

## The population genealogy of the infinitely-many neutral alleles model

Peter Donnelly and Simon Tavaré

<sup>1</sup> Department of Statistical Science, University College, Gower Street, London WC1E 6BT, England

<sup>2</sup> Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

**Abstract.** A process analogous to Kingman's coalescent is introduced to describe the genealogy of populations evolving according to the infinitely-many neutral alleles model. The process records population frequencies in old and new classes, and labels the new classes in order of decreasing age. Its marginal distribution is characterized in a form which is amenable to explicit calculations and the transition densities of the associated  $K$ -allele models follow readily from this representation.

**Key words:** Coalescent —  $K$ -allele diffusions — Population genealogy

### 1. Introduction

Genealogical processes have received considerable attention recently and found a large number of applications within population genetics. Kingman (1982a, c) introduced the  $n$ -coalescent as a robust method of describing the genealogy of a sample of fixed size,  $n$ , taken from a large haploid population evolving according to one of a wide class of reproductive models. It also provides an exact description of the genealogy of the Moran model. With respect to an ancestral population some time  $t$  in the past, the  $n$ -coalescent partitions the sample into equivalence classes; individuals in the same class share a common ancestor in the ancestral population. As  $t$  increases (and the reference time recedes), the number of ancestors of the sample will decrease and the equivalence classes of the  $n$ -coalescent coalesce, until, for large enough  $t$ , the whole sample shares a single common ancestor, and forms one equivalence class. This process was extended by Watterson (1984) in the infinitely-many neutral alleles setting, to a process with two types of equivalence class, old and new. Individuals in the same old class share a common ancestor in the reference ancestral population and have no mutations in their line of descent from that ancestor (and so are all of the same genetic type as the ancestor). Individuals in the same new equivalence class share a common ancestor more recently (that is, less far into the past) than the reference ancestral population, that ancestor itself being a mutant (and so of a novel genetic type), with no further mutation in their lines of descent between the ancestor and the present time. Thus each individual in a particular new class

will share the same genetic type, and different new classes correspond to different types. Donnelly and Tavaré (1986) observed that it is possible to keep track of the order in which new classes arise and as a consequence to study the age structure of the alleles explicitly.

Each of these processes evolves on a discrete state space and describes the genealogy of *samples* from an evolving population. Indeed the robustness result (Kingman (1982c), Donnelly and Tavaré (1986)) relies heavily on the fact that the sample size is fixed as the population size tends to infinity. With the exception of the Moran model, less is known of the genealogy of the whole population. Kingman's (1982b) coalescent is a process in which are embedded  $n$ -coalescents for each value of  $n$  and it may be thought of as a description of the population genealogy. Our purpose here is to describe an analogous (though more complicated) population process in the infinitely-many alleles framework. In the infinite population diffusion time-scale limit the infinite alleles model is usually specified in terms of a particular diffusion model on the simplex of sequences of ordered non-negative real numbers which sum to one (see for example Ethier and Kurtz (1981)) or, more recently, as a measure-valued diffusion (cf. Ethier and Kurtz (1986)). The process described in this paper should be seen as an alternative way of thinking about (or specifying) the model, and would seem to have a number of advantages. Like the sample coalescents to which it is closely related (in particular that in Donnelly and Tavaré (1986)), the process enjoys a simple structure and its transitions always involve one of only two possible types of change of state. The marginal distribution of the process at any fixed time, and its equilibrium distribution, have a particularly tractable representation in a form which greatly facilitates direct calculation. These distributions, and indeed our whole approach, are closely allied to those of Griffiths (1980). In contrast to that work, and to the usual diffusion model, the genealogical process also keeps track of the age ordering of the new alleles. As well as making the process more informative, this device is largely responsible for the simplicity, both in structure and in distribution, which it enjoys.

In the next section we describe the process and give its marginal distribution and transition mechanism. The appendix is used to prove that such a process actually exists. The final section shows how to embed the usual  $K$ -allele models in the genealogical process and recovers their transition densities as an illustration of the ease with which calculations may be performed.

## 2. The process

Throughout, we will work on the so called diffusion (or continuous) time scale and consider a hypothetically infinite population. In the usual way the (scaled) mutation parameter will be denoted by  $\theta$ , which we assume to be positive. One could (at least conceptually) label the individuals from the set  $\mathbb{N} = \{1, 2, 3, \dots\}$  and describe genealogy by means of equivalence relations on this set of labels. This is the approach of Kingman's coalescent. However it is in fact equivalent (Kingman (1982b)) and for our purposes more convenient to specify the frequency of the population in each of the equivalence classes.

