THEORETICAL POPULATION BIOLOGY 26, 119–164 (1984)

Line-of-Descent and Genealogical Processes, and Their Applications in Population Genetics Models

SIMON TAVARÉ

Department of Statistics, Colorado State University, Fort Collins, Colorado 80523

Received June 22, 1983

A variety of results for genealogical and line-of-descent processes that arise in connection with the theory of some classical selectively neutral population genetics models are reviewed. While some new results and derivations are included, the principle aim is to demonstrate the central importance and simplicity of genealogical Markov chains in this theory. Considerable attention is given to "diffusion time scale" approximations of such genealogical processes. A wide variety of results pertinent to (diffusion approximations of) the classical multiallele single-locus Wright–Fisher model and its relatives are simplified and unified by this approach. Other examples where such genealogical processes play an explicit role, such as the infinite sites and infinite alleles models, are discussed. © 1984 Academic Press, Inc.

1. INTRODUCTION

Recent years have seen the development of a wide variety of results about the *genealogical* behavior of generations of individuals whose *genetic* composition is described by the classical population genetics models.

Information obtained about genealogical processes and lines of descent is certainly interesting and important in its own right, but it also contains consequences for the underlying population genetic models.

The purpose of this paper is to review some of the pertinent mathematical details about genealogical processes, lines of descent, and their structure and demonstrate how such ideas can be applied to produce a wide variety of classical results in the mathematical theory of population genetics. I hope that this review will correlate many existing results, and further exhibit the central role played by the genealogical process in explicit form.

The most complete and elegant analyses of such genealogical processes have been obtained for single-locus multiple-alleles systems. It is on such cases that this review concentrates. Some new results and some alternative derivations of known results are included. Throughout, the emphasis will be on obtaining and studying "diffusion time scale" approximations to genealogical models, since most classical population genetic results are couched in terms of diffusion approximations.

0040-5809/84 $3.00

The organization of the paper is as follows. Section 2 contains an overview of the Moran and Wright–Fisher multiple-allele single-locus models, and an introduction to genealogical Markov chains in a discrete-time finite population size setting. In Section 3, we look at Kingman's coalescent process, which describes the family tree of a sample of individuals and their ancestors. Methods for approximating behavior of this discrete Markov chain by a more tractable continuous-time process are given, and applied to the ancestral chains of Section 2.

In Section 4, we introduce an analogous process which takes into account the effect of mutation. Instead of studying the number of distinct ancestors of a sample in preceeding generations, we now look at lines of descent. A line of descent from a given individual is taken to be the descendents of that individual, but excluding any new mutants and their descendents. The diffusion time-scale approximating process, which has the structure of a Markovian death process, and which applies to a variety of underlying discrete reproduction mechanisms, is given. In Sections 5 and 6 some explicit results are given for the transition densities and related properties of these continuous-time death processes. Section 7, which forms the major part of the article, is devoted to applications of the line-of-descent and ancestral processes. We look at rate of loss of alleles, further properties of the coalescent, bivariate genealogical processes (which arise from subsampling the original sample of individuals), and properties of the diffusion processes that occur in the study of gene frequencies.

Section 8 looks briefly at the infinite alleles model, with particular emphasis on sampling properties and ages of alleles, while in Section 9 the role of genealogy in the infinite sites models is discussed.

Section 10 describes briefly some other issues arising from genealogical models, and tries to indicate some areas for future research. In order to maintain continuity in the text, two short appendixes of a more technical nature are included.

2. Setting the Scene

Consider a populatin of fixed size $2N$ in every generation, and a single locus at which K alleles are possible. Each individual in the population is one of the K allelic types, denoted by $A_1, A_2, \ldots, A_K$. Let $X_n^{(j)}$ be the number of individuals of allelic type A_j at time n. Clearly,

$$X_n^{(j)} \geqslant 0; \qquad \sum_{j=1}^{K} X_n^{(j)} = 2N, \qquad n = 0, 1, 2, \ldots. \tag{2.1}$$

Several Markov chain models are used to describe the evolution of

$\mathbf{X}_n = (X_n^{(1)},..., X_n^{(K)})$. We will assume selective neutrality, and allow mutation between the types. To this end, define for $i \neq j$,

$$m_{ij} = \mathbb{P}\{\text{allele of type } A_i \text{ mutates to } A_j\}$$

and set

$$m_{ii} = 1 - \sum_{j \neq i} m_{ij}. \tag{2.2}$$

For notational convenience, let

$$\Delta = \left\{\mathbf{i} = (i_1,..., i_K) : i_j \geqslant 0, \sum_{j=1}^{K} i_j = 2N\right\}$$

and

$$M = (m_{ij}).$$

If the current value of $\mathbf{X}_n$ is $\mathbf{i} \in \Delta$, then the fraction of allelic type A_j in the gene pool after mutation is

$$\pi_j = (2N)^{-1}(\mathbf{i}M)_j, \qquad j = 1, 2,..., K. \tag{2.3}$$

The Wright–Fisher model (cf. Ewens (1979, Chap 3) precribes the transition probabilities $p_{\mathbf{i},\mathbf{j}} = \mathbb{P}\{\mathbf{X}_{n+1} = \mathbf{j} \mid \mathbf{X}_n = \mathbf{i}\}$, $(\mathbf{i}, \mathbf{j} \in \Delta)$ by

$$p_{\mathbf{i},\mathbf{j}} = \binom{2N}{\mathbf{j}} \pi_1^{j_1} \cdots \pi_K^{j_K}, \tag{2.4}$$

corresponding to random mating and multinomial sampling of the gene pool which is divided into fractions π_j of allelic type A_j, $j = 1,..., K$. This model has nonoverlapping generations. An analogous process due to Moran, in which generations overlap, can be described as follows (see, e.g., Karlin and McGregor, 1967; Kelly, 1976). For $\mathbf{i} \in \Delta$, define

$$T_{lk}\mathbf{i} = (i_1,..., i_l + 1,..., i_k - 1,..., i_K), \qquad l \neq k.$$

The transition probabilities are then determined by

$$\begin{aligned} p_{\mathbf{i},T_{lk}\mathbf{i}} &= \frac{i_k}{2N}\pi_l, \qquad l \neq k, \\ p_{\mathbf{i},\mathbf{i}} &= 1 - \sum_k \sum_{l \neq k} \frac{i_k}{2N}\pi_l. \end{aligned} \tag{2.5}$$

These two models will be the basis of this paper, although, as will be indicated later, the "exchangeable" models of Cannings (1974) also fit into the framework of what follows.

Suppose now that there are no mutation pressures in the process. Then $\{\mathbf{X}_n, n \geqslant 0\}$ describes the effects of "random drift" of allele frequencies, and eventually the population must comprise just one allelic type. Among many fundamental questions that are asked about this process is the following: "If a sample of i individuals (genes) is chosen without replacement from the $2N$ genes at time (or generation) n, how many distinct allelic types will be represented in the sample?" This question was addressed in a limited setting by Kempthorne (1967), and more fully by Felsenstein (1971); see also Karlin (1968a).

Felenstein's analysis reveals that if

$$P_{ij}^{(n)} = \mathbb{P}\{\text{sample of } i \text{ alleles taken at time } n \text{ contains } j \text{ distinct allelic types}\} \tag{2.6}$$

then

$$P_{ij}^{(n)} = \sum_{k=1}^{2N} g_{ik} P_{kj}^{(n-1)}, \qquad 1 \leqslant j \leqslant i \leqslant 2N, \tag{2.7}$$

where

$$g_{ik} = \mathbb{P}\{i \text{ individuals randomly selected without replacement have } k \text{ distinct parents}\}, \qquad 1 \leqslant k \leqslant i \leqslant 2N. \tag{2.8}$$

Under the assumption of no selection effects, and of constancy of the reproduction mechanism over time (certainly exhibited by the Wright–Fisher and Moran models), Eq. (2.7) may be iterated to give

$$P_{ij}^{(n)} = \sum_{k=j}^{i} (G^n)_{ik} P_{kj}^{(0)}, \tag{2.9}$$

where $G = (g_{ij})$.

The matrix G is itself the one-step transition matrix of a time-homogeneous Markov chain with state space $\{1, 2, \ldots, 2N\}$. The i,jth element $g_{ij}^{(n)}$ of G^n has the interpretation that for any $s \geqslant 0$,

$$g_{ij}^{(n)} = \mathbb{P}\{i \text{ individuals chosen at generation } n+s \text{ without replacement have exactly } j \text{ distinct ancestors in generation } s\}. \tag{2.10}$$

The chain with transition matrix G will be denoted by $\{A_n, n \geqslant 0\}$. This chain describes the genealogy of individuals in the population, and it will be referred to as the *ancestral* or *genealogical* process. Notice that the behavior of lines of descent in a sample of any size in the population can be analyzed by studying the single Markov chain $\{A_n, n \geqslant 0\}$.

The intimate connection between the genetic process X_n and the ancestral process A_n has been exploited in an elegant series of articles by Gladstien (1976, 1977, 1978), in which the effects of variable population size are also studied. The explicit form of G is known for a wide class of reproduction models whose structure can be described as follows (Cannings, 1974).

Individual i produces a random number Y_i of offspring (an "offspring" of an individual may be that individual himself if there are overlapping generations). Because the total population size is assumed fixed at $2N$ individuals, we must have $Y_1 + \cdots + Y_{2N} = 2N$. Reproductive symmetry of individuals is reflected in the fact that the random variables $(Y_1,..., Y_{2N})$ are assumed exchangeable, and we also assume that their joint distribution is invariant over time.

Under these assumptions, Gladstien (1978) has shown that

$$g_{ij} = \frac{\binom{2N}{j}}{\binom{2N}{i}} \sum_{\mathbf{s} \in A(i,j)} \mathbb{E} \prod_{r=1}^{i} \binom{Y_r}{s_r}, \qquad j = 1,..., i, \tag{2.11}$$

where $A(i,j)$ is the set of j-tuples of strictly positive integers, whose sum is i.

In the case of the Wright–Fisher model, $(Y_1,..., Y_{2N})$ have a joint multinomial distribution with

$$\mathbb{P}[Y_i = y_i,\ 1 \leqslant i \leqslant 2N] = \binom{2N}{\mathbf{y}} \left(\frac{1}{2N}\right)^{2N}.$$

Then (Watterson, 1975), we have

$$g_{ij} = (2N)^{-i}\, 2N(2N-1)\cdots(2N-j+1)\, \mathscr{S}_i^{(j)}, \qquad 1 \leqslant j \leqslant i, \tag{2.12}$$

where $\mathscr{S}_i^{(j)}$ is a Stirling number of the second kind. Specifically, $\mathscr{S}_i^{(j)}$, $j = 0, 1,..., i$, satisfy

$$x^i = \sum_{j=0}^{i} \mathscr{S}_i^{(j)} x(x-1)\cdots(x-j+1).$$

The form of (g_{ij}) for the Wright–Fisher model confirms that each individual chooses his parent independently, uniformly and at random from the individuals of the previous generation.

