

## ON THE GENEALOGY OF NESTED SUBSAMPLES FROM A HAPLOID POPULATION

IAN W. SAUNDERS,\* *CSIRO Division of Mathematics and Statistics*

SIMON TAVARÉ,\*\* *Colorado State University*

G. A. WATTERSON,\*\*\*, *Monash University*

### Abstract

For the haploid genetic model of Moran, the joint distribution of the numbers of distinct ancestors of a collection of nested subsamples is derived. These results are shown to apply to the diffusion approximations of a wide variety of other genetic models, including the Wright–Fisher process. The results allow us to relate the ancestries of populations sampled at different times. Analogous results for a line-of-descent process that incorporates the effect of mutation are given. Some results about the ages of alleles in an infinite-alleles model are described.

MORAN MODEL; WRIGHT–FISHER MODEL; INFINITE ALLELES; AGE OF ALLELES;  
LINES OF DESCENT; DIFFUSION APPROXIMATION

### 1. Introduction

Consider the evolution of a haploid population comprising  $2N$  individuals in each generation. Assuming for the moment that generations are distinct, each individual in generation  $s+1$  is the offspring of exactly one individual in generation  $s$ . The number of offspring born to individual  $j$  in generation  $s$  is, however, a random variable  $Y_j$ , say, constant population size being maintained by the requirement that  $Y_1 + Y_2 + \dots + Y_{2N} = 2N$ . The random variables  $Y_1, \dots, Y_{2N}$  are assumed to be exchangeable, with a joint distribution that is invariant over time. Suppose that we now choose, without replacement, a random sample of size  $i$  from the population at some time labelled 0. Labelling time in the reverse direction, we observe the number of distinct ancestors  $A_1(n)$  of the sample  $n$  generations before now. Thus  $A_1(1)$  is the number of

---

Received 20 June 1983; revision received 10 November 1983.

\* Postal address: CSIRO Division of Mathematics and Statistics, Private Bag 10, Clayton, VIC 3168, Australia.

\*\* Postal address: Department of Statistics, Colorado State University, Fort Collins, CO 80532, USA.

\*\*\* Postal address: Department of Mathematics, Monash University, Clayton, VIC 3168, Australia.

distinct parents of our sample of size  $i$ ,  $A_1(2)$  the number of distinct grandparents, and so on.

Several authors have studied the behavior of the process  $\{A_1(n), n \geq 0\}$ . In particular, Felsenstein (1971) uses the process to determine the rate of loss of alleles in a class of neutral haploid models with no mutation. For our purposes, it is enough to record that  $\{A_1(n), n \geq 0; A_1(0) = i\}$  is a Markov chain with state space  $\{1, 2, \dots, i\}$  whose one-step transition probabilities

$$(1.1) \quad g_{ij} \equiv g_{ij}(1) = \mathbb{P}(A_1(n+1) = j \mid A_1(n) = i)$$

can be computed explicitly: see Gladstien (1978). We shall cite one important example.

The Wright–Fisher genetic model asserts that  $(Y_1, \dots, Y_{2N})$  have a joint multinomial distribution with

$$(1.2) \quad \mathbb{P}(Y_1 = y_1, \dots, Y_{2N} = y_{2N}) = \binom{2N}{\mathbf{y}} \left(\frac{1}{2N}\right)^{2N}.$$

The prescription (1.2) is equivalent to each individual choosing his parent independently uniformly at random from the parent population. Watterson (1975) used the fact that in this case

$$(1.3) \quad g_{ij} = (2N)^{-l} \binom{2N}{j} \Delta^j 0^l, \quad 1 \leq j \leq l,$$

where

$$\Delta^j f(x) = \sum_{n=0}^j (-1)^{j+n} \binom{j}{n} f(x+n).$$

The formula (1.3) will be familiar as the distribution of the number of occupied cells (parents) when  $l$  balls (individuals) are dropped at random into  $2N$  cells.

Several choices of distribution for  $(Y_1, \dots, Y_{2N})$  result in models in which the generations may be considered overlapping. One such is due to Moran (1958), where

$$(1.4) \quad (Y_1, \dots, Y_{2N}) = \begin{cases} (0, 2, 1, \dots, 1), \text{ or a permutation} \\ \text{of this, with probability } 1 - (2N)^{-1} \\ (1, 1, 1, \dots, 1) \text{ with probability} \\ (2N)^{-1} \end{cases}$$

(cf. Cannings (1974)). In this model, the ancestor of an individual may be that individual itself if it survives. While the natural time-scale for this process is in units of birth–death events, we will for uniformity continue to call these

generations. It follows from (1.4) that

$$(1.5) \quad g_{ii} = 1 - \frac{l(l-1)}{4N^2} = 1 - g_{i,l-1}.$$

In all these models state 1 is absorbing, corresponding to the observation that eventually the sample, or population, can be traced back to a single ancestor.

For the process specified by (1.5) it is a straightforward matter to compute the transition probabilities  $g_{ii}(n) = \mathbb{P}(A_1(n) = l \mid A_1(0) = i)$ . We obtain

$$(1.6) \quad g_{ii}(n) = \sum_{r=1}^i \lambda_r^n \frac{(-1)^{r-l} (2r-1) i_{[r]} l_{(r-1)}}{l! (r-l)! i_{(r)}}, \quad 1 \leq l \leq i,$$

where  $\lambda_r = 1 - (r(r-1)/4N^2)$ , and we use the notation

$$a_{[r]} = a(a-1) \cdots (a-r+1), \quad a_{(r)} = a(a+1) \cdots (a+r-1).$$

It is interesting to note that the distribution of the number of distinct ancestors of a sample of any size  $i \leq 2N$  can be found by studying the single process  $A_1(\cdot)$ . Thus far, however, we have considered the ancestry of only a single sample.

In this paper, we study the joint distribution of the number of distinct ancestors of a sequence of nested subsamples taken from a population reproducing according to the Moran model (1.4). Such nested subsamples arise directly when making inferences about a large sample from ancestral information concerning a subsample. They also arise indirectly in the study of the relationship between the ancestries of distinct generations. See Watterson (1982a). This joint distribution is derived in Section 2, and some examples of its application are given in Section 3.

The corresponding exact results for more complicated discrete time models such as the Wright–Fisher process (1.2) are unmanageable, and so in Section 4, we use results of Kingman (1982a) to find large population approximations in these cases. The jump chain of the continuous-time approximating process has the structure of the discrete-time Moran model, and the results of Section 2 become applicable.

In Section 5, we introduce a process incorporating mutation, which is an extension of a model due to Griffiths (1980), and study subsampling in the lines of descent for this process. Some application to the infinite-alleles model, and the ages of alleles, are given.

## 2. The ancestry of a subsample

Suppose that at generation 0 we randomly select a sample of  $i$  individuals from the population and then select a further subsample of size  $j$  from the  $i$ .