As usual fix  $t > 0$  and consider the genealogy of the population at the present time (time 0) with respect to the ancestral population time  $t$  in the past (time  $-t$ ). A certain proportion of the population,  $x_1$  say, may share a common ancestor, without intervening mutation, at time  $-t$  (in a coalescent type terminology these individuals would comprise an old equivalence class). A further proportion,  $x_2$  say, may share a different common ancestor, again without intervening mutation between time 0 and time  $-t$ , and so on. Suppose there are  $k$  such groups (for  $t > 0$ , we will see that  $k$  is almost surely finite) having frequencies  $x_1, \dots, x_k$  written in an arbitrary order (so that later probability statements will refer to the sequence  $x_1, \dots, x_k$  and not to the unordered set  $\{x_1, \dots, x_k\}$ ). Each such group will be of the same genetic type as the corresponding ancestor. There will also be a number of alleles in the current population which do not appear in the ancestral population (recall that  $\theta > 0$  and that the infinitely-many alleles assumption means that each mutation results in a completely novel allele). In fact, we will see that for  $t > 0$ , there will (almost surely) be an infinite number of distinct new alleles. All the individuals of a particular new allele (in coalescent terminology) form a new class—they are descended without mutation from a single common ancestor who was himself a mutant at some time  $-r$ ,  $0 < r < t$ . Each new allele arises at a different time (recall our diffusion type limit) and so we may order them by age, i.e. by time since first appearance in the population, with the oldest new allele arising at some time  $-r_1$ , the second oldest at  $-r_2$ , and so on, with  $t > r_1 > r_2 > \dots > 0$ . Denote the frequencies of the oldest, second oldest, ... allele (i.e. the age ordered frequencies in the new classes) by  $x_{k+1}, x_{k+2}, \dots$ . For this value of  $t$  we may abbreviate the above genealogical information by  $(k; x_1, x_2, \dots, x_k, x_{k+1}, x_{k+2}, \dots)$  for some  $k = 0, 1, 2, \dots, 0 \leq x_i \leq 1$  and  $\sum_{i=1}^{\infty} x_i = 1$ . As  $t$  increases this will give rise to a stochastic process describing the genealogy of the population, and we denote this process by  $\{M_t, t \geq 0\}$ .

When  $t = 0$  every individual in the population is its own (distinct) common ancestor. We denote this by  $(\infty; 0, 0, 0, \dots)$  and choose this as the value of  $M_0$ . The process  $\{M_t, t \geq 0\}$  thus has a state space which is a subset of  $\mathbb{N}_{\infty} \times \Delta$ , where we have written  $\mathbb{N}_{\infty}$  for  $\mathbb{N} \cup \{0, \infty\}$ , and  $\Delta$  for the set of sequences  $(x_1, x_2, \dots)$  with each  $x_i \in [0, 1]$ , and  $\sum x_i \leq 1$ .

Before describing the process, a few words about strategy are in order; there are a number of alternatives. At one extreme we could simply *define* the process  $\{M_t, t \geq 0\}$ , derive its properties, and then state that statements about the (infinite population) “infinitely-many alleles model” are those which necessarily follow from its having a genealogy described by  $\{M_t, t \geq 0\}$ . After all, many genetic models are often defined by specifying (via its generator for example) an appropriate diffusion. In our setting it would perhaps be appropriate to show that the “model” defined via  $\{M_t, t \geq 0\}$  is consistent with (or equivalent to) that which arises from the usual diffusion. At the other extreme one could look at the genealogical process associated with a finite population (infinitely-many alleles) Wright-Fisher, or more general, model and define the genealogical process associated with the infinite population model to be a suitable (weak) limit of the finite process as the population size increases, if such a limit exists. We will steer a middle course, arguing that on the basis of existing results concerning both the diffusion and finite genealogies, the process  $\{M_t, t \geq 0\}$  must behave in a certain

way. That the usual  $K$ -allele diffusions arise as consequences of the genealogy specified by  $\{M_t, t \geq 0\}$  is further convincing evidence for the consistency of the various models.