In the case of the Moran model (2.5), we find that

$$\begin{aligned} \mathbf{Y} &= (1, 1,..., 1) && \text{with probability } 1/2N \\ &= (0, 2, 1,..., 1), && \text{or a permutation of this, with} \\ & && \text{probability } 1 - 1/2N. \end{aligned} \tag{2.13}$$

From this and (2.11) we see that

$$g_{ii} = 1 - \frac{i(i-1)}{4N^2} = 1 - g_{ii-1}, \qquad 1 \leqslant i \leqslant 2N. \tag{2.14}$$

In anticipation of later results, we remark that (2.14), which is in a sense a "degenerate" genealogy, spcifies a death chain, in that the only possible one-step transitions are to the neighboring lower state. In particular, explicit results are available for G^n, and hence for the properties of the ancestral process. We will not pursue the details here, because our interest will focus on more appropriate "diffusion time-scale" approximations. Eventually, of course, A_n must reach state 1, corresponding to Malécot's observation that eventually everybody in the population can be traced back to a single ancestor.

3. A Closer Look at Family Trees. The Coalescent

The genealogical process $\{A_n, n \geqslant 0\}$ of Section 2 describes the behavior of the number of distinct ancestors of a group of individuals chosen from the population at a given time. Kingman (1982a, b) has studied the behavior of a richer process that describes the entire family tree structure of the population. Briefly, here is how this works:

Choose and fix a particular generation n, and consider a sample of i individuals (labeled $I_1,..., I_i$) chosen without replacement from this generation. Kingman has shown that the family tree of these individuals and their ancestors can be described by means of a discrete-time Markov chain $\{\mathscr{A}_r^{(N)}, r \geqslant 0\}$ whose state space is the set $\mathscr{E}_i$ equivalence relations on the set $\{1, 2,..., i\}$. The chain is defined by saying that $(l, m) \in \mathscr{A}_r^{(N)}$ if I_l and I_m have the same ancestor in generation $n - r$. (We are tacitly assuming here that the reproduction structure is extended indefinitely far into the past.) The distinct equivalence classes of $\mathscr{A}_r^{(N)}$ then correspond to those members of generation $n - r$ who give rise to $I_1,..., I_i$. We have $\mathscr{A}_0^{(N)} = \{(l, l): l = 1,..., i\}$, and, since eventually everyone is traced back to a single ancestor, $\mathscr{A}^{(N)}$ must reach an absorbing state at $\{(l, m): 1 \leqslant l, m \leqslant i\}$.

Now for $\alpha \in \mathscr{E}_i$, let $|\alpha|$ be the number of equivalence classes in α. Then the process

$$A_r = |\mathscr{A}_r^{(N)}|, \qquad r \geqslant 0; \quad A_0 = i, \tag{3.1}$$

is precisely the ancestral chain described in Section 2.

While it is possible to compute the transition probabilities $\{P_{\alpha\beta}, \alpha, \beta \in \mathscr{E}_i\}$ of $\mathscr{A}^{(N)}$, their form is rather complicated. However, there is an important

and useful way of approximating the process for large values of N. For $\alpha, \beta \in \mathscr{E}_i$ write $\beta < \alpha$ if β is obtained from α by joining two of its equivalence classes. Then Kingman (1982a) has shown that for the Wright-type reproduction scheme, we have

$$P_{\alpha\beta} = \delta_{\alpha\beta} + r_{\alpha\beta}(2N)^{-1} + O(N^{-2}), \qquad N \to \infty, \tag{3.2}$$

where

$$\begin{aligned} r_{\alpha\beta} &= -\tfrac{1}{2}k(k-1) && \text{if} \quad \alpha = \beta, \quad |\alpha| = k, \\ &= 1 && \text{if} \quad \beta < \alpha, \\ &= 0 && \text{otherwise,} \end{aligned} \tag{3.3}$$

and $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$; $= 0$ otherwise.

It then follows that if $P_N = \{p_{\alpha\beta}, \alpha, \beta \in \mathscr{E}_i\}$,

$$\lim_{N\to\infty} P_N^{[2Nt]} = e^{Rt}, \tag{3.4}$$

where $R = \{r_{\alpha\beta}\}$ is the infinitesimal generator of a continuous-time Markov chain $\{\mathscr{A}_t, t \geqslant 0\}$ on $\mathscr{E}_i$. Hence $\mathscr{A}_{[N\cdot]}^{(N)}$ converges weakly to $\mathscr{A}_\cdot$ as $N \to \infty$. Kingman calls the process $\{\mathscr{A}_t, t \geqslant 0; |\mathscr{A}_0| = i\}$ the *i-coalescent*, and he describes some of its properties. In particular, the form of the transition rates in (3.3) is robust under a variety of other reproduction schemes (e.g., many of the exchangeable-type models introduced by Cannings (1974)). A "degenerate," but important, case arises from the Moran model, where (3.2) is to be replaced by

$$P_{\alpha\beta} = \delta_{\alpha\beta} + r_{\alpha\beta}(2N^2)^{-1} \tag{3.5}$$

and (3.4) by

$$\lim_{N\to\infty} P_N^{[2N^2t]} = e^{Rt}. \tag{3.6}$$

The difference in time scale exhibited by (3.4) and (3.6) is attributable to allowing for conversion to generations in the Moran model (which operates in terms of birth–death events), and the extra factor of 2 follows from differences in the variance of the offspring distributions $\mathbf{Y}$ of the generations. As might be expected, these time-scale changes are ones that are used to approximate the genetic process $(2N)^{-1}\mathbf{X}_n$ by a diffusion process.

Kingman's approximation result applies immediately to the ancestral chain of Section 2 also. Using either (2.11) and (2.12) or (3.1), we can see that as $N \to \infty$

$$A_{[l_N t]} \equiv |\mathscr{A}_{[l_N t]}^{(N)}| \Rightarrow |\mathscr{A}_t| \equiv A_t, \qquad t \geqslant 0, \tag{3.7}$$

where

$$l_N = 2N \qquad \text{(Wright case)},$$
$$= 2N^2 \qquad \text{(Moran case)}.$$

Now $\{A_t, t \geqslant 0; A_0 = i\}$ is a continuous-time Markov chain with state space $\{1, 2, \ldots, i\}$ and infinitesimal generator $Q = (q_{ij})$ specified by

$$\begin{aligned} q_{jj} &= -\binom{j}{2}; \qquad j = 1, 2, \ldots, i, \\ q_{jk} &= \binom{j}{2}, \qquad k = j-1, \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned} \tag{3.8}$$

Thus A_t is a death process starting from $A_0 = i$, and ending at 1.

We will return to a variety of applications of properties of A_t in Section 7. But first we need to see what happens to lines of descent in the presence of mutation.

4. Lines of Descent with Mutation

Griffiths (1980a) uses diffusion methods to study lines of descent in neutral Wright–Fisher processes with mutation. Here a line of descent is defined as an inverted tree starting with a single individual at time 0, with branches at each generation where genes are produced from a parent in the line. It is important in what follows to suppose that new mutations are *not* included in the line of descent of their parents, but are considered to begin *new* lines of descent.

While it is possible to analyze lines of descent for the general mutation structure (2.2), it is considerably simpler to study the special case in which

$$m_{ii} = 1 - m \qquad \text{for all } i. \tag{4.1}$$

In this case, the mutation rate m away from any allelic type is the same for all types. This symmetrizing assumption means that we do not need to know the allele frequencies in the sampling generation.

We now turn to the explicit computation of the distribution of the number of lines of descent in a sample of size i taken without replacement from a particular generation of the Moran model described in (2.5).

Recall from (2.13) that the offspring vector $\mathbf{Y}$ is either $(1, 1, \ldots, 1)$ or (a permutation of) $(0, 2, 1, \ldots, 1)$. In the first case, a random sample of i individuals will have i lines of descent going back one generation if the new

individual is in the sample, and does not mutate, or if the new individual is not in the sample. Thus

$$\mathbb{P}(i \to i \mid \mathbf{Y} = (1, 1, \ldots, 1)) = \frac{i}{2N}(1-m) + 1 - \frac{i}{2N} = 1 - \frac{im}{2N}. \quad (4.2)$$

In the second case, we have i lines of descent going back one generation if we pick the new individual but not his parent, and the new individual does not mutate, or if we do not pick the new individual. Thus

$$\mathbb{P}(i \to i \mid \mathbf{Y} \equiv (0, 2, 1, \ldots, 1)) = 1 - \frac{i}{2N} + (1-m)\frac{i}{2N}\left(1 - \frac{i-1}{2N}\right). \quad (4.3)$$

If we now define

$$h_{ij} = \mathbb{P}(\text{sample of } i \text{ genes has } j \text{ lines of descent in previous generation}),$$

then averaging (4.2) and (4.3) over the distribution of $\mathbf{Y}$ given in (2.13) shows that

$$h_{ii} = 1 - h_{i,i-1} = 1 - \frac{i(i-1)}{4N^2} - \frac{mi}{2N}\left(1 - \frac{i-1}{2N}\right). \quad (4.4)$$

If $m = 0$, (4.4) reduces, as it must, to the ancestral probabilities (2.14). Now notice that by mutational symmetry and the equivalence and time-independence of reproductive behavior of individuals, we may extend the behavior of the sampling back through time to obtain the line-of-descent chain $\{B_n, n \geqslant 0\}$ which allows for mutation. The interpretation of the n-step transition probabilities $h_{ij}^{(n)}$ is

$$h_{ij}^{(n)} = \mathbb{P}(i \text{ individuals in generation } n+s \text{ have } j \text{ lines of descent in generation } s),$$

for any $s \geqslant 0$. Notice that 0 is an absorbing state, since eventually all original lines of descent must disappear due to the effects of sampling and mutation.

While parameters (4.4) admit simple derivations of $h_{ij}^{(n)}$ in terms of independent, but not identically distributed, geometric random variables, our interest is again in asymptotic results valid for large N. We make the usual order-of-magnitude assumption that

$$2Nm \to \theta \geqslant 0 \qquad \text{as} \quad N \to \infty. \quad (4.5)$$

Writing H_N for the transition matrix of B_n starting from $B_0 = i$, we see from (4.4) that

$$H_N = I + (2N^2)^{-1} Q + O(N^{-3}), \tag{4.6}$$

where $Q = (q_{ij})$ is given by

$$\begin{aligned} q_{jj} &= -j(j+\theta-1)/2, & j &= 0, 1, \ldots, i, \\ q_{jj-1} &= j(j+\theta-1)/2, & j &= 1, \ldots, i, \\ q_{jk} &= 0, & &\text{otherwise.} \end{aligned} \tag{4.7}$$

Kingman's argument (1982a, p. 31) then shows that

$$\lim_{N\to\infty} H_N^{([2N^2t])} = e^{Qt}, \tag{4.8}$$

effectively establishing that $B_{([2N^2t])}$ converges to a process B_t as $N \to \infty$, where B_t is a Markov chain on $\{0, 1, \ldots, i\}$ with infinitesimal generator (4.7).

We have described the effects of loss of original lines of descent using the Moran model mainly because the form of H_N is more readily computed in this case. Griffiths (1980a) studied a similar problem for the Wright–Fisher model by entirely different means. He essentially showed that if

$$\lim_{N\to\infty} 4Nm = \theta; \qquad l_N = 2N, \tag{4.9}$$

then $B_{([2Nt])}$ converges to the process B_t described above. So once again, a stability (or robustness) result emerges about the form of B_t for different reproduction mechanisms.