How is the process of ancestors of the subsample related to that of the whole sample?

We shall use the following notation:

$A_1(n)$  = number of distinct ancestors of members of the sample  $n$  generations before the present.

$A_2(n)$  = corresponding number for the subsample.

$P_n(i, j; l_1, l_2) = \mathbb{P}(A_1(n) = l_1, A_2(n) = l_2 \mid A_1(0) = i, A_2(0) = j)$

$g_{il}(n) = \mathbb{P}(A_1(n) = l \mid A_1(0) = i)$  (given by (1.6))

$N_r(l) = \min \{n : A_r(n) = l\}, \quad r = 1, 2$

$A_2^*(l) = A_2(N_1(l)).$

Thus  $A_2^*(l)$  is the number of distinct ancestors of the subsample at the most recent time that the sample has  $l$  distinct ancestors. By the nature of the process,  $A_2^*(l)$  is also the number of distinct ancestors of the subsample whenever the sample has  $l$  distinct ancestors.

The following lemma is the basis of most of our results.

*Lemma 1.* Given  $A_1(0) = i, A_2(0) = j, \{A_2^*(i - l_1), l_1 = 0, \dots, i - 1\}$  forms an inhomogeneous Markov chain with transition probabilities given by

$$\begin{aligned}
 \mathbb{P}(A_2^*(l-1) = k-1 \mid A_2^*(l) = k) \\
 &= 1 - \mathbb{P}(A_2^*(l-1) = k \mid A_2^*(l) = k) \\
 (2.1) \quad &= \frac{k(k-1)}{l(l-1)}.
 \end{aligned}$$

*Proof.* Since the  $A_2(n)$  ancestors of the subsample are a subset of the  $A_1(n)$  ancestors of the sample it is clear that  $A_2(n+1) = A_2(n)$  whenever  $A_1(n+1) = A_1(n)$ . Further, since the  $j$  members of the subsample were randomly selected from the sample, their  $A_2(n)$  ancestors form a random subset of the  $A_1(n)$  ancestors of the sample. Thus if  $A_1(n+1) = A_1(n) - 1$ , so that two of the  $A_1(n)$  ancestors have a common parent, the probability that these two are both in  $A_2(n)$ , so that  $A_2(n+1) = A_2(n) - 1$ , is

$$\binom{A_2(n)}{2} \bigg/ \binom{A_1(n)}{2} = \frac{A_2(n)(A_2(n)-1)}{A_1(n)(A_1(n)-1)}.$$

That is,

$$\begin{aligned}
 \mathbb{P}(A_2(n+1) = k-1 \mid A_1(n) = l, A_1(n+1) = l-1, A_2(n) = k) \\
 (2.2) \quad &= \frac{k(k-1)}{l(l-1)}.
 \end{aligned}$$

But by the definition of  $A_2^*$ , and since  $A_2$  changes only when  $A_1$  does, this is precisely

$$\mathbb{P}(A_2^*(l-1) = k-1 \mid A_2^*(l) = k) = \frac{k(k-1)}{l(l-1)}.$$

The probability (2.2) is unaffected by conditioning on the values of  $A_1(0), \dots, A_1(n-1)$  and  $A_2(0), \dots, A_2(n-1)$ , and so the Markov property holds for the sequence

$$A_2^*(i), A_2^*(i-1), \dots, A_2^*(1).$$

Clearly  $A_2^*(l) - A_2^*(l-1)$  is at most 1, and so the complementary probabilities in (2.1) follow immediately.

Note that we have nowhere used the transition probabilities for the process  $\{A_1(n)\}$ . Thus these results will hold for a variety of such processes and in particular when  $\{A_1\}$  is a continuous-time process. This fact will be used in Section 4 below.

Using Lemma 2.1 we obtain the conditional distribution of  $A_2(n)$  given  $A_1(n)$ , which is the main result of this section.

*Theorem 2.*

$$(2.3) \quad \begin{aligned} &\mathbb{P}(A_2(n) = l_2 \mid A_1(n) = l_1, A_1(0) = i, A_2(0) = j) \\ &= \frac{(i-j)! (i-l_1)! j! (j-1)! l_1! (l_1-1)! (i+l_2-1)!}{(j-l_2)! (l_1-l_2)! i! (i-1)! l_2! (l_2-1)! (l_1+j-1)! (i+l_2-l_1-j)!} \end{aligned}$$

*Remark.* Surprisingly,  $j$  and  $l_1$  play interchangeable roles in this expression, and it does not depend on  $n$ .

*Proof.* Clearly the required probability is

$$\begin{aligned} &\mathbb{P}(A_2^*(l_1) = l_2 \mid A_2(0) = j, A_1(0) = i) \\ &= \phi_{ij}(l_1, l_2), \text{ say.} \end{aligned}$$

Usually we shall suppress the dependence of  $\phi$  on  $i$  and  $j$ . The forward equations now give the recurrence

$$(2.4) \quad \phi(l_1, l_2) = \phi(l_1+1, l_2+1) \frac{l_2(l_2+1)}{l_1(l_1+1)} + \phi(l_1+1, l_2) \left(1 - \frac{l_2(l_2-1)}{l_1(l_1+1)}\right)$$

with boundary conditions

$$\begin{aligned} \phi(i, l_2) &= 1 & l_2 = j \\ &= 0 & \text{otherwise.} \end{aligned}$$

The recurrence can be solved recursively for  $\phi(i-1, l_2)$ ,  $l_2=1, \dots, j$ ,  $\phi(i-2, l_2)$ ,  $l_2=1, \dots, j$  and so on. A simple inductive argument shows that (2.3) gives the general solution.

Combining this result with (1.6) we obtain the joint distribution of  $A_1(n)$  and  $A_2(n)$  as

$$(2.5) \quad P_n(i, j; l_1, l_2) = g_{il_1}(n)\phi(l_1, l_2).$$

The bivariate process  $\{(A_1(n), A_2(n))\}$  itself forms a Markov chain with transition probabilities

$$(2.6) \quad P_1(i, j; l_1, l_2) = \begin{cases} 1 - \frac{i(i-1)}{4N^2}, & l_1 = i, l_2 = j \\ \frac{i(i-1) - j(j-1)}{4N^2}, & l_1 = i-1, l_2 = j \\ \frac{j(j-1)}{4N^2}, & l_1 = i-1, l_2 = j-1 \\ 0 & \text{otherwise,} \end{cases}$$

cf. Watterson (1982b).

Note that  $A_1(N_2(k))$  is the number of distinct ancestors of the sample when the subsample first has exactly  $k$  distinct ancestors. That is

$$A_1(N_2(k)) = \max \{l : A_2^*(l) = k\}.$$

Using (2.3) we can obtain the distribution of this quantity.