We first describe the transition mechanism of  $\{M_t, t \geq 0\}$ . Griffiths (1980) studied lines of descent in the diffusion approximation of neutral Wright-Fisher models. This is tantamount to studying the genealogy induced by the diffusion. Each of our old classes corresponds, in his terminology, to a line of descent with a root at time zero and he shows that for  $t > 0$  there are (almost surely) only a finite number of such. Suppose then that for some  $t > 0$ , we have  $M_t = (k; x_1, x_2, \dots, x_k, x_{k+1}, x_{k+2}, \dots)$  for some  $k < \infty$ . How can  $M_t$  change? Each new class corresponds to a particular new allele which arose after time  $-t$ , and so as  $t$  increases *the frequencies in these classes cannot change*. The classes themselves will be relabeled though, as extra new alleles arise. Each old class corresponds to a particular individual in the population at time  $-t$ . Tracing the genealogy of those members of the population (at time 0) who are in old classes with respect to the population at time  $-t$  is equivalent to tracing the genealogy of the  $k$  ancestors (one for each class) from  $-t$  further into the past. But this latter genealogy is much studied — it is that of a sample of fixed size  $k$  from an infinite population and is described by a  $k$ -coalescent and in particular by the process of Donnelly and Tavaré (1986), which takes account of mutation and the age structure. (The preceding argument is an example of Kingman's (1982a) temporal coupling.) The time until the next change occurs depends only on  $k$  and is exponentially distributed with mean  $2/(k(k + \theta - 1))$ . Given that a change occurs it will involve a particular pair of the  $k$  individuals sharing a common ancestor (and so the coalescence of the two corresponding old classes) with probability  $2/k(k + \theta - 1)$ , or a particular one of the  $k$  individuals being the result of a mutation (and so the corresponding old class becoming the oldest of the new classes) with probability  $\theta/(k(k + \theta - 1))$ . In each case the number of old classes decreases by exactly one.

We have thus specified the initial distribution of  $\{M_t, t \geq 0\}$  and its behavior from any positive time onwards. It can be shown that this is enough to characterize the process. We omit the details however and instead define the appropriate process and verify that it enjoys the required properties.

Let  $\{D_t, t \geq 0\}$  be a death process with death rate  $k(k + \theta - 1)/2$  from state  $k$ , and an entrance boundary at infinity. That is,  $D_0 = \infty$ , and  $D_t$  waits in state  $k$  for a period  $\tau_k$  which is independent of the past history of the process, and exponentially distributed with mean  $\mu_k = 2/k(k + \theta - 1)$ . Since, for any  $n$ , the time to first reach state  $n$ ,  $\sum_{k > n} \tau_k$ , has finite mean  $\sum_{k > n} \mu_k$ , it is almost surely finite. Intuitively, infinitely many of the waiting times  $\tau_k$  will have occurred in any finite interval  $[0, t)$  and once it reaches a finite value,  $D_t$  behaves as an "ordinary" death process. The value of this death process will correspond to the number of old classes in  $M_t$ . Notice that  $D_t$  is almost surely finite for  $t > 0$ , and given that  $D_t = n$ ,  $\{D_{t+s}, s \geq 0\}$  behaves as a death process with initial state  $n$  and the above death rates. Let  $\alpha_k$  be the indicator of the event that a new class is formed when the number of old classes drops from  $k + 1$  to  $k$ . The  $\alpha_k$  are independent, with  $\mathbb{P}(\alpha_k = 1) = \theta/(\theta + k)$ . The Borel-Cantelli Lemma then ensures that for any  $j$ ,  $\sum_{k \geq j} \alpha_k = \infty$  almost surely. Since  $\mathbb{P}(D_t < \infty) = 1$  for any  $t > 0$ , it follows that

$\sum_{k \geq D_t} \alpha_k = \infty$  almost surely. Thus for any  $t > 0$ ,  $M_t$  will contain an infinite number of new classes. Griffiths (1980) also shows that for any  $t > 0$ , the frequencies in the old lines are almost surely positive, and since each new class was an old class at some time  $s$ ,  $0 < s < t$ , it will have non-zero frequency with probability one.

We are now in a position to give a formal definition of the process  $\{M_t, t \geq 0\}$ . Let  $\{D_t, t \geq 0\}$  be the death process defined above. Let  $\{\mathfrak{M}_k, k = 0, 1, 2, \dots\}$  be a discrete time Markov chain on  $\mathbb{N}_\infty \times \Delta$ , independent of  $\{D_t, t \geq 0\}$ , with

$$\mathbb{P}(\mathfrak{M}_{k-1} = (k-1; x_1, \dots, x_{i-1}, x_i + x_j, \dots, x_{j-1}, x_{j+1}, \dots, x_k, x_{k+1}, \dots) |$$

$$\mathfrak{M}_k = (k; x_1, \dots, x_k, x_{k+1}, \dots)) = 1/k(k + \theta - 1), \quad 1 \leq i < j \leq k;$$

$$\mathbb{P}(\mathfrak{M}_{k-1} = (k-1; x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_j + x_i, x_{j+1}, \dots, x_k, x_{k+1}, \dots) |$$

$$\mathfrak{M}_k = (k; x_1, \dots, x_k, x_{k+1}, \dots)) = 1/k(k + \theta - 1), \quad 1 \leq i < j \leq k.$$

$$\mathbb{P}(\mathfrak{M}_{k-1} = (k-1; x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k, x_i, x_{k+1}, x_{k+2}, \dots) |$$

$$\mathfrak{M}_k = (k; x_1, \dots, x_k, x_{k+1}, \dots)) = \theta/k(k + \theta - 1), \quad 1 \leq i \leq k.$$