From now on, we will be interested in properties of B_t (and of A_t, obtained from B_t by setting $\theta = 0$) observed on their natural time scale.

5. Properties of $\{B_t, t \geqslant 0\}$

We will briefly look at some of the properties of the process B_t in its own right. In Sections 7 and 8, a variety of applications of these results will be given.

We will use the following notation throughout:

$$\begin{aligned} a_{(j)} &= a(a+1)\cdots(a+j-1), & j &\geqslant 1; \quad a_{(0)} = 1, \\ a_{[j]} &= a(a-1)\cdots(a-j+1), & j &\geqslant 1; \quad a_{[0]} = 1. \end{aligned} \tag{5.1}$$

5.1. *The Distribution of B_t*

The distribution $h_{ij}(t) = \mathbb{P}[B_t = j \mid B_0 = i]$ is most expeditiously computed by finding the spectral expansion of the generator Q defined in (4.7). This is straightforward (see Appendix I), and leads eventually to

$$h_{ij}(t) = \sum_{k=j}^{i} \rho_k(t) \frac{(-1)^{k-j}(2k+\theta-1)(j+\theta)_{(k-1)}\, i_{[k]}}{j!\,(k-j)!\,(i+\theta)_{(k)}}, \qquad 1 \leqslant j \leqslant i, \tag{5.2}$$

$$h_{i0}(t) = 1 + \sum_{k=1}^{i} \rho_k(t) \frac{(-1)^k(\theta)_{(k-1)}\, i_{[k]}(2k+\theta-1)}{k!\,(i+\theta)_{(k)}}, \tag{5.3}$$

where

$$\rho_k(t) = \exp\{-k(k+\theta-1)\,t/2\}. \tag{5.4}$$

As one possible interpretation of these results, consider an infinite alleles model in which new mutations are to new types, and all initial types are different. Then $h_{ij}(t)$ is the probability that a sample of i genes taken at time t contains j of the original types.

5.2. *Lines of Descent in the Whole Population*

Of great importance later on is the distribution of the number of lines of descent in the whole population. That is, we need to evaluate from (4.6)

$$h_j(t) = \lim_{N\to\infty} h_{2N,j}^{([2N^2t])},$$

which is the probability that j lines of descent survive to time t. Griffiths' (1981, p. 47) method shows that

$$\lim_{i\to\infty} \lim_{N\to\infty} h_{ij}^{([2N^2t])} = \lim_{N\to\infty} h_{2N,j}^{([2N^2t])},$$

so that $h_j(t)$ may be computed from (5.2) and (5.3) by taking $i \to \infty$. This gives

$$\begin{aligned} h_j(t) &= \sum_{k=j}^{\infty} \rho_k(t) \frac{(-1)^{k-j}(2k+\theta-1)(j+\theta)_{(k-1)}}{j!\,(k-j)!}, && j \geqslant 1, \\ &= 1 + \sum_{k=1}^{\infty} \rho_k(t) \frac{(2k+\theta-1)(-1)^k\,\theta_{(k-1)}}{k!}, && j = 0, \end{aligned} \tag{5.5}$$

as found by Griffiths (1980a) by a different approach.

The line-of-descent chain B_t for the whole population is a Markov process

that starts from an entrance boundary at infinity, and ends eventually at 0, at which time all original lines of descent have been lost due to sampling or the effects of mutation. The structure of B_t, starting from $B_0 = i$, $1 \leqslant i \leqslant \infty$, is simple to describe. Given that currently $B_t = j$, the waiting time for the loss of the next line has an exponential distribution with mean $2[j(j+\theta-1)]^{-1}$. This implies in particular that

$$\int_0^\infty h_{ij}(t)\,dt = \frac{2}{j(j+\theta-1)}, \qquad 1 \leqslant j \leqslant i \leqslant \infty, \tag{5.6}$$

and hence that the mean time to loss of the original lines, starting from $B_0 = i$, is

$$m_i = \sum_{k=1}^{i} \frac{2}{k(k+\theta-1)}, \qquad 1 \leqslant i \leqslant \infty. \tag{5.7}$$

See also Griffiths (1980a).

5.3. *Approximations and Bounds*

It is clear from (5.2) that for $1 \leqslant m \leqslant i$,

$$\mathbb{P}[B_t \geqslant m \mid B_0 = i] = O(\rho_m(t)), \qquad t \to \infty,$$

and that

$$\lim_{t\to\infty} \rho_m^{-1}(t)\,\mathbb{P}[B_t \geqslant m \mid B_0 = i] = \binom{i}{m} \frac{(m+\theta)_{(m)}}{(i+\theta)_{(m)}}.$$

However, a simple argument (provided in Appendix I) allows us to find bounds uniform in $i = B_0$. Indeed, for $1 \leqslant m \leqslant i$,

$$\begin{aligned} \rho_m(t) \leqslant \mathbb{P}[B_t \geqslant m \mid B_0 = i] &\leqslant \rho_m(t) \frac{(m+\theta)_{(m)}}{m!} \frac{i_{[m]}}{(i+\theta)_{(m)}} \\ &\leqslant \rho_m(t) \frac{(m+\theta)_{(m)}}{m!}. \end{aligned} \tag{5.8}$$

Setting $m = 1$ in (5.8) gives

$$e^{-\theta t/2} \leqslant 1 - h_{i0}(t) \leqslant \frac{i(1+\theta)}{i+\theta} e^{-\theta t/2} \leqslant (1+\theta)\,e^{-\theta t/2}, \qquad i \geqslant 1, \tag{5.9}$$

which provides rather shap bounds for the probability that any of the original lines of descent survive to time t, at least when θ is small.

Writing (5.9) in the form

$$1 - h_{i0}(t) \leqslant \min\left(1, \frac{i(1+\theta)}{i+\theta} e^{-\theta t/2}\right)$$

and integrating this from 0 to ∞ gives bounds for the mean time m_i to loss of lines in the form

$$\frac{2}{\theta} \leqslant m_i \leqslant \frac{2}{\theta}\left\{1 - \ln\left(\frac{i+\theta}{i+i\theta}\right)\right\}, \qquad i \geqslant 1. \tag{5.10}$$

so that $m_i \sim 2/\theta$ as $\theta \downarrow 0$, as found by Li and Nei (1977).

5.4. *Moments*

From the probability generating function of B_t, given $B_0 = i$ given in Appendix I, we find that

$$\mathbb{E}((B_t)_{[n]} \mid B_0 = i) = \sum_{l=n}^{i} \rho_l(t)(2l+\theta-1)\binom{l-1}{n-1}\frac{(\theta+l)_{(n-1)}\, i_{[l]}}{(i+\theta)_{(l)}}$$

and in particular when $n = 1$,

$$\mathbb{E}(B_t \mid B_0 = i) = \sum_{l=1}^{i} \rho_l(t)(2l+\theta-1)\frac{i_{[l]}}{(i+\theta)_{(l)}}. \tag{5.11}$$

Letting $i \to \infty$ gives Griffiths' (1980a) result

$$\mathbb{E}(B_t \mid B_0 = \infty) = \sum_{l=1}^{\infty} \rho_l(t)(2l+\theta-1).$$

5.5. *Approximations for Large θ*

For fixed $i < \infty$, and large θ, the infinitesimal parameters (4.7) show that B_t, starting from $B_0 = i$, runs like a pure linear death process. It is straightforward to verify that the process $D_\theta(t) = B_{2t/\theta}$ converges to a process D_t as $\theta \to \infty$, where D_t is a pure linear death process of rate 1 starting from $D_0 = i$.

6. The Genealogical Chain $\{A_t, t \geqslant 0\}$

We now turn to the study of the ancestral process $\{A_t, t \geqslant 0\}$ with infinitesimal generator defined at (3.8). Many of these results follow from those of Section 5 letting $\theta \downarrow 0$.

Thus, letting $g_{ij}(t) = \mathbb{P}[A_t = j \mid A_0 = i]$, we have

$$g_{ij}(t) = \sum_{k=j}^{i} \rho_k^0(t) \frac{(2k-1)(-1)^{k-j} j_{(k-1)} i_{[k]}}{j!\,(k-j)!\, i_{(k)}}, \qquad 2 \leqslant j \leqslant i, \tag{6.1}$$

where $\rho_k^0(t) = \exp\{-k(k-1)\,t/2\}$. For the case $j = 1$, we have

$$g_{i1}(t) = 1 - \sum_{k=2}^{i} \rho_k^0(t) \frac{(2k-1)(-1)^k\, i_{[k]}}{i_{(k)}}, \tag{6.2}$$

as in Griffiths (1979b) and Watterson (1982b). When $i = \infty$, we obtain (by the same reasoning as applied to the process B_t of Section 5) the distribution of the number of distinct ancestors at time 0 of the population at time t:

$$g_j(t) = \sum_{k=j}^{\infty} \rho_k^0(t) \frac{(2k-1)(-1)^{k-j} j_{(k-1)}}{(k-j)!\, j!}, \qquad 2 \leqslant j < \infty, \tag{6.3}$$

while the probability of the whole population at time t being descended from a single individual is

$$g_1(t) = 1 - \sum_{k=2}^{\infty} \rho_k^0(t)(2k-1)(-1)^k, \tag{6.4}$$

as found by Littler (1975), Griffiths (1980a), and Kingman (1982a). The expected length of time until a sample of i individuals has a single common ancestor is (as in (5.7))

$$m_i^0 = \sum_{k=2}^{i} \frac{2}{k(k-1)} = 2\left[1 - \frac{1}{i}\right]$$

and the variance is

$$v_i^0 = \frac{1}{i^2} + \sum_{k=1}^{i-1} \frac{2}{k^2} + 2\left(\frac{1}{i} - 1\right) - (m_i^0)^2.$$

As $i \to \infty$, we find that the expected time to trace the population back to a single ancestor is $m_\infty^0 = 2$, and the variance is $4\pi^2/3 - 12 \approx 1.16$ (Watterson, 1982b).

Bounds for the distributions of A_t can be found by the same method that leads to (5.8). Indeed, for $m \leqslant i$,

$$\rho_m^0(t) \leqslant \mathbb{P}[A_t \geqslant m \mid A_0 = i] \leqslant \rho_m^0(t) \frac{m_{(m)}\, i_{[m]}}{m!\, i_{(m)}} \leqslant \rho_m^0(t) \frac{m_{(m)}}{m!}. \tag{6.5}$$

When $m = 2$, we get

$$e^{-t} \leqslant 1 - g_{i1}(t) \leqslant 3e^{-t}, \qquad i = 2, 3, \ldots, \tag{6.6}$$

which bounds the distribution of time until the sample (or population) can be traced back to a single common ancestor. This result is due to Kingman (1982a). An analogous argument for the discrete-time Wright–Fisher model is provided in Kingman (1980, Appendix II).

The moments of A_t follow immediately from (5.11) by letting $\theta \downarrow 0$. We record in particular that

$$\mathbb{E}(A_t \mid A_0 = i) = \sum_{l=1}^{i} \rho_l^0(t) \frac{(2l-1)\, i_{[l]}}{i_{(l)}} \tag{6.7}$$

(Griffiths, 1981, p. 47).