*Lemma 3.*

$$(2.7) \quad \begin{aligned} \mathbb{P}(A_1(N_2(k)) = l \mid A_1(0) = i, A_2(0) = j) \\ = \frac{(i-j)! (i-l-1)! j! (j-1)! l! (l-1)! (i+k)!}{(j-k-1)! (l-k)! i! (i-1)! k! (k-1)! (l+j)! (i+k-l-j)!}. \end{aligned}$$

*Proof.*

$$\begin{aligned} \mathbb{P}(A_1(N_2(k)) = l \mid A_1(0) = i, A_2(0) = j) \\ = \mathbb{P}(A_2^*(l+1) = k+1, A_2^*(l) = k) \\ = \phi(l+1, k+1) \frac{k(k+1)}{l(l+1)}. \end{aligned}$$

The result follows from (2.3).

Taking  $k = 1$  we obtain the conditional distribution of  $A_1$  when  $A_2$  first hits state 1.

Corollary 4.

$$\mathbb{P}(A_1(N_2(1)) = l \mid A_1(0) = i, A_2(0) = j) = (j - 1)(i + 1) \frac{l! j!}{(l + j)!} \frac{(i - l - 1)! (i - j)!}{(i - 1)! (i - l - j + 1)!}.$$

This is the quantity  $y_{i,j;l}$  of Watterson (1982b).

The formula (2.3) extends in an obvious manner to an extended model where we take  $q$  nested samples  $S_1 \supseteq S_2 \supseteq \dots \supseteq S_q$  with  $|S_r| = i_r$ . If we now write

$$\begin{aligned} A_r(n) &= \text{number of ancestors of sample } S_r \text{ } n \text{ generations back,} \\ N_r(l) &= \min \{n : A_r(n) = l\}, \\ A_{r+1}^*(l) &= A_{r+1}(N_r(l)) \end{aligned}$$

then Lemma 1 shows that  $\{A_r^*\}$  is a Markov chain. Clearly, given  $\{A_r^*\}$ ,  $\{A_{r-1}^*\}$  is independent of  $\{A_{r+1}^*\}$  and so by repeated application of Theorem 2 we have the following result.

Theorem 5.

$$\begin{aligned} \mathbb{P}(A_r(n) = l_r, r = 2, \dots, q \mid A_1(n) = l_1, A_r(0) = i_r, r = 1, \dots, q) \\ = \frac{(i_1 - l_1)! i_q! (i_q - 1)! l_1! (l_1 - 1)!}{(i_q - l_q)! i_1! (i_1 - 1)! l_q! (l_q - 1)!} \\ \times \prod_{r=1}^{q-1} \frac{(i_r - i_{r+1})! (i_r + l_{r+1} - 1)!}{(l_r - l_{r+1})! (l_r + i_{r+1} - 1)! (i_r - l_r - i_{r+1} + l_{r+1})!}. \end{aligned}$$

The joint distribution of the Markov chain  $(A_1(n), \dots, A_q(n))$  can be obtained as an expression analogous to (2.5).

3. Applications of the subsampling formula

In this section we use formula (2.3) to investigate the relationships between the ancestries of a sample and a subsample and of populations observed at different times.

Example 1. Suppose we are interested in the time  $N_1(1)$  back to a single ancestor for a sample of size  $i$ , but are able to determine only the time  $N_2(1)$  for a smaller subsample of size  $j$ . What is the distribution of  $N_1(1)$  given  $N_2(1)$ ?

From Corollary 4 we have

$$\begin{aligned} \mathbb{P}(N_1(1) = n_1 + n_2 \mid N_2(1) = n_2) \\ = \sum_l \mathbb{P}(N_1(1) = n_1 + n_2 \mid A_1(n_2) = l) \mathbb{P}(A_1(n_2) = l \mid N_2(1) = n_2) \\ = \sum_l \frac{1}{2N^2} g_{l2}(n_1 - 1) y_{ij;l} \quad \text{for } n_1 \geq 1, \end{aligned}$$

while for  $n_1 = 0$  we obtain

$$(3.1) \quad \begin{aligned} \mathbb{P}(N_1(1) = n_2 \mid N_2(1) = n_2) &= y_{ij;1} \\ &= \frac{j-1}{j+1} \frac{i+1}{i-1}, \end{aligned}$$

giving even for small  $j$  a very high probability that  $N_1(1) = N_2(1)$ .

For  $n_1 > 0$  the probabilities are of order  $N^{-2}$ ; for example

$$\mathbb{P}(N_1(1) = N_2(1) + 1) = \frac{1}{N^2} \frac{(j-1)(i+1)(i-j)}{(j+2)(j+1)(i-1)(i-2)}.$$

*Example 2.* The above examples indicate that the common ancestor of a subsample has a high probability of being also the common ancestor of the whole sample. When this occurs, it implies that in some generation the sample and the subsample had the same number  $l > 1$  of ancestors. Clearly the  $l$  ancestors of the subsample were the same individuals as those of the sample and so the numbers of ancestors must be equal in all earlier generations: the ancestries have ‘coupled’. Note that if  $l = 1$  then the common ancestor of the sample occurred prior to that of the subsample and we do not consider this to be coupling. The distribution of the common number  $L$  of ancestors in the most recent coupled generation is

$$\begin{aligned} \mathbb{P}(L = l) &= c_l \\ &= \phi_{i,j}(l, l) - \phi_{i,j}(l+1, l+1) \\ &= 2l(i-j) \frac{j! (j-1)! (i-l-1)! (i+l-1)!}{i! (i-1)! (j+l)! (j-l)!}, \quad \text{by (2.3),} \end{aligned}$$

and  $c_2 + \dots + c_j$  is given by (3.1). Considering  $c_l/c_{l-1}$  we find a mode of the distribution of  $L$  to be at the largest integer  $\hat{l}$  such that

$$\hat{l}(\hat{l}-1) \leq \frac{ij}{2(i-j)-1}.$$

When  $i$  is large and  $j/i = \alpha$  we find

$$\hat{l} \approx \sqrt{\frac{\alpha}{2(1-\alpha)}} \sqrt{i}.$$

and

$$\begin{aligned} c_1 &= 1 - \mathbb{P}(\text{coupling}) \\ &\approx \frac{2(1-\alpha)}{\alpha i} \\ &\approx \hat{l}^{-2}. \end{aligned}$$

*Example 3.* Watterson (1982b) considers, for the Wright–Fisher model, the sequence  $T_n$  where  $T_n$  is the most recent generation in which the individuals of



generation  $n$  all had a common ancestor. This sequence is related to the process of mutant substitutions. Here we consider the analogous quantities in the Moran model.