(That such a process exists will be shown in the Appendix.) Finally, define

$$M_0 = (\infty; 0, 0, \dots) \quad \text{and} \quad M_t = \mathfrak{M}_{D_t}, \quad (t > 0).$$

Now we give the distribution of  $\{M_t\}$ . It is easiest to give the distribution of  $\{D_t\}$  and then the distribution of  $M_t$  conditional on  $D_t$ , i.e., the distribution of  $\mathfrak{M}$ . The former is well known; see for example Tavaré (1984, equation (5.5)):

$$\mathbb{P}(D_t = k) = \begin{cases} \sum_{j=k}^{\infty} \exp(-j(j + \theta - 1)t/2) \frac{(-1)^{j-k} (2j + \theta - 1) \Gamma(k + \theta + j - 1)}{k!(j-k)! \Gamma(k + \theta)}, & k \geq 1 \\ 1 + \sum_{j=1}^{\infty} \exp(-j(j + \theta - 1)t/2) \frac{(-1)^j (2j + \theta - 1) \Gamma(\theta + j - 1)}{j! \Gamma(\theta)}, & k = 0. \end{cases}$$

In the appendix we will show that the distribution,  $\nu_k$  say, of  $\mathfrak{M}_k$  is concentrated on  $\{k\} \times \Delta^1$ , where  $\Delta^1$  is the subset of sequences in  $\Delta$  which sum to unity, and has the following representation. Let  $V_k$  be a random variable having the beta density  $f_k$  given by

$$f_k(x) = \frac{\Gamma(k + \theta) x^{k-1} (1-x)^{\theta-1}}{\Gamma(\theta)(k-1)!}, \quad 0 \leq x \leq 1. \tag{2.1}$$

Let  $(U_1, U_2, \dots, U_k)$  be a random  $k$ -vector having a uniform distribution on the simplex  $\{(u_1, u_2, \dots, u_k); u_i \geq 0, u_1 + u_2 + \dots + u_k = 1\}$  (i.e. density proportional to Lebesgue measure on the simplex), and let  $Z_1, Z_2, \dots$  be independent and identically distributed random variables with density  $f$  given by

$$f(x) = \theta(1-x)^{\theta-1}, \quad 0 \leq x \leq 1,$$

and choose  $(U_1, U_2, \dots, U_k)$ ,  $V_k$ , and  $\{Z_j\}$  to be mutually independent. Finally define the random element  $X^{(k)}$  of  $\Delta^1$  by

$$X^{(k)} = (V_k U_1, V_k U_2, \dots, V_k U_k, (1 - V_k) Z_1, (1 - V_k)(1 - Z_1) Z_2, (1 - V_k)(1 - Z_1)(1 - Z_2) Z_3, \dots).$$

When  $k = 0$ , put

$$X^{(0)} = (Z_1, (1 - Z_1) Z_2, (1 - Z_1)(1 - Z_2) Z_3, \dots). \quad (2.2)$$

Now for any Borel subset  $A$  of  $\Delta$ , define  $\nu_k$  by

$$\nu_k(\{k\} \times A) = \mathbb{P}(X^{(k)} \in A), \quad k = 0, 1, 2, \dots$$

*Remark 1.* At equilibrium ( $t \rightarrow \infty$ ) there will be no old classes and the age ordered frequencies of the new classes will have the representation (2.2). (For two alternative derivations and numerous applications of this representation see Donnelly and Tavaré (1986) and Donnelly (1986a).) Perhaps the easiest way of thinking about the distribution  $\nu_k$  of  $\mathcal{M}_k$  is as follows. The proportion of the population who are in old classes,  $V_k$ , has the distribution  $f_k$  given by (2.1). Conditional on this, the joint distribution of the proportions in each old class is uniform on the appropriate space and the joint distribution of the age ordered proportion in each of the new classes has the representation (2.2). That is, conditional on the total size of the new classes they behave like a stationary population of that size.

*Remark 2.* Griffiths (1980) obtained the distribution (2.1) of  $V_k$ , and the above distribution of  $(U_1, \dots, U_k)$  as the marginal distribution of the total and individual frequencies respectively of genes in old lines of descent in the diffusion approximation, conditional on there being  $k$  such old lines. That the distributions are identical to those above is hardly surprising. It is a consequence of the fact that our genealogical process  $\{M_t, t \geq 0\}$  and the usual infinite alleles diffusion really describe the same model. But note that our result does not automatically follow from Griffiths'. The novelty of the current approach lies in constructing and characterizing a genealogical process (rather than its distribution for fixed  $t$ ). Furthermore, the labeling of new alleles by age is informative, and equally importantly leads to the very tractable distribution represented at (2.2). Previous approaches labeled the new alleles in decreasing order of class size and lead to the Poisson-Dirichlet distribution. For the relationship between the two, see for example Donnelly (1986a).