7. Applications and Examples

Sections 5 and 6 focused on the properties of genealogical and line-of-descent processes in their own right. The underlying discrete-time models were approximated by continuous-time Markov processes whose structure is easy to describe. The passage from discrete time to continuous time corresponds to "infinite population size" models, the time scale being in units of $2N$ generations (for Wright–Fisher models) or in units of $2N^2$ birth–death events for the Moran model. These time scales are the ones usually associated with diffusion approximations of the underlying genetic models described in Section 2. In this section, we will use the results about genealogical chains to evaluate properties of these approximating diffusion processes. While many of the results described here are well known, their derivation by considering ancestry and lines of descent serves to highlight the central role these processes should and do play.

7.1. *Rate of Loss of Alleles*

Consider the case of the K allele model $\mathbf{X}_n$ defined in (2.1)–(2.5) in which there is no mutation. By suitably scaling time (as in (3.7)) the process $(2N)^{-1}\mathbf{X}_n$ is approximated by a diffusion process $\mathbf{X}(t)$, where $\mathbf{X}(t) = (X_1(t),\ldots, X_{K-1}(t))$, with

$$X_j(t) = \text{fraction of allele type } A_j \text{ at time } t \tag{7.1}$$

satisfying $X_j(t) \geqslant 0$, $\sum_{j=1}^{K} X_j(t) = 1$. The analog of Felsenstein's sampling equation (2.9) shows that the probability $P_{ij}(t)$ that i individuals sampled at time t contain j allelic types is

$$P_{ij}(t) = \sum_{k=j}^{i} g_{ik}(t)\, P_{kj}(0), \tag{7.2}$$

and, for the whole population, the probability that there are j distinct allelic types at time t will be

$$P_j(t) = \sum_{k=j}^{\infty} g_k(t)\, P_{kj}(0), \tag{7.3}$$

where $g_{ik}(t)$ and $g_k(t)$ are given by (6.1) and (6.3), respectively.

If a sample of size k is taken from the initial generation in which $X_j(0) = p_j$, $j = 1,..., K$, then the distribution of the number of alleles of type A_j in the sample is multinomial, with parameters $k, p_1,..., p_K$. The $P_{kj}(0)$ is then the probability that j types are represented in the sample at time 0. That is,

$$P_{kj}(0) = \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \sum\nolimits_s (p_{i_1} + \cdots + p_{i_s})^k, \qquad j = 1, 2,..., \min(k, K) \tag{7.4}$$

where $\sum_s$ is the sum over all choices $1 \leqslant i_1 < i_2 < \cdots < i_s \leqslant K$. Now consider the case of a two-allele model ($K = 2$). For an allele starting with initial frequency p, the probability $f(p; t)$, say, that the allele is fixed by time t, is clearly

$$f(p; t) = \sum_{k=1}^{\infty} g_k(t)\, p^k. \tag{7.5}$$

If we now substitute (7.5) and (7.4) into (7.3), we obtain

$$P_j(t) = \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \sum\nolimits_s f(p_{i_1} + \cdots + p_{i_s}; t), \tag{7.6}$$

a result derived by Littler (1975) by different means. See also Griffiths (1979).

The mean number of types represented in a sample of size k at time 0 (i.e., the mean of the distribution $P_{kj}(0), j = 1,..., k$) is

$$\sum_{j=1}^{K} \{1 - (1 - p_j)^k\}$$

and it follows that the expected number of allelic types surviving at time t is, from (7.3),

$$K - \sum_{j=1}^{K} f(1 - p_j; t). \tag{7.7}$$

If T_r represents the time at which for the first time there are exactly r allelic types in the population ($r = 1, 2,..., K - 1$), then

$$\mathbb{P}[T_r > t] = \sum_{j=r+1}^{K} P_j(t)$$
$$= \sum_{s=1}^{r} (-1)^{r-s} \binom{K-1-s}{r-s} \sum_s \sum_{n=2}^{\infty} g_n(t)\{(p_{i_1} + \cdots + p_{i_s})$$
$$- (p_{i_1} + \cdots + p_{i_s})^n\}, \tag{7.8}$$

the last following from (7.4) and (7.3). Integrating (7.5) from $t = 0$ to $t = \infty$ gives, using (5.6) with $\theta = 0$,

$$\mathbb{E}T_r = -2 \sum_{s=1}^{r} (-1)^{r-s} \binom{K-1-s}{r-s} \sum_s \{1 - p_{i_1} - \cdots - p_{i_s}\}$$
$$\times \ln(1 - p_{i_1} - \cdots - p_{i_s}).$$

When $r = 1$, this reduces to

$$\mathbb{E}T_1 = -2 \sum_{j=1}^{K} (1 - p_j) \ln(1 - p_j).$$

See Littler (1975) and Holgate (1979).

Rates of loss of alleles may be obtained from the behavior of $g_i(t)$ from (6.3); see Littler (1975) and Kimura (1955b) for example. The discrete-time version of this rate of loss of alleles is studied by Karlin (1968a), Felsenstein (1979), and Burrows and Cockerham (1974).

Finally, we mention that the time to fixation distribution (7.5) is precisely the probability generating function of A_t (starting from $A_0 = \infty$) evaluated at p. From Appendix I, (A7), we get

$$f(p; t) = p + \sum_{l=1}^{\infty} (2l+1)(-1)^l \rho_{l+1}^0(t)\, p(1-p)\, F(l+2, 1-l; 2; p)$$

as per Kimura (1955b). See also Ewens (1979, p. 141).

7.2. *The Distribution of the i-Coalescent*

We return now to the i-coalescent $\mathscr{A}_t$ described in Section 3. This process has state space $\mathscr{E}_i$, the set of equivalence relations on $\{1, 2, \ldots, i\}$. Despite its apparently complicated structure, it is possible to compute the distribution of $\mathscr{A}_t$. The process $\mathscr{A}_t$ moves through a sequence of equivalence relations

$$E_i \prec E_{i-1} \prec \cdots \prec E_1,$$

where $E_i = ((l, l): l = 1, \ldots, i)$, $E_1 = ((l, m): 1 \leqslant l, m \leqslant i)$, and $|E_l| = l$, spending an exponentially distributed amount of time (with mean $2/l(l-1)$) in E_l.

The Markov chain $\{E_r, r = i, i-1,\ldots, 1\}$, the jump chain of the process $\mathscr{A}_t$, has transition probabilities given by

$$\mathbb{P}(E_{k-1} = \beta \mid E_k = \alpha) = \frac{2}{k(k-1)}, \qquad \beta \prec \alpha, \quad |\alpha| = k,$$
$$= 0, \qquad \text{otherwise.}$$

This corresponds to choosing two of the equivalence classes of α at random, and merging them to form β. Kingman (1982b) established that E_k and A_t are independent, and hence the representation

$$\mathscr{A}_t = E_{A_t}, \qquad t \geqslant 0,$$

implies that

$$\mathbb{P}(\mathscr{A}_t = \alpha) = \mathbb{P}(A_t = k \mid A_0 = i)\, \mathbb{P}(E_k = \alpha)$$
$$= g_{ik}(t)\, \mathbb{P}(E_k = \alpha), \tag{7.9}$$

where $g_{ik}(t)$ is given explicitly by (6.1), while

$$\mathbb{P}(E_k = \alpha) = \frac{(i-k)!\, k!(k-1)!}{i!(i-1)!}\, \lambda_1! \cdots \lambda_k!, \tag{7.10}$$

where $\lambda_1,\ldots, \lambda_k$ are the sizes of the equivalences classes of α.

We now focus on another process which describes the family sizes of our sample of i individuals (cf. Kingman, 1982a, (5.2); Kendall, 1975). To describe this process let

$$\mathscr{P}_i = \left\{(m_1,\ldots, m_i): m_j \geqslant 0, \sum_{j=1}^{i} j m_j = i\right\}.$$

We define the family size process $\{\mathscr{F}_t, t \geqslant 0\}$ with state space $\mathscr{P}_i$ by collapsing the coalescent $\mathscr{A}_t$ as follows. For $\alpha \in \mathscr{E}_i$, $\mathbf{m} = (m_1,\ldots, m_i) \in \mathscr{E}_i$, let $f(\alpha) = \mathbf{m}$ if

$$\alpha \equiv 1^{m_1} 2^{m_2} \cdots i^{m_i},$$

the notation on the right indicating that α has m_j equivalence classes of size j, $j = 1,\ldots, i$. Let

$$\mathscr{F}_t = f(\mathscr{A}_t).$$

We call $\{\mathscr{F}_t, t \geqslant 0\}$ the family-size process; $\mathscr{F}_0 = (i, 0,\ldots, 0)$, and eventually

we must have $\mathscr{F} = (0, 0, \ldots, 0, 1)$. It follows from (7.9) and (7.10) that if $\sum_{j=1}^{i} m_j = k$, i.e., there are k families at time t, then

$$\mathbb{P}(\mathscr{F}_t = \mathbf{m}) = g_{ik}(t) \binom{i-1}{k-1}^{-1} \binom{k}{\mathbf{m}}. \tag{7.11}$$

Given that there are k families, the sizes of these families have distribution

$$\binom{i-1}{k-1}^{-1} \binom{k}{\mathbf{m}}, \qquad \sum m_j = k, \qquad \mathbf{m} \in \mathscr{S}_i.$$

This last follows from (7.10) by multiplying by the number of α having the given $\lambda_1, \ldots, \lambda_k$; the latter distribution is familiar in the context of occupancy problems (cf. Feller, 1968, pp. 38–40).

In principle the family size process $\mathscr{F}_t$ can be used to study the distribution of gene frequencies in a *sample* from the models discussed in Section 7.1. For the flavor of this, see Griffiths (1979b, p. 335).

7.3. *Joint Distribution of Number of Distinct Ancestors in Nested Subsamples*

Here we will consider the following subsampling scheme. Suppose that a sample of size i is taken at some time which we will label 0 for convenience. From this sample of size i we extract without replacement a further random sample of size j. We now record the number of distinct ancestors $(A_1(n), A_2(n))$ of our sample and the subsample, respectively, n generations earlier. The bivariate process $\{(A_1(n), A_2(n)),\ n \geqslant 0\}$ is a Markov chain with state space $\mathscr{S} = \{l_1 = 1, \ldots, i;\ l_2 = 1, 2, \ldots, \min(j, l_1)\}$. Explicit results are available for the transition structure of $(A_1(n), A_2(n))$ for the Wright–Fisher reproduction scheme (see Watterson, 1982b), and a comprehensive analysis of the Moran version is given in Saunders *et al.* (1984).