It is clear that the ancestors of generation 0 in generation  $m$  are a subset of those of generation 1, and so on. Since by the Markov property the ancestry prior to generation  $m$  is independent of that subsequent to generation  $m$  we can consider the ancestors of generation 0, generation 1,  $\dots$ , generation  $m - 1$  in generation  $m$ , to form nested random subsets selected from generation  $m$ . Given the sizes of these subsets we can obtain the joint distribution of  $T_0, \dots, T_m$  from repeated application of Theorem 2. We assume that  $m < 2N$ , so that  $T_0 > m$ . If there are  $I_r^{(m)}$  ancestors of generation  $r$  in generation  $m$  and  $A_r(n)$  denotes the number of ancestors of generation  $r$  in generation  $m + n$ , we have, given  $I_r^{(m)} \equiv A_r(0) = i_r$ ,

$$\begin{aligned}
 & \mathbb{P}(T_0 = m + t_0, \dots, T_m = m + t_m \mid A_r(0) = i_r, r = 0, \dots, m) \\
 &= \mathbb{P}(A_r(t_r - 1) = 2, A_r(t_r) = 1, r = 0, \dots, m) \\
 &= \sum_{l_s: r < s} \mathbb{P}(A_s(t_s - 1) = l_{rs}, s > r, A_r(t_r - 1) = 2, \\
 &\quad A_r(t_r) = 1, r = 0, \dots, m) \\
 (3.2) \quad &= \frac{1}{(2N^2)^m} \sum_{l_s: r < s} g_{i_m l_{m0}}(t_0 - 1) \phi_{i_m i_{m-1}}(l_{m0}, l_{m-10}) \dots \phi_{i_1 i_0}(l_{10}, 2) \\
 &\quad \times g_{l_{m0} - 1 l_{m1}}(t_1 - t_0 - 1) \phi_{l_{m0} - 1 l_{m-10} - 1}(l_{m1}, l_{m-11}) \\
 &\quad \times \dots \times \phi_{l_{20} - 1 l_{10} - 1}(l_{21}, 2) \\
 &\quad \times \dots \times g_{l_{m-1} - 1, 2}(t_m - t_{m-1} - 1)
 \end{aligned}$$

where the sum is taken over the set  $\{l_{rs} : 2 \leq l_{r+1} \leq \dots \leq l_m, l_{rs} \leq i_s - i_r + 1, r = 0, \dots, m - 1, s = r + 1, \dots, m\}$ . To obtain the unconditional joint distribution of  $T_0, T_1, \dots$ , we must average (3.2) over the distribution of the  $I_r^{(m)}$ . Note that the one-step transition probabilities of the Markov chain  $(A_q(n), \dots, A_0(n))$  where now  $i_q \geq \dots \geq i_0$ , are, from Theorem 5,

$$\begin{aligned}
 & \mathbb{P}(A_r(n + 1) = l_r, r = 0, \dots, q \mid A_r(n) = i_r, r = 0, \dots, q) \\
 &= 1 - \frac{i_q(i_q - 1)}{4N^2}; \quad l_j = i_j, \quad j = 0, \dots, q \\
 (3.3) \quad &= \frac{1}{4N^2} \{i_{r+1}(i_{r+1} - 1) - i_r(i_r - 1)\}; \\
 &\quad l_j = i_j, j = 0, \dots, r; = i_j - 1, j = r + 1, \dots, q \\
 &= \frac{1}{4N^2} i_0(i_0 - 1); \quad l_j = i_j - 1, \quad j = 0, \dots, q \\
 &= 0 \quad \text{otherwise.}
 \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}(I_r^{(k+1)} = l_r, r = 0, \dots, k + 1 \mid I_r^{(k)} = i_r, r = 0, \dots, k) \\ = 0 \quad \text{if } l_{k+1} \neq 2N \end{aligned}$$

and otherwise is given by (3.3).

Thus if we write

$$\psi_k(i_0, \dots, i_k) = \mathbb{P}(I_r^{(k)} = i_r, r = 0, \dots, k)$$

then the  $\psi_k$  satisfy the recurrence

$$\begin{aligned} (3.4) \quad \psi_k(i_0, \dots, i_{k-1}, 2N) &= \frac{1}{2N} \psi_{k-1}(i_0, \dots, i_{k-1}) \\ &+ \frac{1}{4N^2} \sum_{r=0}^{k-1} \psi_{k-1}(i_0, \dots, i_r, i_{r+1} + 1, \dots, i_{k-1} + 1) \\ &\times \{(i_{r+1} + 1)i_{r+1} - i_r(i_r - 1)\} \\ &+ \frac{1}{4N^2} \psi_{k-1}(i_0 + 1, \dots, i_{k-1} + 1)i_0(i_0 + 1). \end{aligned}$$

In principle, (3.4) together with (3.2) gives the joint distribution of  $T_0, \dots, T_m$  for any  $m < 2N$ . In view of the complexity of these formulae, however, we shall consider some special cases where more explicit answers are possible.

*Example 4.* When  $t_0 = t_m = t > 0$  we have, since  $T_0 \leq T_1 \leq \dots \leq T_m$ ,

$$\begin{aligned} \mathbb{P}(T_0 = T_1 = \dots = T_m = t + m) \\ &= \mathbb{P}(T_0 = T_m = t + m) \\ &= \mathbb{P}(T_0 = T_m \mid T_m = t + m)\mathbb{P}(T_m = t + m) \\ &= \mathbb{P}(A_2^*(2) = 2 \mid T_m = t + m, A_2(0) = A_0(m), \\ &\quad A_1(0) = 2N)\mathbb{P}(T_m = t + m) \\ &= \sum_j \phi_{2N,j}(2, 2)g_{2N,j}(m)\mathbb{P}(T_m = t + m) \\ &= \sum_j \frac{2N + 1}{2N - 1} g_{2N,j}(m) \frac{j - 1}{j + 1} \mathbb{P}(T_m = t + m) \\ &= \lambda \frac{m}{2} \mathbb{P}(T_m = t + m) \end{aligned}$$

since as noted by Kingman (1982a)

$$r_2 = \left( \frac{j - 1}{j + 1} \right)$$

is an eigenvector of  $(g_{ij})$ . Its eigenvalue is

$$\lambda_2 = \left(1 - \frac{2}{4N^2}\right)$$

so that

$$\mathbb{P}(T_0 = T_m = t + m) = \frac{1}{2N^2} \left(1 - \frac{1}{2N^2}\right)^m g_{2N,2}(t - 1).$$

Summing over  $t > 0$  gives

$$(3.5) \quad \mathbb{P}(T_0 = T_1 = \dots = T_m) = \left(1 - \frac{1}{2N^2}\right)^m.$$

The latter two results are consistent with  $(T_m, m = 0, 1, 2, \dots)$  being a Markov chain. In fact, however, such is not the case; it may be verified that  $\mathbb{P}(T_2 = t_2 | T_1 = t_1, T_0 = t_0)$  is not independent of  $t_0$ , for instance.

*Example 5.* Extending the result of the above example, we can consider the distribution of  $T_m - T_0$ . Watterson (1982b) showed how this distribution is related to the number of substitutions of mutant nucleotides in generations  $0, \dots, m - 1$ , and obtained an approximation for the Wright–Fisher model. Here we obtain an exact distribution for the Moran model.