*Remark 3.* One of the advantages of focusing on genealogy is that the above characterization lends itself readily to calculations. One could study at length the transient age structure of the population, and the joint distribution of population and sample frequencies (i.e. subsampling genealogies). The ideas are analogous to Donnelly (1986a) and we omit the details. Instead, to illustrate calculations, we study  $K$  allele models.

### 3. $K$ -allele models

Now (and throughout this section) consider the case in which each individual in the population is of one of  $K$  possible types, and label these alleles

$A_1, A_2, \dots, A_K$ . Suppose that the (scaled) mutation rate from alleles of type  $A_i$  to  $A_j (i \neq j)$  is given by  $\epsilon_{ij} = \epsilon_j$  for each  $i$ . Kingman (1980) gives a biological justification for this assumption. Historically, many of the fundamental results about the infinite alleles model were obtained by taking suitable limits in a  $K$ -allele model with symmetric mutation ( $\epsilon_j = \epsilon/K$ ) as  $K \rightarrow \infty$ . Here we observe that the structure of the  $K$ -allele models can be recovered from the infinite alleles model and in particular from the genealogical process  $\{M_t, t \geq 0\}$ .

The first step is to give an equivalent version of the  $K$ -allele model. Define  $\theta = \epsilon_1 + \dots + \epsilon_K$ . Now consider a  $K$ -allele model in which mutations occur at rate  $\theta$ , and given that a mutation occurs, the resulting offspring will be of type  $A_j$  with probability  $\epsilon_j/\theta, j = 1, 2, \dots, K$ . This is readily seen to be equivalent to the original model; we have simply introduced the device of mutation from a type to itself. (A more detailed version of this argument in the finite population setting is given in Donnelly (1986b).) The point is that the process can now be constructed genealogically. Simply put down the genealogy for an infinite alleles model with mutation rate  $\theta$ , then consider each novel allele (in the infinite alleles model) in turn and independently reclassify it to be of type  $A_j$  with probability  $\epsilon_j/\theta, j = 1, 2, \dots, K$ .

The distribution of the  $K$ -allele model evolving forward in time now follows from our knowledge of  $\{M_t, t \geq 0\}$ . Fix  $t > 0$ . The joint distribution of the proportions of the population who are of types  $A_1, \dots, A_K$  respectively, as the result of the mutation in  $(0, t)$ , will be the same as the joint distributions of  $Y_1, Y_2, \dots, Y_K$  say, where for  $j = 1, 2, \dots, K, Y_j$  is the sum of the frequencies of the new classes of  $M_t$  which are relabelled  $j$  in the above (multinomial  $(\epsilon_1/\theta, \dots, \epsilon_K/\theta)$ ) relabelling. Take a collection of frequencies  $(x_1, x_2, \dots)$  with distribution of the form (2.2). As we are going to relabel the points anyway, rewrite these in descending order as a point  $(y_1, y_2, \dots), y_1 \geq y_2 \geq y_3 \geq \dots \geq 0$  of  $\Delta^1$ . In this form the point has a Poisson-Dirichlet distribution and (among other things) may be represented as  $(z_1/\sigma, z_2/\sigma, \dots)$  where  $z_1 \geq z_2 \geq z_3 \geq \dots$  are the points of a non-homogeneous Poisson process on  $(0, \infty)$  with rate  $\theta z^{-1} e^{-z}$ , and

$$\sigma = \sum_{i=1}^{\infty} z_i,$$

which is almost surely finite. See Kingman (1978). Now independently label each point of the Poisson process in such a way that  $\mathbb{P}(\text{point } z_i \text{ is labelled } j) = \epsilon_j/\theta, j = 1, 2, \dots, K; i = 1, 2, 3, \dots$ . Denote by  $(z_1^{(j)}, z_2^{(j)}, z_3^{(j)}, \dots)$  the points labelled  $j, j = 1, 2, \dots, K$ , again written in decreasing order. It is well known in the theory of Poisson processes (see for example Karr (1986)) that the  $K$  collections of points  $(z_1^{(1)}, z_2^{(1)}, \dots), (z_1^{(2)}, z_2^{(2)}, \dots), \dots, (z_1^{(K)}, z_2^{(K)}, \dots)$  form  $K$  independent Poisson processes rates  $\epsilon_1 z^{-1} e^{-z}, \epsilon_2 z^{-1} e^{-z}, \dots, \epsilon_K z^{-1} e^{-z}$  respectively.