From the latter paper, we extract the following results. After transforming the time scale as described by (3.7), we can approximate the behavior of the discrete process by a continuous-time Markov process $\{(A_1(t), A_2(t)),\ t \geqslant 0;\ A_1(0) = i,\ A_2(0) = j\}$ whose infinitesimal generator $\{q(i, j; l_1, l_2)\}$ is determined by

$$\begin{aligned} q(i,j;l_1,l_2) &= -i(i-1)/2, && l_1 = i, \\ & && l_2 = j, \\ &= (i(i-1) - j(j-1))/2, && l_1 = i-1, \\ & && l_2 = j, \\ &= j(j-1)/2, && l_1 = i-1, \\ & && l_2 = j-1, \end{aligned}$$

other entries being zero. The joint distribution of $(A_1(t), A_2(t))$ is given by

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

where $g_{il_1}(t) = \mathbb{P}(A_1(t) = l_1 \mid A_1(0) = i)$ is given explicitly in (6.1) and (6.2), and the conditional probability $\phi(l_1, l_2) = \mathbb{P}(A_2(t) = l_2 \mid A_1(t) = l_1,\ A_1(0) = i,\ A_2(0) = j)$ is given by

$$\phi(l_1, l_2) = \frac{(i-j)!(i-l_1)!\,j!(j-1)!\,l_1!(l_1-1)!(i+l_2-1)!}{\left[\begin{array}{r}(j-l_2)!(l_1-l_2)!\,i!(i-1)! \\ l_2!(l_2-1)!(l_1+j-1)!(i+l_2-l_1-j)!\end{array}\right]}.$$

If we let $i \to \infty$ in the above expression, we obtain the conditional probability that a sample of size j has l_2 distinct ancestors time t ago, given that the whole population then has l_1 distinct ancestors. Denoting this probability by $P_{l_1}(l_2 \mid j)$, which is independent of t, we have

$$P_{l_1}(l_2 \mid j) = \binom{l_1}{l_2}\binom{j-1}{l_2-1}\Big/\binom{l_1+j-1}{l_1-1}, \qquad l_2 = 1,\ldots, \min(j, l_1). \quad (7.12)$$

This conditional distribution admits a simple explanation. The denominator is the number of ways in which j objects (individuals) can be dropped into l_1 cells (ancestors). The first term in the numerator is the number of ways the l_2 ancestors can be chosen from the l_1, and the second factor is the number of ways the j individuals can be assigned to l_2 ancestors, each ancestor being assigned at least one of the i indviduals. Equation (7.12) was also found by Griffiths (1980a) by a combinatorial argument.

Another useful consequence of (7.12) is that the transition functions $g_{ij}(t)$ and $g_n(t)$ of Section 6 are related by

$$g_{ij}(t) = \sum_{n=j}^{\infty} g_n(t)\binom{n}{j}\binom{i-1}{j-1}\Big/\binom{n+i-1}{n-1}, \qquad j = 1,\ldots, i, \quad (7.13)$$

obtained by conditioning first on the number of distinct ancestors of the whole population.

An analogous series of results also applies to subsampling in the lines-of-descent process (Saunders *et al.*, 1984). We record only those results for the approximating continuous-time process. If we take, as earlier, a sample of size i, and from the i a further subsample of size j and record the number of lines of descent $B_1(t)$ in the sample and $B_2(t)$ in the subsample time t ago, then $\{(B_1(t), B_2(t)),\ t \geqslant 0;\ B_1(0) = i,\ B_2(0) = j\}$ is a continuous-time Markov process on $\mathscr{S} = \{l_1 = 0, 1,\ldots, i;\ \ l_2 = 0, 1,\ldots, \min(l_1, j)\}$ with infinitesimal generator given by

$$
\begin{aligned}
q(i,j;l_1,l_2) &= -i(i+\theta-1)/2, && l_1 = i,\\
&&& l_2 = j,\\
&= (i-j)(i+j+\theta-1)/2, && l_1 = i-1, \qquad (7.14)\\
&&& l_2 = j,\\
&= j(j+\theta-1)/2, && l_1 = i-1,\\
&&& l_2 = j-1,
\end{aligned}
$$

other entries being zero.

The conditional distribution analogous to (7.12) giving the probability that a sample of j individuals at time t has l_2 lines of descent at time 0, given that the whole population at time t has l_1 is

$$
P_{l_1}(l_2 \mid j) = \binom{l_1}{l_2}\binom{j+\theta-1}{j-l_2}\Big/\binom{j+l_1+\theta-1}{j},
$$
$$
l_2 = 0, 1, \ldots, \min(j, l_1).
$$

Again we link $h_{ij}(t)$ and $h_n(t)$ of Section 5 by

$$
h_{ij}(t) = \sum_{n=j}^{\infty} h_n(t)\binom{n}{j}\binom{i+\theta-1}{i-j}\Big/\binom{n+i+\theta-1}{i},
$$
$$
j = 0, 1, \ldots, i. \qquad (7.15)
$$

Further applications of the subsampling results of this section will be given in respect of ages of alleles in Section 8.

7.4. *Transition Densities for Two-Allele Diffusion Model with No Mutation*

The simplest genetic process is the so-called random drift model with two possible alleles A and a, say. Let $X(t)$ be the fraction of allele A at time t. Questions about the time to loss or fixation of A are covered by the results of Section 7.1. The remaining questions revolve around the properties of the density of $X(t)$, so let

$$
f(t,p;y)\,dy = \mathbb{P}[X(t) \in dy \mid X(0) = p].
$$

The density f is well known, being derived by Kimura (1955a, b). In this section, we give the density in a form which makes the role of genealogy explicit. First, we need to relate the moments of $X(t)$ to those of the genealogical process. This is an elementary sampling problem, and we have

$$
\begin{aligned}
\mathbb{E}(X^n(t) \mid X(0) = p) &= \mathbb{P}(\text{sample of } n \text{ alleles taken at time } t\\
&\qquad \text{are all } A\text{'s})\\
&= \sum_{l=1}^{n} g_{nl}(t)\,p^l,
\end{aligned}
$$

the last being obtained by conditioning on the number of distinct ancestors the n individuals have at time 0. Thus

$$\mathbb{E}(X^n(t) \mid X(0)=p) = \mathbb{E}(p^{A_t} \mid A_0 = n). \tag{7.16}$$

The expectation on the left is given in Crow and Kimura (1970, Eq. (7.4.37)), and it agrees with the form of the probability generating function of A_t given in Appendix I, (A7), with $s=p$, $i=n$. Notice that because of (7.13), we have

$$\mathbb{E}(X^n(t) \mid X(0)=p) = \sum_{r=1}^{\infty} g_r(t) \sum_{j=1}^{\min(n,r)} \binom{r}{j}\binom{n-1}{n-j} p^j \Big/ \binom{n+r-1}{n}. \tag{7.17}$$

It now follows (see Appendix II) that the density f is given by

$$f(t,p;y) = \sum_{r=2}^{\infty} g_r(t) \sum_{j=1}^{r-1} \binom{r}{j} p^j (1-p)^{r-j} \times \frac{(r-1)!\, y^{j-1}(1-y)^{r-j-1}}{(j-1)!(r-j-1)!}. \tag{7.18}$$

This form of the density is due to Griffiths (1979c). Equation (7.18) shows how the genealogy comes into play: Given that there are r distinct ancestors at time 0 of the population at time t, the number of those that are type A is a binomial (r,p) random variable. Given further that j of the ancestors were type A, the frequency of A in the j lines of descent has a Beta distribution.

It is perhaps worth saying that although (7.18) follows from (7.17) directly, it could also be derived as a limit from the transition function of the Moran model by looking at the discrete coalescent process, and grouping the lines according to the type of ancestor.

Of course, densities of the process $X(t)$ conditional on either loss or fixation of the A allele follow immediately. For example, if T_1 denotes the time to fixation of A, then

$$\mathbb{P}[T_1 \leqslant t \mid X(0)=p,\, T_1 < \infty] = \sum_{n=1}^{\infty} g_n(t)\, p^{n-1}.$$

Using (A7) in Appendix I shows that this agrees with Kimura's (1970) expression. See also Ewens (1979, Eq. (5.30)). Finally,

$$\mathbb{E}(T_1 \mid X(0)=p;\, T_1 < \infty) = \sum_{n=2}^{\infty} \frac{2(1-p^{n-1})}{n(n-1)} = 2p^{-1}(1-p)\ln(1-p), \tag{7.19}$$

this last being attributed to Kimura and Ohta (1969).

7.5. *One-Way Mutation Models*

We now examine the two-allele model in which there is mutation from the A-allele to the a-allele, and no mutation in the reverse direction. The fraction $Y(t)$ of the a type at time t is a diffusion process on $[0, 1]$ with generator L_a given by

$$L_a = \frac{x(1-x)}{2}\frac{d^2}{dx^2} + \frac{\theta(1-x)}{2}\frac{d}{dx}. \tag{7.20}$$

This process arises as the continuous-time approximation to the corresponding discrete models specified by (2.4) or (2.5), θ being the (appropriately scaled) mutation rate; $\theta = \lim_{N\to\infty} 4Nm$ (Wright–Fisher model), $\theta = \lim_{N\to\infty} 2Nm$ (Moran model).

The lines-of-descent process B_t was derived under the assumption of equal mutation rates away from each allelic type. One might then expect that B_t will play a crucial role in the analysis of the $Y(t)$ process, since mutation occurs to the a-type. Notice that $\mathbb{E}(Y^n(t) \mid Y(0) = p)$ is the probability that a sample of size n contains only a-type genes. This can be computed by conditioning on the number of lines of descent B_t going back to time 0. All such lines must originate with an a-type ancestor (probability p^{B_t}), and any line which originates after that time does so following a mutation, i.e., with an a-type gene. Hence

$$\mathbb{E}(Y^n(t) \mid Y(0) = p) = \mathbb{E}(p^{B_t} \mid B_0 = n). \tag{7.21}$$

An expression for the left-hand side of (7.21) is given by Crow and Kimura (1970, (8.5.12)), and is in agreement with the expression on the right given in Appendix I, (A5) for the pgf of B_t.

If T_1 denotes the time taken until the a-allele fixes in the population, then (by conditioning on the number of lines of descent at time t), we have

$$\begin{aligned}\mathbb{P}(T_1 > t \mid Y(0) = p) &= \sum_{n=1}^{\infty} h_n(t)\{1 - p^n\} \\ &= \sum_{n=1}^{\infty} \rho_n(t)(-1)^{n-1}\frac{(2n+\theta-1)\,\theta_{(n-1)}}{n!} \\ &\quad \times F(-n; \theta + n - 1; \theta; p),\end{aligned} \tag{7.22}$$

this following from Appendix I, (A5), and (5.5). See also Crow and Kimura (1956, 1970, p. 394). The transition density f_a of $Y(t)$, given $Y(0) = p$, follows from (7.21) and the crucial sampling equation (7.15) as

$$\begin{aligned}f_a(t, p; y) = \sum_{r=1}^{\infty} h_r(t) \sum_{j=0}^{r-1} \binom{r}{j} p^j (1-p)^{r-j} \\ \times \frac{\Gamma(r+\theta)}{\Gamma(j+\theta)\,\Gamma(r-j)} y^{j+\theta-1}(1-y)^{r-j-1}.\end{aligned} \tag{7.23}$$

A simple derivation of this is given in Appendix II; this formulation can be found in Griffiths (1979c), who derived it by a limiting argument. See also Crow and Kimura (1956, 1970, p. 394). Equation (7.23) admits a similar interpretation to (7.18) in terms of sampling from lines of descent.