The joint distribution of  $T_0$  and  $T_m$ , when  $m < 2N$ , is, for  $t_m > t_0$ ,

$$\begin{aligned} \mathbb{P}(T_0 = t_0, T_m = t_m) &= \sum_k \sum_j g_{2N,j}(m) P_{t_0-m-1}(2N, j; k, 2) P_1(k, 2; k - 1, 1) g_{k-1,2}(t_m - t_0 - 1) g_{21} \\ &= \sum_k \sum_j P_m(2N, k; j, 2) g_{2N,k}(t_0 - m - 1) P_1(k, 2; k - 1, 1) g_{k-1,2}(t_m - t_0 - 1) g_{21} \\ &\quad \text{since } \phi_{2N,j}(k, 2) = \phi_{2N,k}(j, 2) \\ &= \sum_k \frac{1}{4N^4} g_{k,2}(m) g_{2N,k}(t_0 - m - 1) g_{k-1,2}(t_m - t_0 - 1). \end{aligned}$$

Thus for  $t > 0$

$$\begin{aligned} \mathbb{P}(T_m - T_0 = t) &= \sum_{t_0=2N-1}^{\infty} \mathbb{P}(T_0 = t_0, T_m = t_0 + t) \\ &= \frac{1}{4N^4} \sum_{t_0=2N-1}^{\infty} \sum_{k=3}^{m+2} g_{k,2}(m) g_{2N,k}(t_0 - m - 1) g_{k-1,2}(t - 1) \\ &\quad \text{where the range of } k \text{ is chosen to include} \\ &\quad \text{all positive terms,} \\ &= \frac{1}{4N^4} \sum_{k=3}^{m+2} g_{k,2}(m) g_{k-1,2}(t - 1) \sum_{t_0=2N-m-2}^{\infty} g_{2N,k}(t_0). \end{aligned}$$

Now since  $k \leq m + 2$ ,

$$\sum_{t_0=2N-m-2}^{\infty} g_{2N,k}(t_0) = \mathbb{E} \left\{ \sum_{t=0}^{\infty} I(A_1(t) = k) \mid A_1(0) = 2N \right\},$$

the mean number of generations that  $A_1(\cdot)$  spends in state  $k$ . But since  $g_{kk-1} = k(k-1)(2N)^{-2}$ , this expectation is clearly  $(2N)^2/k(k-1)$ . That is

$$\mathbb{P}(T_m - T_0 = t) = \frac{1}{N^2} \sum_{k=3}^{m+2} g_{k,2}(m) g_{k-1,2}(t-1) / k(k-1).$$

Recall that we have assumed  $m < 2N$ . If  $m \geq 2N$ , further terms are needed on the right.

#### 4. Approximations to other models

One natural question to ask is how the results of Sections 2 and 3 apply to other reproduction schemes, as typified by the Wright–Fisher model in (1.2). Kingman (1982a,b), as part of his analysis of a related genealogical process—the coalescent—showed that the Markov chain with transition probabilities (1.3) is well approximated by a death process in continuous time.

Temporarily denoting the discrete-time process by  $A_1^{(N)}(n)$  to exhibit the dependence on the population size  $2N$ , Kingman showed that as  $N \rightarrow \infty$

$$(4.1) \quad A_1^{(N)}(\lfloor 2Nt \rfloor) \Rightarrow A_1(t), \quad t \geq 0$$

where  $A_1(t)$  is a death process on  $\{1, 2, \dots, i\}$  with infinitesimal generator  $Q = (q_{ij})$  given by

$$(4.2) \quad q_{ll} = \frac{-l(l-1)}{2} = -q_{l,l-1}, \quad 1 \leq l \leq i.$$

Kingman (1982a), Section 4 also discusses the convergence of other discrete ancestral chains to the process determined by (4.2). Thus, viewed on their natural time-scale, many genetic models have the same ancestral process, whose structure differs only by the choice of time-scale from the Moran genealogical model (1.5). It follows immediately from (1.6) that the transition functions  $g_{il}(t) = \mathbb{P}(A_1(t) = l \mid A_1(0) = i)$  are given by

$$(4.3) \quad g_{il}(t) = \sum_{r=l}^i \frac{\rho_r(t) (-1)^{r-l} (2r-1) i_{[r]} l_{(r-1)}}{l! (r-l)! i_{(r)}}, \quad 1 \leq l \leq i,$$

where

$$\rho_r(t) = \exp \{-r(r-1)t/2\}.$$

Notice that in the case of the chain specified by (1.5), the time-scaling required to achieve (4.2) is in units of  $2N^2$  generations.

The ancestral probabilities for the whole population may be obtained by letting  $i \rightarrow \infty$  in (4.3). The process then starts from an entrance boundary at  $\infty$ ; in this case the total population size is infinite, since  $N \rightarrow \infty$ . Cf. Watterson (1982b), (3.6).

As far as the subsampling schemes studied in Section 2 are concerned, we note that in Theorem 2, the derivation depended only on the behavior of  $A_2(\cdot)$  at the times  $A_1(\cdot)$  changed state. Hence the result there applies also in the present (continuous-time) framework. In particular, (2.3) gives the conditional distribution of  $A_2(t)$  given  $A_1(t)$ . Letting  $i \rightarrow \infty$  in (2.3) we obtain

$$(4.4) \quad \begin{aligned} &\mathbb{P}(A_2(t) = l_2 \mid A_1(t) = l_1, A_1(0) = \infty, A_2(0) = j) \\ &= \binom{l_1}{l_2} \binom{j-1}{l_2-1} / \binom{l_1+j-1}{l_1-1}, \quad l_2 = 1, 2, \dots, \min(j, l_1). \end{aligned}$$

(4.4) is the conditional distribution of the number of ancestors at time  $t$  of a random sample of  $j$  individuals, given that the whole population had  $l_1$  distinct ancestors at time  $t$ . The conditional distribution (4.4) admits a simple interpretation. The denominator is the number of ways in which the  $j$  individuals can be assigned to  $l_1$  ancestors. The first term in the numerator is the number of ways of choosing the  $l_2$  ancestors of the  $j$  from  $l_1$  possible ancestors, and the second term is the number of ways the  $j$  individuals can be assigned to  $l_2$  ancestors, each ancestor being assigned at least one of the individuals. (4.4) was found by Griffiths (1980) by a different argument.

The joint distribution of  $(A_1(t), A_2(t))$  follows as in (2.5), using (4.3) in place of (1.6), and the generalisation to the multiple subsamples  $(A_1(t), \dots, A_q(t))$  follows from (4.3) and Theorem 5. Continuous-time analogues of the problems considered in Section 3 can be studied in a similar manner; here we highlight only one of them, the continuous-time counterpart of Example 4.