We now derive the distribution of

$$\sigma_j = \sum_{i=1}^{\infty} z_i^{(j)}.$$

Note that

$$E(e^{-t\sigma_j}) = E\left(\prod_{i=1}^{\infty} \exp(-tz_i^{(j)})\right),$$

and this last is just the probability generating functional of the process evaluated at the function  $\exp(-t \cdot)$ . Hence

$$E \left[ \prod_{i=1}^{\infty} \exp(-tz_i^{(j)}) \right] = \exp \left[ \int_0^{\infty} (1 - e^{-tz}) \varepsilon_j z^{-1} e^{-z} dz \right] = (1+t)^{-\varepsilon_j},$$

and thus it follows that  $\sigma_j$  has a Gamma distribution with index  $\varepsilon_j$  and scale parameter 1. Hence  $(\sigma_1/\sigma, \sigma_2/\sigma, \dots, \sigma_K/\sigma)$  has a Dirichlet distribution with parameters  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K)$ . (See Johnson and Kotz (1972, p. 321, and recall that  $(\sigma_1, \sigma_2, \dots, \sigma_K)$  are mutually independent).

In our  $K$ -allele population at time  $t$  there will be a random number,  $l$  say (with the same distribution as  $D_t$ ) of old classes, each of which consists of individuals who are descended without mutation from a single ancestor in the population at time 0. Denote the frequencies of the  $l$  classes by  $X_1, X_2, \dots, X_l$ . All of the individuals of a particular old class will be of the same genetic type as the corresponding ancestor. It follows from the above and the earlier characterization of the distribution of  $\mathfrak{M}_l$  that

$$(X_1, X_2, \dots, X_l; Y_1, Y_2, \dots, Y_K) =^d (U_1 V_l, \dots, U_l V_l; (1 - V_l) W_1, \dots, (1 - V_l) W_K),$$

where  $(U_1, \dots, U_l)$  and  $(W_1, \dots, W_K)$  have independent Dirichlet distributions with parameters  $(1, 1, \dots, 1)$  and  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K)$  respectively, each independent of  $V_l$  which has the distribution (2.1) with  $k$  replaced by  $l$ , and  $=^d$  denotes equality in distribution.

In the sequel it will be convenient to have a compact notation for Dirichlet distributions. Random variables  $W_1, \dots, W_p$  will be said to have a Dirichlet distribution  $D_p(\tau_1, \dots, \tau_p)$  if  $W_p = 1 - W_1 - \dots - W_{p-1}$ , and  $W_1, \dots, W_{p-1}$  have joint probability density

$$\frac{\Gamma(\tau_1 + \dots + \tau_p)}{\Gamma(\tau_1) \dots \Gamma(\tau_p)} w_1^{\tau_1-1} \dots w_{p-1}^{\tau_{p-1}-1} (1 - w_1 - \dots - w_{p-1})^{\tau_p-1}$$

on  $\{0 < w_i, i = 1, \dots, p-1; w_1 + \dots + w_{p-1} < 1\}$ . It then follows by a change of variables argument that

$$(X_1, \dots, X_l, Y_1, \dots, Y_K) \text{ has } D_{l+K}(1, \dots, 1, \varepsilon_1, \dots, \varepsilon_K) \tag{3.1}$$

distribution.

Now assume that at time  $t$  the number of old classes is  $D_t = l$ . Suppose the initial distribution assigns labels (types) independently to each member of the population in such a way that at time zero any particular individual will be of type  $A_1, A_2, \dots, A_K$  with probability  $p_1, p_2, \dots, p_K$  respectively. Since individuals in the old classes inherit the type of their appropriate ancestor, each old class at time  $t$  will consist of individuals all of whom are of type  $A_j$  with probability  $p_j, j = 1, 2, \dots, K$ , independently for each class. Let  $Y'_i$  be the total of the frequencies of the old classes which are assigned label  $i, i = 1, 2, \dots, K$ . The number of classes so labelled has a multinomial distribution, and if  $m_i (\geq 1)$  are labelled  $i$ , a simple symmetry argument in (3.1) shows that  $(Y'_1, \dots, Y'_K, Y_1, \dots, Y_K)$  has the  $D_{2K}(m_1, \dots, m_K, \varepsilon_1, \dots, \varepsilon_K)$  distribution.



Hence  $(Y'_1 + Y_1, \dots, Y'_K + Y_K)$  has the  $D_K(m_1 + \varepsilon_1, \dots, m_K + \varepsilon_K)$  distribution. This last statement holds even if some of the  $m_i$  are zero. Thus the probability density of the  $K$  allele frequencies at time  $t$ , given initial frequencies  $p_1, \dots, p_K$ , and given that  $D_i = l$ , has the form

$$\sum \frac{l!}{m_1! \dots m_K!} p_1^{m_1} \dots p_K^{m_K} \frac{\Gamma(\theta + l)}{\Gamma(\varepsilon_1 + m_1) \dots \Gamma(\varepsilon_K + m_K)} x_1^{m_1 + \varepsilon_1 - 1} \dots s_K^{m_K + \varepsilon_K - 1},$$

the sum being over all non-negative  $m_1, \dots, m_K$  with sum  $l$ . This last expression, together with the distribution of  $\{D_i\}$  given earlier, gives the transition density of the  $K$ -allele model forward in time. It was obtained by other means (for the diffusion model) in Griffiths and Li (1983 equation (8)); see also Griffiths (1979).