While the process $Y(t)$ can be analyzed in terms of the lines-of-descent process B_t, the fraction $X(t) = 1 - Y(t)$ of A-alleles is more intimately related to the ancestral process A_t. Here $X(t)$ is a diffusion process on $[0, 1]$, with generator L_A given by

$$L_A = \frac{x(1-x)}{2}\frac{d^2}{dx^2} - \frac{\theta}{2}x\frac{d}{dx}. \tag{7.24}$$

Choosing the functions $f_n(x) = x^n$ leads to

$$L_A f_n(x) = \frac{-n(n-1)}{2}(x^n - x^{n-1}) - n\frac{\theta}{2}x^n, \qquad n \geqslant 1,$$

showing that

$$\mathbb{E}[X^n(t) \mid X(0) = p] = \mathbb{E}\left(p^{A_t}\exp\left(-\frac{\theta}{2}\int_0^t A_u\,du\right) \;\middle|\; A_0 = n\right) \tag{7.25}$$

which shows the Poisson nature of the mutation mechanism in the *ancestral* process A_t. The probability that n alleles chosen at time t from the population are all type A is computed by conditioning on the number of distinct ancestors A_t at time 0, and multiplying by the probability that no mutations occur in the interval $(0, t)$. Given the process A_u, $0 \leqslant u \leqslant t$, this last probablity is precisely the exponential term on the right of (7.25); mutations occur in each line of ancestors independently according to a Poisson process of rate $\theta/2$. The total length of the ancestral lines going back to time 0 is $\int_0^t A_u\,du$, and so the probability of no mutations is $\exp\{-\theta/2\int_0^t A_u\,du\}$.

In particular,

$$\mathbb{E}(X(t) \mid X(0) = p) = pe^{-\theta t/2}.$$

Of course, the density of $X(t)$ follows immediately from (7.23) by transformation, and the distribution of T_0, the time to loss of allele A, from (7.22). In particular, we have

$$\begin{aligned}\mu(p) \equiv \mathbb{E}(T_0 \mid X(0) = p) &= \int_0^\infty \sum_{n=1}^\infty h_n(t)(1 - (1-p)^n)\,dt \\ &= \sum_{n=1}^\infty \frac{2(1-(1-p)^n)}{n(n+\theta-1)},\end{aligned} \tag{7.26}$$

this last following from (5.6). Equation (7.26) was given by Littler and Good (1978) as a computationally efficient way to calculate by term by term integration the mean time to loss in its usual form

$$2\int_0^p (1-\xi)^{-\theta}\int_\xi^1 (1-\eta)^{\theta-1}\,\eta^{-1}\,d\eta\,d\xi,$$

given by Ewens (1964). See also Li and Nei (1977) and Griffiths (1980a).

7.6. *Survival of Allele Types under Mutation*

Consider a K-allele model in which the mutation rate away from any allele is m. Whenever a mutation occurs, the resultant type will be denoted by A_0; eventually, all the types $A_1,\ldots,A_K$ will disappear, the population then comprising only A_0 alleles. The K-dimensional diffusion approximation of this process is $X(t)=(X_1(t),\ldots,X_K(t))$, where $X_j(t)$ is the fraction of allele type j, $j=1,\ldots,K$. The generator is

$$L=\frac{1}{2}\sum_{j=1}^{K}\sum_{i=1}^{K}(\delta_{ij}x_i-x_ix_j)\frac{\partial^2}{\partial x_i\,\partial x_j}-\frac{\theta}{2}\sum_{i=1}^{K}x_i\frac{\partial}{\partial x_i},$$

where θ is, as earlier, the (suitably scaled) mutation rate and the state space is $\{(x_1,\ldots,x_K): x_i\geqslant 0,\ x_1+\cdots+x_K\leqslant 1\}$.

Several authors have studied this process, since it can be used to describe disappearance of allelic types in the infinite alleles model, where new mutations are lumped together as allele A_0. The interest here will be in the properties of samples taken from the population at time t, say. We will assume that the initial frequencies are $X_j(0)=p_j$, where $p_j\geqslant 0$, $p_1+\cdots+p_K=1$.

The assumption of equal mutation rates away from each allelic type j, $j=1,\ldots,K$, is that used to derive the line-of-descent process $\{B_t\}$ of Section 5, and so the results there will apply to give quantities analogous to those of Section 7.1.

For example, the probability that none of types $A_1,\ldots,A_K$ survive at time t is $h_0(t)$, given by (5.5), while for $j=1,\ldots,K$ the probability $P_j(t)$ that exactly j of the K types survive is

$$P_j(t)=\sum_{l=j}^{\infty}h_l(t)\,P_{lj}(0), \tag{7.27}$$

where $P_{lj}(0)$ is given by (7.4). Analogous results hold for a sample of size i, $h_l(t)$ being replaced by $h_{il}(t)$.

If we define

$$T_k=\inf\{t:\ \text{no more than } k-1 \text{ of } A_1,\ldots,A_K \text{ remain at time } t\}$$

then

$$\mathbb{P}(T_k \leqslant t) = \sum_{j=0}^{k-1} P_j(t),$$

which reduces to Littler and Good's (1978) result:

$$\mathbb{P}(T_k \leqslant t) = \sum_{r=K-k+1}^{K} (-1)^{r-K+k-1} \binom{r-1}{K-k} \sum_r f(p_{i_1} + \cdots + p_{i_r}; t),$$

where

$$f(p; t) = \sum_{n=0}^{\infty} h_n(t)(1-p)^n \tag{7.28}$$

is the distribution function of the time to loss of an allele A described by the one–way mutation model with generator L_A defined by (7.24), starting from initial frequency p; (7.28) follows from (7.22). The mean of T_k follows as

$$\mathbb{E}T_k = \sum_{r=K-k+1}^{K} (-1)^{r-K+k-1} \binom{r-1}{K-k} \sum_r \mu(p_{i_1} + \cdots + p_{i_r}), \tag{7.29}$$

where $\mu(p)$ is defined by (7.26). Numerical values of $\mu(p)$ may be obtained from Littler and Good's Table I, multiplying each entry by 2 to allow for different scalings.

The final question of interest we study in this section concerns the frequency of alleles in the lines of descent at time t. Since allele type A_0 is described by a diffusion process $\{X_0(t),\ t \geqslant 0\colon X_0(0) = 0\}$ with generator (7.20), we have immediately from (7.23) with $p = 0$ that the density of $X_0(t)$ is

$$\sum_{r=1}^{\infty} h_r(t) \frac{\Gamma(r+\theta)}{(r-1)!\,\Gamma(\theta)} y^{\theta-1}(1-y)^{r-1}, \qquad 0 < y < 1.$$

Hence, given that r lines of descent survive to time t ($r \geqslant 1$), the conditional distribution of the total frequency $X_1(t) + \cdots + X_K(t)$ of alleles in these lines of descent is

$$\frac{\Gamma(r+\theta)}{(r-1)!\,\Gamma(\theta)} (1-y)^{\theta-1} y^{r-1}, \qquad 0 < y < 1.$$

This is due to Griffiths (1980a, (18)). He shows further that given that r lines of descent survive to time t, the joint density of $U_1, \ldots, U_r$, the gene frequencies in the lines, is given by

$$\frac{\Gamma(r+\theta)}{\Gamma(\theta)} \left(1 - \sum_1^r u_j\right)^{\theta-1}, \qquad 0 < u_j < 1, \quad 0 < \sum_1^r u_j < 1.$$

8. The Infinite Alleles Model, and Recurrent Mutation

We consider now the case in which all mutations result in novel allelic types. This corresponds to the infinite alleles model, that has been studied in great detail by many authors. Detailed accounts of the structure of the infinite alleles model may be found in the books of Ewens (1979), Kelly (1979), and Kingman (1980). We will focus on just two aspects of this extensive theory: sampling properties, and the ages of alleles. Once more, we concentrate on diffusion time-scale results.

To this end, let m be the probability that an allele mutates (to a novel type), and assume that, as in (4.5), $2Nm \to \theta$ as $N \to \infty$ in the case of the Moran reproduction scheme (or, as in (4.9), $4Nm \to \theta$ for the Wright–Fisher model) and that the process is now on its continuous time scale.

8.1. *Sampling in the Infinite Alleles Model*

Suppose that a sample of size i is taken from a stationary infinite alleles population. The sample can be represented in the following way. Let $\mathbf{m} \in \mathscr{P}_i = \{(m_1,..., m_i): m_j \geqslant 0,\ \sum_{j=1}^{i} j m_j = i\}$. We say that the sample has configuration $\mathbf{m}$ if there are m_j allele types represented j times for $j = 1,..., i$. Ewens (1972) established that the distribution of $\mathbf{m}$, $p(\mathbf{m})$, say, is given by

$$p(\mathbf{m}) = \binom{\theta + i - 1}{i}^{-1} \prod_{j=1}^{i} \left(\frac{\theta}{j}\right)^{m_j} \frac{1}{m_j!}, \qquad \mathbf{m} \in \mathscr{P}_i. \tag{8.1}$$

The Ewens sampling formula (8.1) has been drived many times. Karlin and McGregor (1972) used the ancestral probabilities g_{ij} of (2.10) for the Wright–Fisher model and then applied a limiting argument as $4Nm \to \theta$ as $N \to \infty$. Watterson (1976b) found (8.1) by using the stationary distribution of the infinite alleles diffusion model. More recently, Kingman (1982a) showed that (8.1) is a consequence of mutation in the coalescent process. Specifically, suppose that mutations occur according to a Poisson process of rate $\theta/2$, independently of $\mathscr{A}_t$ (cf. 7.25).) Now group the individuals in the sample as follows. We say that two individuals are in the same equivalence class if they are descended from a single individual, and no mutations occur in the lines down to this common ancestor. Kingman gives the distribution of the resulting equivalence relation $\Theta \in \mathscr{E}_i$ as

$$\mathbb{P}(\Theta = \alpha) = \frac{\Gamma(\theta)\,\theta^k}{\Gamma(i+\theta)} \prod_{l=1}^{k} (\mu_l - 1)!,$$

where $\mu_1,..., \mu_k$ are the sizes of the k equivalence classes in α. If we multiply this by the number of α which have the given sizes $\mu_1,..., \mu_k$, i.e., by

$$i!/\mu_1! \cdots \mu_k!\, m_1! \cdots m_i!,$$

where m_j is the number of $\mu_1,\ldots,\mu_k$ equal to $j, j = 1,\ldots, i$, we obtain (8.1). The sampling formula (8.1) also arises in other systems. See, e.g., Karlin and McGregor (1967), Kendall (1975), Kelly (1977), and Kingman (1980, 1982c). Arguments analogous to Karlin and McGregor's have recently been used by Lewis and Pollak (1982, Sect. 4) to assess the accuracy of some diffusion approximations of Nagylaki (1980) for a geographically structured population.