From Corollary 4, we have

$$\mathbb{P}(A_1(N_2(1)) = 1 \mid A_1(0) = \infty, A_2(0) = j) = \frac{j-1}{j+1},$$

and hence the probability that the common ancestor of a sample of size  $i$  taken at time 0 is also the common ancestor of the whole population at  $t$  time units before 0 is given by

$$(4.5) \quad \sum_{j=2}^i g_{ij}(t) \frac{j-1}{j+1} = e^{-t} \frac{i-1}{i+1}.$$

The case  $i = \infty$  is due to Watterson (1982b), where further details relating to

approximations to the results of Example 5 in Section 3 may be found. (The divisor  $j+2$  in Watterson (1982b), (3.9), is a misprint for  $j+1$ .)

**5. Lines of descent and the effect of mutation**

The genealogical process  $A_1(\cdot)$  used in the earlier part of this paper can be used to describe the *genetic* composition of a sample from a population in which no mutation occurs between the allelic types (cf. Felsenstein (1971), and the review article of Tavaré (1984), and the references contained therein). We now consider the case in which mutation can occur between the allelic types. We make the simplifying assumption that the mutation rate away from any allelic type is the same for all types. That is, for  $0 < m < 1$ ,

$$(5.1) \quad \mathbb{P}(\text{an allele does not mutate}) = 1 - m.$$

The line of descent from a given individual is now taken to be the descendants of that individual, but *excluding* any new mutants and their descendants (cf. Griffiths (1980)).

The mathematical tools used to analyse such a line-of-descent process are similar in spirit to those of Sections 2 and 4. However, the questions naturally asked about such a process are somewhat different, and we have therefore separated its analysis from that of the earlier models.

We shall use notation similar to that of Sections 1 and 2. In particular:

- $A_1(n)$  = number of lines of descent from the members of the sample going back  $n$  generations from the present time, 0.
- $A_2(n)$  = same quantity for the subsample of size  $A_2(0)$  chosen at random without replacement from the  $A_1(0)$ .
- $h_{il}(n) = \mathbb{P}(A_1(n) = l \mid A_1(0) = i), 0 \leq l \leq i$ .

We shall use the Moran reproduction scheme specified by Equation (1.2), and we assume that only the new-born individual can mutate. The one-step transition probabilities  $h_{il} \equiv h_{il}(1)$  are then given by

$$(5.2) \quad h_{il} = \begin{cases} 1 - \frac{i(i + \theta - 1)}{2N(2N + \theta)}, & l = i \\ \frac{i(i + \theta - 1)}{2N(2N + \theta)}, & l = i - 1 \\ 0, & \text{otherwise} \end{cases}$$

where

$$\theta = \frac{2Nm}{1 - m},$$

cf. Tavaré (1984). We remark that here state 0, rather than state 1, is the absorbing state, and  $A_1(n) = 0$  when each member of the sample is descended from a mutant occurring subsequent to time  $(-n)$ . We turn now to the computation of the joint distribution of  $(A_1(\cdot), A_2(\cdot))$ . If we define

$$P_n(i, j; l_1, l_2) = \mathbb{P}(A_1(n) = l_1, A_2(n) = l_2 \mid A_1(0) = i, A_2(0) = j)$$

then

$$(5.3) \quad P_1(i, j; l_1, l_2) = \begin{cases} 1 - \frac{i(i+\theta-1)}{2N(2N+\theta)}, & l_1 = i, l_2 = j \\ \frac{(i-j)(i+j+\theta-1)}{2N(2N+\theta)}, & l_1 = i-1, l_2 = j \\ \frac{j(j+\theta-1)}{2N(2N+\theta)}, & l_1 = i-1, l_2 = j-1 \\ 0, & \text{otherwise.} \end{cases}$$

These transition probabilities can be computed by conditioning on the ancestry of the whole sample and then taking a random sample of size  $j$  from that. The state space of the Markov chain  $(A_1(n), A_2(n))$  is  $\mathcal{S} = \{l_1 = 0, 1, \dots, i; l_2 = 0, 1, \dots, \min(j, l_1)\}$ . The probabilities in (5.3) reduce to those in (5.2) when  $i = j$ , and to those of (2.6) when  $m = 0$ .

Analogous to Theorem 2, we have the following result.

**Theorem 6.**

$$(5.4) \quad \begin{aligned} &\mathbb{P}(A_2(n) = l_2 \mid A_1(n) = l_1, A_1(0) = i, A_2(0) = j) \\ &= \frac{(i-j)! \Gamma(j+\theta) j! (i-l_1)! \Gamma(i+l_2+\theta) l_1! \Gamma(l_1+\theta)}{i! \Gamma(i+\theta) (i-l_1-j+l_2)! l_2! (l_1-l_2)! (j-l_2)! \Gamma(l_2+\theta) \Gamma(j+l_1+\theta)}. \end{aligned}$$

*Proof.* From (5.3), and ideas analogous to those in the proof of Lemma 1, we see that  $\{A_2^*(i-l_1), l_1 = 0, \dots, i\}$  is a non-homogeneous Markov chain with transition probabilities given by

$$(5.5) \quad \begin{aligned} \mathbb{P}(A_2^*(l-1) = k-1 \mid A_2^*(l) = k) &= 1 - \mathbb{P}(A_2^*(l-1) = k \mid A_2^*(l) = k) \\ &= \frac{k(k+\theta-1)}{l(l+\theta-1)}. \end{aligned}$$

As in the proof of Theorem 2, the probability on the left of (5.4) is just

$$\phi(l_1, l_2) = \mathbb{P}(A_2^*(l_1) = l_2 \mid A_2(0) = j, A_1(0) = i),$$

and the forward equations yield in place of (2.4) the recurrence

$$\begin{aligned} \phi(l_1, l_2) &= \phi(l_1 + 1, l_2) \frac{(l_1 - l_2 + 1)(l_1 + l_2 + \theta)}{(l_1 + \theta)(l_1 + 1)} \\ &\quad + \phi(l_1 + 1, l_2 + 1) \frac{(l_2 + 1)(l_2 + \theta)}{(l_1 + 1)(l_1 + \theta)} \end{aligned}$$

with boundary conditions  $\phi(i, l_2) = 1$  if  $l_2 = j; = 0$  otherwise. The solution of this is (5.4), as can be verified by substitution. The joint distribution of  $(A_1(n), A_2(n))$  follows immediately from (5.4) as

$$(5.6) \quad P_n(i, j; l_1, l_2) = h_{il_1}(n)\phi(l_1, l_2).$$

An explicit expression for the transition function  $h_{ij}(n)$  is given in the remark after (5.10).

We shall later require the distribution of  $A_1(\cdot)$  at the time when  $A_2(\cdot)$  first hits state  $k$ . This follows by the same method of proof as Lemma 3, using (5.5).