*Acknowledgements.* The authors were supported in part by NSF grants DMS 85-01763 and DMS 86-08857.

**Appendix**

Our purpose here is to prove the existence of the discrete time Markov chain  $\mathfrak{M}$ , described in Sect. 2 and to verify that the marginal distribution of  $\mathfrak{M}_k$  is in fact given by  $\nu_k, k = 0, 1, 2, \dots$ . Since the state space  $\mathbb{N}_\infty \times \Delta$  of  $\mathfrak{M}_k$  is compact, both of these conclusions will follow from the Kolmogorov extension theorem if we can show that the distributions  $\nu_k$  are consistent with the transition mechanism of  $\mathfrak{M}$ .

Suppose that there are  $k$  old classes, with total frequency  $V_k$  having a beta distribution with density

$$f_k(x) = \frac{\Gamma(k + \theta)}{\Gamma(k)\Gamma(\theta)} x^{k-1}(1-x)^{\theta-1}, \quad 0 < x < 1. \tag{A1}$$

The frequencies in the  $k$  old classes have representation

$$V_k(U_1, U_2, \dots, U_k) \tag{A2}$$

where  $(U_1, U_2, \dots, U_k)$  has a Dirichlet distribution  $D_k(1, 1, \dots, 1)$ , independent of  $V_k$ , whereas the new classes, in decreasing age order, have the representation

$$(1 - V_k)(Z_1, (1 - Z_1)Z_2, (1 - Z_1)(1 - Z_2)Z_3, \dots) \tag{A3}$$

where the  $\{Z_i\}$  are independent, identically distributed random variables with the beta density  $\theta(1-x)^{\theta-1}, 0 < x < 1$ , independent of  $(U_1, U_2, \dots, U_k)$  and  $V_k$ .

We will show first that when the number of old classes decreases by one to  $k-1$ , then the total frequency  $V_{k-1}$  in the old classes has the distribution (A1) with  $k$  replaced by  $k-1$ .

A transition from  $k$  to  $k-1$  can occur in two essentially different ways. For some  $1 \leq i < j \leq k$ , the  $i$ th and  $j$ th old classes may coalesce. This event,  $E_{ij}$  say, has probability  $2/(k(k + \theta - 1))$ , and in this case  $V_{k-1} = V_k$ . Alternatively, the  $i$ th old class is chosen to become the oldest of the new classes. This event,  $F_i$  say, occurs with probability  $\theta/k(k + \theta - 1)$ , and in this case  $V_{k-1} = V_k(1 - U_i)$ . Hence the  $r$ th moment of  $V_{k-1}$  is

$$\begin{aligned} \mathbb{E}(V_{k-1}^r) &= \frac{k-1}{k+\theta-1} \mathbb{E}(V_k^r) + \sum_{i=1}^k \frac{\theta}{k(k+\theta-1)} \mathbb{E}((V_k(1-U_i))^r) \\ &= \frac{k-1}{k+\theta-1} \mathbb{E}(V_k^r) + \frac{\theta}{k+\theta-1} \mathbb{E}(V_k^r) \mathbb{E}((1-U_1)^r) \\ &= \left( \frac{k-1}{k+\theta-1} + \frac{\theta}{k+\theta-1} \frac{k-1}{k+r-1} \right) \mathbb{E}(V_k^r) \\ &= \frac{k-1}{k+\theta-1} \frac{k+r+\theta-1}{k+r-1} \frac{\Gamma(k+\theta)\Gamma(k+r)}{\Gamma(k)\Gamma(r+k+\theta)} \end{aligned}$$

$$= \frac{\Gamma(k-1+\theta)\Gamma(k-1+r)}{\Gamma(k-1)\Gamma(r+k-1+\theta)}, \quad r=0, 1, \dots$$

These are the moments of the density  $f_{k-1}(x)$ ; since this density is determined by its moments, we are finished.