It follows in particular from (8.1) that the number H_i, say, of allele types observed in a sample of size i has probability generating function (pgf) given by

$$\mathbb{E}(s^{H_i}) = (\theta s)_{(i)}/\theta_{(i)}, \tag{8.2}$$

the mean number of types being

$$\mathbb{E}H_i = \theta \sum_{j=0}^{i-1} \frac{1}{\theta + j}. \tag{8.3}$$

Griffiths (1979b, Eqs. (2.2), (2.10)) has found the sampling distribution analogous to that in (8.1) at any time t, and also the mean number of allele types in that sample. Rather than recording these results, we will discuss others which give a rather detailed picture of the way in which new mutations arise and are lost in the infinite alleles model.

Consider a stationary infinite alleles population, and let $N_i(t)$ be the number of allele types in common between a sample of size i taken at time t later, and the original stationary population. The number of lines of descent B_t of the i individuals going back to time 0 has distribution $\{h_{ij}(t), j = 0, 1,\ldots, i\}$ given explicitly by (5.2) and (5.3), and the individuals at the roots of these lines form a random sample from a stationary population. Hence from (8.2) we obtain

$$\mathbb{E}(s^{N_i(t)}) = \sum_{j=0}^{i} h_{ij}(t) \frac{(\theta s)_{(j)}}{(\theta)_{(j)}}, \qquad i \geqslant 1. \tag{8.4}$$

and, from (8.3),

$$\mathbb{E}(N_i(t)) = \theta \sum_{j=1}^{i} h_{ij}(t) \left(\sum_{l=0}^{j-1} \frac{1}{\theta + l} \right). \tag{8.5}$$

Corresponding results for the number of allele types in common between two populations at time t apart may be found by letting $i \to \infty$ in (8.4) and (8.5), replacing $h_{ij}(t)$ by $h_j(t)$ in (5.5). See also Griffiths (1979b, Eq. (3.11), 1980a). Note that from (5.2),

$$\mathbb{E}(N_i(t)) = \frac{i(1+\theta)}{i+\theta} e^{-\theta t/2} + O(e^{-(1+\theta)t}), \qquad t \to \infty,$$

as found by Griffiths (1979b); a similar approximation applies to the distribution of $N_i(t)$ determined by (8.4). Further details covering the cases of (finite) samples taken from the two populations, and application to genetic distance between populations separated time t ago may be found in the latter paper of Griffiths.

8.2. *The Age of Alleles in an Infinite Alleles Model*

The previous analysis of these genetic models has been concerned with *prospective* behavior. However, to understand the nature of extant variations, we need to study the *retrospective* behavior of such processes. Problems concerning the ages of alleles seem to have been studied first by Kimura and Ohta (1973). Ewens (1979, Chaps. 3, 9) reviews some of the later approaches to this problem. The focus here will be on properties of the ages of the oldest alleles in a sample, or the population, that can be derived from the line of descent processes of Sections 5 and 7.3. We will restrict attention to "diffusion time scale" results once more, and we refer the reader to Kelly (1977, 1979) and Watterson (1976a) for some exact results for the (finite population size) Moran model.

Recall from Section 5 that a line of descent is all descendents of the same type as the ancestor, and that $B_t = 0$ means that all the individuals in the sample (or population) consist of types that first arose by mutation at most t time units ago. It follows that we can reinterpret the quantities $h_{i0}(t)$ in (5.3) and $h_0(t)$ in (5.5) as age distributions. Specifically, let X_i (respectively, X) denote the age of the oldest allele in a sample of size i (respectively, population) from a stationary infinite alleles model. Then

$$\begin{aligned} \mathbb{P}(X_i \leqslant t) &= h_{i0}(t), \qquad t \geqslant 0, \\ \mathbb{P}(X \leqslant t) &= h_0(t), \qquad t \geqslant 0. \end{aligned} \tag{8.6}$$

The bounds derived in (5.10) can also be reinterpreted in terms of ages, and it follows that

$$\frac{2}{\theta} \leqslant \mathbb{E}X_i \leqslant \frac{2}{\theta}\left(1 - \ln\left(\frac{i+\theta}{i(1+\theta)}\right)\right), \qquad i = 1, 2, \ldots.$$

The age of the oldest allele in a (stationary) infinite alleles model also satisfies, from (5.9),

$$e^{-\theta t/2} \leqslant \mathbb{P}(X > t) \leqslant (1+\theta)\, e^{-\theta t/2}.$$

It is perhaps surprising to note that the age of the oldest allele is independent of the frequency of that allele (Kelly, 1977). However, the probability that an allele *is* the oldest is not independent of its frequency, but equal to it

(Watterson and Guess, 1977) For other approaches to this problem see Sawyer (1977) and Littler and Good (1978).

Suppose now that we choose a sample of size i from the (stationary) population, and from this sample, we extract a further subsample of size j. What is the probability that the oldest allele in the sample of i is included in the subsample? This question can be answered by a more detailed analysis of the bivariate line of descent process $(B_1(t), B_2(t))$ considered in Section 7.3. The probability we require is the probability that at the time $B_2(\cdot)$ reaches 0, $B_1(\cdot)$ reaches 0 for the first time. This is found by Saunders *et al.* (1984), and we have

$$\mathbb{P}(\text{oldest in a sample of } i \text{ is included in a subsample of size } j) = j(i+\theta)/i(j+\theta). \tag{8.7}$$

In the special case $i = \infty$, we obtain

$$\mathbb{P}(\text{oldest allele in the population is included in a sample of size } j) = j/(j+\theta). \tag{8.8}$$

Equation (8.7) is due to Kelly (1977), and (8.8) to Watterson and Guess (1977), where further related results may be found. It is interesting to note, however, that these results are a simple consequence of the analysis of the bivariate line of descent process. A further consequence of (8.7) is the distribution of a number of representatives F_i of the oldest allele in a sample of size i drawn from a stationary infinite alleles model. By considering the probability of *not* choosing the oldest when a random subsample of size j is taken, we obtain

$$\mathbb{P}(F_i = n) = \frac{\theta}{n}\binom{i-1}{n-1}\Big/\binom{i+\theta-1}{n}, \qquad n = 1, 2, \ldots, i. \tag{8.9}$$

The mean and variance of F_i are

$$\mathbb{E}F_i = \frac{i+\theta}{1+\theta}, \qquad \operatorname{var}(F_i) = \frac{(i+\theta)\,\theta(i-1)}{(1+\theta)^2(2+\theta)},$$

respectively. Equation (8.9) is due to Kelly (1977).

As $i \to \infty$ in such a way that $n/i \to x$, $0 < x < 1$, we obtain the density $f(x)$ of the relative frequency of the oldest allele in a stationary population:

$$f(x) = \theta(1-x)^{\theta-1}, \qquad 0 < x < 1.$$

This is also the density of the relative frequency at equilibrium of the allelic type of randomly selected individual. Compare Sawyer (1977).

We conclude this section by finding the distribution of the number of allelic types, N_j say, in the population that are older than the oldest allelic type in a sample of size j. This was obtained by Saunders *et al.* (1984) by first finding the distribution of the number of lines of descent in the population at the time the subsample of size j first reaches 0, and then averaging over the distribution of the number of types remaining. They obtained a geometric distribution:

$$\mathbb{P}(N_j = n) = \frac{j}{j+\theta}\left(\frac{\theta}{j+\theta}\right)^n, \qquad n = 0, 1, 2, \ldots. \tag{8.10}$$

Notice that when $n = 0$, $\mathbb{P}(N_j = 0) = j/(j+\theta)$, as in (8.8).

8.3. *K-Allele Models with Recurrent Mutation*

We now concentrate on the transition densities of the diffusion approximations to the K-allele mutation models described in Section 2. The transition densities have only been found for a special class of mutation parameters, namely, those for which the mutation probabilities m_{ij} of (2.2) satisfy

$$m_{ij} = m_j; \qquad i, j = 1, 2, \ldots, K. \tag{8.11}$$

If we make the usual order of magnitude assumptions about the mutation rates m_j that

$$\begin{aligned} &4Nm_j \to \varepsilon_j \qquad \text{(Wright–Fisher model (2.4))}, \\ &2Nm_j \to \varepsilon_j \qquad \text{(Moran model (2.5))}, \end{aligned} \tag{8.12}$$

and we scale time in units of $2N$ generations for the Wright–Fisher process, and in units of $2N^2$ birth–death events for the Moran model, then the discrete process $(X_n^{(1)}, \ldots, X_n^{(K-1)})$ of (2.1) is approximated by a diffusion process $\mathbf{X}(t) = (X_1(t), \ldots, X_{K-1}(t))$, where $X_j(t)$ is the relative frequency of the jth allelic type at time t, and $X_1(t) + \cdots + X_K(t) = 1$. The $\mathbf{X}(t)$ has a density p, say, that satisfies the forward diffusion equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i=1}^{K-1}\frac{\partial}{\partial y_i}\left[\sum_{j=1}^{K-1}\frac{\partial V_{ij}\, p}{\partial y_j} - M_i\, p\right], \tag{8.13}$$

where

$$M_i = \varepsilon_i - \theta y_i, \qquad V_{ij} = y_i(\delta_{ij} - y_j) \tag{8.14}$$

and where (in this subsection only)

$$\theta = \varepsilon_1 + \cdots + \varepsilon_K. \tag{8.15}$$

A general description of the form of (8.13) may be found, e.g., in Ewens (1979, Sec. 5.10). Several authors have found eigenfunction expansions for the density p of (8.13). Kimura (1956) studied the case $K = 3$ with no mutation effects, and Littler and Fackerell (1975) extended Kimura's analysis. Results for the two-allele mutation model are well known (cf. Crow and Kimura, 1956, 1970), and these were extended by Karlin and McGregor (1964). The results of this section are due to Griffiths (1979c) who used a different approach.

Suppose, then, that $X_j(0) = x_j$, $j = 1,..., K$, and that $\varepsilon_j > 0$, $j = 1,..., K$. Then Griffiths shows that the density $p(t, \mathbf{x}; \mathbf{y})$ is given by

$$p(t, \mathbf{x}; \mathbf{y}) = \sum_{m=0}^{\infty} h_m(t) \sum\nolimits_m \binom{m}{\mathbf{l}} x_1^{l_1} \cdots x_K^{l_K} \times \frac{\Gamma(\theta + m)}{\Gamma(l_1 + \varepsilon_1) \cdots \Gamma(l_K + \varepsilon_K)} y_1^{l_1 + \varepsilon_1 - 1} \cdots y_K^{l_K + \varepsilon_K - 1}, \tag{8.16}$$

for $0 < y_i < 1$, $y_1 + \cdots + y_K = 1$, and where $h_m(t)$, the probability that there are m lines of descent surviving to time t, is given explicitly in (5.5), and $\sum_m$ is taken over all integers $l_1,..., l_K \geqslant 0$, $l_1 + \cdots + l_K = m$.

The expression in (8.16) is particularly useful for simulation purposes, as has been pointed out by Griffiths and Li (1983), since it shows that $(X_1(t),..., X_K(t))$ may be viewed probabilistically as the result of first sampling from the line-of-descent process B_t, then, given the value of B_t, m say, sampling from a multinomial distribution to find the numbers of the different alleles at the roots of the lines of descent, and finally generating an observation from a Dirichlet distribution.