*Lemma 7.*

$$\begin{aligned} \mathbb{P}(A_1(N_2(k)) = l \mid A_1(0) = i, A_2(0) = j) \\ = \frac{(i - j)! \Gamma(j + \theta) j! (i - l - 1)! \Gamma(i + k + 1 + \theta) l! \Gamma(l + \theta)}{i! \Gamma(i + \theta) (i - l + k - j)! k! (l - k)! (j - k - 1)! \Gamma(k + \theta) \Gamma(j + l + \theta + 1)}. \end{aligned}$$

In the special case  $k = 0$ , we obtain

$$(5.7) \quad \begin{aligned} \mathbb{P}(A_1(N_2(0)) = l \mid A_1(0) = i, A_2(0) = j) \\ = \frac{(i + \theta)(i - j)! (i - l - 1)! j \Gamma(j + \theta) \Gamma(l + \theta)}{i! (i - l - j)! \Gamma(\theta) \Gamma(j + l + \theta + 1)}, \\ l = 0, 1, \dots, i - j. \end{aligned}$$

5.1. *Continuous-time results.* While it is clear from the previous paragraphs that explicit results for the Moran process determined by (5.3) are available, we shall focus attention from now on on continuous-time results which apply (as approximations) to other reproduction schemes as typified by the Wright-Fisher model (1.2). As in Section 3, such results arise by letting  $N \rightarrow \infty$ . We will assume that  $\theta = 2Nm/(1 - m)$  remains fixed as  $N \rightarrow \infty$ , or, equivalently, that

$$(5.8) \quad m \rightarrow 0, \quad 2Nm \rightarrow \theta \quad \text{as } N \rightarrow \infty.$$

Denoting the discrete-time process by  $\{(A_1^{(N)}(n), A_2^{(N)}(n)), n \geq 0\}$  to highlight the dependence on  $N$ , it follows by examination of the transition probabilities in (5.3) and an argument analogous to that of Kingman (1982a) that as  $N \rightarrow \infty$ ,

$$(A_1^{(N)}(\lfloor 2N^2 \cdot \rfloor), A_2^{(N)}(\lfloor 2N^2 \cdot \rfloor)) \Rightarrow (A_1(\cdot), A_2(\cdot))$$



where  $(A_1(t), A_2(t))$  is a Markov process on  $\mathcal{S}$ , with infinitesimal generator  $\{Q(i, j; l_1, l_2)\}$  determined by

$$(5.9) \quad Q(i, j; l_1, l_2) = \begin{cases} -i(i + \theta - 1)/2; & l_1 = i, l_2 = j \\ (i - j)(i + j + \theta - 1)/2; & l_1 = i - 1, l_2 = j \\ j(j + \theta - 1)/2; & l_1 = i - 1, l_2 = j - 1. \end{cases}$$

We remark that the continuous-time process  $\{A_1(t), t \geq 0\}$  and its generalisation specified by (5.9) arise not only as the ‘diffusion’ approximation for the model (5.3) but also for the other reproduction schemes specified in Section 1. In particular, Griffiths (1980) essentially showed that if in the Wright–Fisher line-of-descent process, we measure time in units of  $2N$  generations, and assume that  $4Nm \rightarrow \theta$  as  $N \rightarrow \infty$ , then  $A_1^{(N)}([2N \cdot]) \Rightarrow A_1(\cdot)$ . The generator of  $A_1(\cdot)$  may be found from (5.9) by setting  $j = i$ . Thus the results of this section also apply as approximations to a variety of other discrete-time models.

We record first some properties of the continuous-time death process  $A_1(t)$ . Defining

$$h_{il}(t) = \mathbb{P}(A_1(t) = l \mid A_1(0) = i), \quad 0 \leq l \leq i,$$

we have

$$(5.10) \quad h_{il}(t) = \begin{cases} \sum_{k=l}^i \frac{\rho_k(t)(-1)^{k-l}(2k + \theta - 1)(l + \theta)_{(k-1)} i_{[k]}}{l!(k-l)!(i + \theta)_{(k)}}, & 1 \leq l \leq i \\ 1 + \sum_{k=1}^i \frac{\rho_k(t)(-1)^k(2k + \theta - 1)\theta_{(k-1)} i_{[k]}}{k!(i + \theta)_{(k)}}, & l = 0 \end{cases}$$

where  $\rho_k(t) = \exp\{-k(k + \theta - 1)t/2\}$ .

*Remark.* The discrete-time transition probabilities  $h_{il}(n)$  for the Moran model specified by (5.2) are given by (5.10) with  $\rho_k(t)$  replaced by  $\rho_k^n$ , say, where

$$\rho_k = 1 - \frac{k(k + \theta - 1)}{2N(2N + \theta)}, \quad \theta = \frac{2Nm}{1 - m}.$$

If we are interested in the distribution of lines-of-descent in the *whole* population, then the process  $A_1(\cdot)$  starts from an entrance boundary at  $\infty$  (the limit  $N \rightarrow \infty$  having been taken). The transition functions

$$h_l(t) = \mathbb{P}(A_1(t) = l \mid A_1(0) = \infty)$$

may be obtained from (5.10) by letting  $i \rightarrow \infty$ . The probabilities  $h_l(t)$  were first found by Griffiths (1980).

We notice that the proof of Theorem 6 makes no assumption about the time-scale chosen, since the crucial probability (5.5) is the same in discrete or

continuous time. Thus the result of Theorem 6, and consequently of Lemma 7, apply in the present continuous-time setting. In particular, we have

$$P_i(i, j; l_1, l_2) \equiv \mathbb{P}(A_1(t) = l_1, A_2(t) = l_2 \mid A_1(0) = i, A_2(0) = j) \\ = h_{il_1}(t)\phi(l_1, l_2),$$

where  $h_{il_1}(t)$  is given explicitly by (5.10), and  $\phi(l_1, l_2)$  by (5.4). As an immediate consequence, we obtain by letting  $i \rightarrow \infty$  in (5.4)

$$(5.11) \quad \mathbb{P}(A_2(t) = l_2 \mid A_1(t) = l_1, A_1(0) = \infty, A_2(0) = j) \\ = \frac{\binom{l_1}{l_2} \binom{j + \theta - 1}{j - l_2}}{\binom{j + l_1 + \theta - 1}{j}}, \quad l_2 = 0, 1, \dots, \min(j, l_1),$$

which is the analog of (4.4). It implies in particular that the transition functions  $h_{jl_2}(t)$  and  $h_l(t)$  are related by

$$(5.12) \quad h_{jl_2}(t) = \sum_{l_1=l_2}^{\infty} h_{il_1}(t) \binom{l_1}{l_2} \binom{j + \theta - 1}{j - l_2} / \binom{j + l_1 + \theta - 1}{j}, \quad l_2 = 0, 1, \dots, j.$$

(Cf. Griffiths (1980), Equation (22).)

