}{\theta_{(n)}} \frac{n!}{\mu_1(\mu_1 + \mu_2) \dots (\mu_1 + \dots + \mu_l)}} \right)$$

$$= E \left(\frac{(n-1)!}{\psi_1 \psi_2 \dots \psi_{l-1}} \bigg/ \frac{n!}{\psi_1 \psi_2 \dots \psi_{l-1} i} \right)$$

$$= \frac{i}{n}.$$