Next we establish that once the number of old classes has decreased from  $k$  to  $k-1$ , the joint distribution of the  $k-1$  old frequencies has the representation

$$V_{k-1}(U'_1, \dots, U'_{k-1})$$

where  $(U'_1, \dots, U'_{k-1})$  is independent of  $V_{k-1}$ , and has Dirichlet distribution  $D_{k-1}(1, \dots, 1)$ . Conditional on a coalescence occurring, we know that  $V_{k-1} = V_k$ , and that  $(U'_1, \dots, U'_{k-1})$  is formed by summing a randomly chosen pair of elements from  $(U_1, \dots, U_k)$ . The resultant  $(U'_1, \dots, U'_{k-1})$  has the required properties, as was shown in Kingman (1982c), p. 244. On the other hand, conditional on the  $i$ th old class being relabelled as the oldest of the new classes the old frequencies become

$$V_k(U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_k) \\ = V_k(1-U_i) \left( \frac{U_1}{(1-U_i)}, \dots, \frac{U_{i-1}}{(1-U_i)}, \frac{U_{i+1}}{(1-U_i)}, \dots, \frac{U_k}{(1-U_i)} \right).$$

It may readily be verified that  $1-U_i$  is independent of the random vector on the right, and that this vector has the distribution  $D_{k-1}(1, \dots, 1)$ . Since  $V_{k-1} = V_k(1-U_i)$ , this establishes our claim.

Finally, we need to show that after  $k$  old classes have been reduced to  $k-1$  the new frequencies have the same probabilistic structure as (A3). This is clear if the jump from  $k$  to  $k-1$  occurred via a coalescence. If the jump occurred from the reclassification of an old class as a new one then  $V_{k-1} = V_k(1-U_i)$  for some  $i, 1 \leq i \leq k$ , and the new frequencies are

$$V_k U_i, (1-V_k)Z_1, (1-V_k)(1-Z_1)Z_2, \dots \\ = (1-V_k(1-U_i)) \left[ \frac{V_k U_i}{1-V_k+V_k U_i}, \left[ 1 - \frac{V_k U_i}{1-V_k+V_k U_i} \right] (Z_1, (1-Z_1)Z_2, \dots) \right] \\ = (1-V_{k-1})(Z_0, (1-Z_0)Z_1, (1-Z_0)(1-Z_1)Z_2, \dots),$$

where

$$Z_0 = \frac{V_k U_i}{1-V_k+V_k U_i}.$$

The random variables  $V_k U_1, \dots, V_k U_k, 1-V_k$  have a Dirichlet distribution with parameters  $1, \dots, 1$  and  $\theta$ , so it is straightforward to show that  $Z_0$  has density  $\theta(1-x)^{\theta-1}, 0 < x < 1$ , and that  $Z_0$  is independent of  $1-V_k+V_k U_i \equiv 1-V_{k-1}$ . Thus the frequencies in the new classes have the required structure, and the proof is complete.

**References**

Donnelly, P.: Partition structures, Polya urns, the Ewens sampling formula and the ages of alleles. *Theor. Popul. Biol.* **30**, 271-288 (1986a)

Donnelly, P.: Dual processes in population genetics. In: Tautu, P. (ed.) *Stochastic spatial processes*, (Lect. Notes Math., Vol. 1212, pp. 94-105) Berlin, Heidelberg, New York: Springer (1986b)

Donnelly, P., Tavaré S.: The ages of alleles and a coalescent. *Adv. Appl. Probab.* **18**, 1-19 (1986)

Ethier, S. N., Kurtz, T. G.: The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429-452 (1981)

Ethier, S. N., Kurtz, T. G.: *Markov processes: characterization and convergence*. New York: Wiley (1986)

Griffiths, R. C.: A transition density expansion for a multi-allele diffusion model. *Adv. Appl. Probab.* **11**, 310-325 (1979)

Griffiths, R. C.: Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theor. Popul. Biol.* **17**, 37-50 (1980)

- Griffiths, R. C., Li, W.-H.: Simulating allele frequencies in a population, and the genetic differentiation of populations under mutation pressure. *Theor. Popul. Biol.* **23**, 19-33 (1983)
- Johnson, N. L., Kotz, S.: Continuous multivariate distributions. New York: Wiley 1972
- Karr, A. F.: Point processes and their statistical inference. New York: Dekker 1986
- Kingman, J. F. C.: Random partitions in population genetics. *Proc. Roy. Soc. A* **361**, 1-20 (1978)
- Kingman, J. F. C.: Mathematics of genetic diversity. CBMS Regional Conference in Applied Mathematics, 34. S.I.A.M. (1980)
- Kingman, J. F. C.: On the genealogy of large populations, *Adv. Appl. Probab.* **19A**, 27-43 (1982a)
- Kingman, J. F. C.: The coalescent. *Stochastic Processes Appl.* **13**, 235-248 (1982b)
- Kingman, J. F. C.: Exchangeability and the evolution of large populations. In: Koch, F., Spizzichino, I. (eds.) Exchangeability in probability and statistics, pp. 97-112. Amsterdam: North Holland (1982c)
- Tavaré, S.: Line of descent and genealogical processes and their application in population genetics models. *Theor. Popul. Biol.* **26**, 119-164 (1984)
- Watterson, G. A.: Lines of descent and the coalescent. *Theor. Popul. Biol.* **26**, 77-92 (1984)

Received January 5/Revised April 24, 1987