**5.2. Some results for the infinite-alleles model.** In this section, we shall assume that all mutations are to new types, this corresponding to the (continuous-time) infinite-alleles model. A line-of-descent is all descendants of the same type as the ancestor, and  $A_1(t) = 0$  means that all the individuals in the sample consist of types that first arose by mutation in the interval  $(-t, 0)$ .

One question asked about this process concerns the ages of the alleles in a sample or in the population. Suppose that the infinite-alleles population is stationary. It is clear that the quantities  $h_{i0}(t)$  and  $h_0(t)$  of (5.10) can be interpreted as the distribution functions of the age of the oldest allele in a sample of size  $i$  and in the whole population, respectively. Note that neither depends on the frequency of the allele in question. (Of course, analogous exact results hold for the discrete-time Moran model.)

If a sample of size  $i$  is chosen from a (stationary) infinite-alleles population, and a further subset of size  $j$  is randomly chosen from the  $i$ , what is the probability that the oldest in the sample is included in the subsample? Using (5.7), this is clearly

$$(5.13) \quad \mathbb{P}(A_1(N_2(0)) = 0 \mid A_1(0) = i, A_2(0) = j) = \frac{j(i + \theta)}{i(j + \theta)}.$$

This is an exact result for the Moran model (5.3) with the appropriate

identification of  $\theta$ . Letting  $i \rightarrow \infty$ , we see that the probability that a sample of size  $j$  includes the oldest in the population is

$$(5.14) \quad \mathbb{P}(A_1(N_2(0)) = 0 \mid A_1(0) = \infty, A_2(0) = j) = \frac{j}{j + \theta}.$$

(5.13) and (5.14) are certainly not new. (5.13) is due to Kelly (1977) in the setting of the discrete-time Moran model, and (5.14) was found by Watterson and Guess (1977). It is interesting to note, however, that these results are simple consequences of our analysis of the bivariate line-of-descent process. A further consequence of (5.13) is the distribution of the number  $F_i$ , say, of the oldest allele in a sample of size  $i$ . By considering the probability of not choosing the oldest when a random subsample of size  $j$  is taken, we see that

$$(5.15) \quad \sum_{n=1}^{i-j} \mathbb{P}(F_i = n) \binom{i-n}{j} / \binom{i}{j} = \frac{(i-j)\theta}{i(j+\theta)}, \quad j = 1, 2, \dots, i-1.$$

Hence  $\mathbb{P}(F_i = n)$  may be found:

$$(5.16) \quad \mathbb{P}(F_i = n) = \frac{\theta}{n} \binom{i-1}{n-1} / \binom{i+\theta-1}{n}, \quad n = 1, 2, \dots, i.$$

From (5.15) with  $j = 1$  and  $j = 2$  respectively,

$$\mathbb{E}F_i = \frac{i + \theta}{1 + \theta}, \quad \text{Var}(F_i) = \frac{(i + \theta)\theta(i - 1)}{(1 + \theta)^2(2 + \theta)}.$$

(5.16) is also due to Kelly ((1977), (1979), Chapter 7) for a related Moran process.

We conclude this section with the following result.

*Theorem 8.* Consider a stationary infinite-alleles model. Then the distribution of the number  $N_j$ , say, of types in the population that are older than the oldest allele in a sample of size  $j$  is geometric, with mean  $\mathbb{E}N_j = \theta/j$ , and

$$(5.17) \quad \mathbb{P}(N_j = n) = \frac{j}{j + \theta} \cdot \left(\frac{\theta}{j + \theta}\right)^n, \quad n = 0, 1, 2, \dots.$$

*Proof.* From (5.7), the distribution of  $A_1(N_2(0))$  is

$$(5.18) \quad \mathbb{P}(A_1(N_2(0)) = l \mid A_2(0) = j, A_1(0) = \infty) = \frac{j\Gamma(j + \theta)\Gamma(l + \theta)}{\Gamma(\theta)\Gamma(j + l + \theta + 1)}.$$

Ewens (1972) showed that the distribution of the number of types in a sample of size  $l$  from a stationary population has probability generating function  $f_l(s)$  given by

$$(5.19) \quad f_l(s) = (\theta s)_{(l)} / \theta_{(l)}, \quad l \geq 1.$$

The number of lines of descent surviving in the population at the time  $A_2(\cdot)$  hits 0 is given by (5.18), and the individuals at the roots of these lines form a random sample from a stationary population. It follows from (5.18) and (5.19) that

$$\begin{aligned} \mathbb{E}(s^{N_i}) &= \frac{j}{j+\theta} + \sum_{l=1}^{\infty} \frac{(\theta s)_{(l)}}{\theta_{(l)}} \cdot \frac{j\Gamma(j+\theta)\Gamma(l+\theta)}{\Gamma(j+l+\theta+1)\Gamma(\theta)} \\ &= \frac{j}{j+\theta} \sum_{l=0}^{\infty} \frac{(\theta s)_{(l)} 1_{(l)}}{(j+\theta+1)_l} \cdot \frac{1}{l!} \\ &= \frac{j}{j+\theta} F(\theta s, 1; j+\theta+1; 1), \end{aligned}$$

where  $F(\cdot)$  is the hypergeometric function (cf. Abramowitz and Stegun (1972), p. 555)

$$\begin{aligned} &= \frac{j}{j+\theta} \cdot \frac{j+\theta}{j+\theta(1-s)} \\ &= \frac{j}{j+\theta(1-s)}, \end{aligned}$$

from which (5.19) follows.

**Acknowledgment**

G. A. Watterson thanks the Center for Demographic and Population Genetics, University of Texas in Houston, for its kind hospitality during the preparation of this paper.

**References**

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

CANNINGS, C. (1974) The latest roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Adv. Appl. Prob.* **6**, 260–290.

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

FELSENSTEIN, J. (1971) The rate of loss of multiple alleles in finite haploid populations. *Theoret. Popn Biol.* **2**, 391–403.

GLADSTIEN, K. (1978) The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. Appl. Math.* **34**, 630–642.

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. (1980) *The mathematics of genetic diversity*. *CBMS-NSF Regional Conf. Ser. Appl. Math.* **34**.

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.

MORAN, P. A. P. (1958) Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–71.

TAVARÉ, S. (1984) Line-of-descent and genealogical processes and their applications in population genetics models. *Theoret. Popn Biol.*

WATTERSON, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theoret. Popn Biol.* **7**, 256–276.

WATTERSON, G. A. (1982a) Substitution times for mutant nucleotides. *J. Appl. Prob.* **19A**, 59–70.

WATTERSON, G. A. (1982b) Mutant substitutions at linked nucleotide sites. *Adv. Appl. Prob.* **14**, 206–224.

WATTERSON, G. A. AND GUESS, H. A. (1977) Is the most frequent allele the oldest? *Theoret. Popn Biol.* **11**, 141–160.