Letting $t \to \infty$ in (8.16), we obtain the well-known limiting (and stationary) distribution of $\mathbf{X}(t)$:

$$p(\infty, \mathbf{x}; \mathbf{y}) = \frac{\Gamma(\theta)}{\Gamma(\varepsilon_1) \cdots \Gamma(\varepsilon_K)} y_1^{\varepsilon_1 - 1} \cdots y_K^{\varepsilon_K - 1},$$

$$0 < y_i < 1, \qquad \sum_1^K y_i = 1.$$

Of more importance perhaps is the fact that (8.16) applies in other interesting cases too. For example, the symmetric mutation model in which

$$\varepsilon_j \equiv \varepsilon = u/(K - 1), \qquad \theta = K\varepsilon \tag{8.17}$$

has been used by a variety of authors as the starting point for their analysis of the infinite alleles model, formally obtained by letting $K \to \infty$. Some of the details may be found in e.g., Griffiths (1979a–c), Kingman (1980), Stewart (1976), Griffiths and Li (1983), and Watterson (1976b). Furthermore, (8.16)

continues to hold in the limit as any $\varepsilon_j \to 0$. Thus (8.16) includes the previously mentioned results for the case of no mutation. In particular, it also applies to the model of Littler and Good (1978) which was studied in some detail in Section 7.6.

9. Infinite Sites Models

The final process we consider is an infinite sites model due to Watterson (1975). We will assume that our idealized cistrons have infinitely many sites, and that there is complete linkage between the sites. Each site is subject to mutation during meiosis, and the number of new mutant sites per cistron per generation is a Poisson random variable with mean v. We assume that the number of mutations occurring in different gametes are independent, and, because there are infinitely many sites, we assume that no two mutations ever occur at the same site. The reproduction mechanisms of the population, which is of fixed size $2N$ as before, is specified by prescribing an ancestral process $\{A_n, n \geqslant 0\}$ as described in Section 2.

Several authors have considered variants of this problem. Ewens (1974, 1979, Chap. 9) describes the model with independent sites. Of particular interest has been the distribution of the random variable $K_i(t)$ defined by

$$K_i(t) = \text{number of segregating sites in a random sample of size } i \text{ taken at time } t. \tag{9.1}$$

As in the earlier sections, we will concentrate on "diffusion time-scale" results inherited by approximating the ancestral processes $\{A_n, n \geqslant 0\}$ by their death-process analogs $\{A_t, t \geqslant 0\}$. Li (1977) finds the time-dependent behavior of $K_2(t)$ for the present model, and Golding and Strobeck (1982), Griffiths (1980a) study the distribution of $K_2(t)$ in a finite sites model. The stationary (and limiting) behavior of the infinite sites model is due to Watterson (1975), where explicit results for discrete models with Moran-type reproduction (and others) are detailed. Chakraborty (1977) studies the distribution of $K_2(t)$ for variable population size models. The distribution of $K_i(t)$ for any i and t was found by Griffiths (1981); the connection between the number of alleles and the number of segregating sites is amplified in Griffiths (1982).

9.1. *Stationary Properties*

Let K_i be the number of segregating sites in a sample of size i from a stationary population. To find the distribution of K_i, we look back at the times T_j during which the i individuals in the sample had exactly j distinct ancestors, $j = i,..., 2$, and compute the distribution of the number of segregating sites picked up and passed on to the sample during these times.

Once again we assume $\theta = \lim_{N\to\infty} 2Nv$ (for Moran-type reproduction; $\theta = \lim_{N\to\infty} 4Nv$ in Wright–Fisher case). Suppose that $T_j = t$. Then, conditional on t, the number of new segregating sites, M_j, contributed to the sample has a Poisson distribution with mean $j\theta t/2$, the sum of j independent Poisson random variables with mean $\theta t/2$ (cf. Watterson, 1975, Sec. 4). Thus

$$\mathbb{E}(s^{M_j} \mid T_j = t) = \exp\{-(\theta jt/2)(1-s)\}. \tag{9.2}$$

From (3.8), T_j has an exponential distribution with parameter $j(j-1)/2$, and hence averaging (9.2) over the distribution of T_j gives

$$\begin{aligned}\mathbb{E}(s^{M_j}) &= \int_0^\infty \exp\left\{-\frac{\theta jt}{2}(1-s)\right\} \frac{j(j-1)}{2} \exp\left\{\frac{-j(j-1)\,t}{2}\right\} dt \\ &= \frac{j-1}{j-1+\theta(1-s)}.\end{aligned} \tag{9.3}$$

Thus M_j has a geometric distribution with mean $\mathbb{E}M_j = \theta/(j-1)$. Hence (Watterson, 1975), K_i can be represented as the sum of independent geometric random variables:

$$K_i = M_i + M_{i-1} + \cdots + M_2. \tag{9.4}$$

The distribution of K_i can readily be found from (9.3) and (9.4) by a partial fraction expansion, giving

$$\mathbb{P}(K_i = n) = \frac{(i-1)}{\theta} \sum_{l=1}^{i-1} (-1)^{l-1} \binom{i-2}{l-1} \left(\frac{\theta}{l+\theta}\right)^{n+1}, \qquad n = 0, 1, \ldots. \tag{9.5}$$

The probability of no segregating sites in the sample is

$$\mathbb{P}(K_i = 0) = \frac{(i-1)!\,\Gamma(1+\theta)}{\Gamma(i+\theta)},$$

which agrees, as it must, with (8.2), the probability that the sample has only one allele type.

The mean of K_i is

$$\mathbb{E}K_i = \theta\left(1 + \frac{1}{2} + \cdots + \frac{1}{i-1}\right). \tag{9.6}$$

This result leads to an unbiased estimator $\tilde{\theta}$, say, of the parameter θ given the observed number of segregating sites:

$$\tilde{\theta} = K_i \Big/ \left(1 + \frac{1}{2} + \cdots + \frac{1}{i-1}\right).$$

Compare Ewens (1974). By writing (9.5) in the form

$$\mathbb{P}(K_i = n) = \int_0^\infty \frac{(\theta x)^n e^{-\theta x}}{n!} (i-1) e^{-x} (1 - e^{-x})^{i-2} \, dx,$$

it can be checked by a total positivity argument (cf. Karlin, 1968b) that for fixed n, $\mathbb{P}(K_i = n)$ is a unimodal function of θ. Computer computation of the (now unique) maximum likelihood estimator $\hat{\theta}$ of θ shows that $\hat{\theta} > \tilde{\theta}$. However the simple asymptotic theory for $\tilde{\theta}$ suggests that the moment estimator is easier to use in practice.

The following interesting problem that further relates the infinite alleles and infinite sites models has been studied by Griffiths. A gene in the infinite alleles model of Section 8 may be viewed as an infinite sequence of completely linked sites which reproduce together. Mutations occur at sites not previously segregating in the population, and different sequences can then be identified as different alleles. Griffiths find the joint probability generating function of the number of segregating sites and alleles in a sample; the method he uses is again based on tracing back the ancestry of the sample. One question that Griffiths' analysis poses is to assess the extra information contained in the sample with regard to estimating θ by using this joint distribution (recall that based on the number of segregating sites, we could use $\tilde{\theta}$ discussed above, whereas the moment estimator based on the number of different alleles is obtained from (8.3); see, e.g., Ewens (1979)). While further investigation seems required, Griffiths' results show that if θ is small, then the number of segregating sites is approximately the number of distinct alleles minus one, so that there seems to be little extra information in using this joint distribution.

9.2. *Loss of Segregating Sites*

We move on now to an analysis of the transient behavior of $K_i(t)$. The number of segregating sites in a sample of size i can be decomposed into two parts; the sites introduced by mutation in $(0, t)$, and the segregatig sites in common between times 0 and t. We focus first on the latter.

Let $O_i(t)$ be the number of segregating sites in a sample of size i taken at time t that are in common with the segregating sites at time 0, and let $f_j(s)$ be the pgf of the number of segregating sites in a sample of size j taken at time 0. By tracing back the ancestry of the sample to time 0, we obtain

$$\mathbb{E}(s^{O_i(t)}) = \sum_{j=1}^{i} g_{ij}(t) f_j(s), \tag{9.7}$$

where the ancestral probabilities $g_{ij}(t)$ are given in Eqs. (6.1) and (6.2), and $f_1(s) \equiv 1$ (since a sample of size one has no segregating sites). The same

conditioning argument applies to the whole population, and, denoting $\lim_{i\to\infty} O_i(t)$ by $O(t)$, we have

$$\mathbb{E}(s^{O(t)}) = \sum_{j=1}^{\infty} g_j(t) f_j(s),$$

$g_j(t)$ being defined by (6.3) and (6.4). It is clear that the original segregating sites must eventually be lost, and the rate of their loss can be computed from the explicit form of $g_j(t)$ and $g_{ij}(t)$. Compare Griffiths (1981, Eqs. (9), (14), and (20)).

9.3. *The Distribution of the Number of New Segregating Sites*

Let $N_i(t)$ be the number of segregating sites that have arisen in the interval $(0, t]$. As in the stationary case, we count the number of new segregating sites passed on via births occurring to ancestors of the sample during those generations in which the sample has exactly j distinct ancestors. This leads to

$$\mathbb{E}(s^{N_i(t)}) = \mathbb{E}\left(\exp\left\{-(\theta/2)(1-s)\int_0^{\min(t,\,T)} A_u\, du\right\} \;\middle|\; A_0 = i\right), \tag{9.8}$$

where T is the time taken to trace the sample of size i back to a single common ancestor. An expression for the pgf on the left of (9.8) can be found by finding a differential equation for $f_i(s;t) \equiv \mathbb{E}(s^{N_i(t)})$. By conditioning on the value of A_h in (9.8), rearranging, dividing by h and letting $h \downarrow 0$, we get

$$\frac{\partial f_i}{\partial t} = \frac{i(i-1)}{2}(f_{i-1} - f_i) - \frac{\theta}{2}(1-s)\,if_i \tag{9.9}$$

subject to $f_i(s;0) = 1$, $i \geqslant 1$, $f_1(s,t) = 1$, $t \geqslant 0$. Equation (9.9) is due to Griffiths (1981); (9.9) is elementary to solve, since when written in matrix form, the generator is lower triangular. Notice that by letting $t \to \infty$ in (9.8), we obtain the limiting (and stationary) pgf of the number of segregating sites given by

$$f_i(s;\infty) = \mathbb{E}\left(\exp\left\{-(\theta/2)(1-s)\int_0^{T} A_u\, du\right\} \;\middle|\; A_0 = i\right). \tag{9.10}$$

The expression on the right of (9.10) can readily be evaluated to give

$$f_i(s;\infty) = \prod_{j=2}^{i} \frac{j-1}{j-1+\theta(1-s)}, \tag{9.11}$$

in agreement with (9.3) and (9.4). The solution of (9.9) is then given by (Griffiths, 1981, (8)):

$$f_i(s;t) = f_i(s;\infty) + i! \sum_{r=2}^{i} \xi_r(t) \binom{i-1}{r-1} \frac{(2r+\alpha-1)}{r\Gamma(i+r+\alpha)} \times \sum_{j=2}^{r} (-1)^{r-j} \binom{r}{j} \frac{\Gamma(j+r+\alpha-1)}{(j-1)!} [1 - f_j(s;\infty)], \quad (9.12)$$

where $\xi_r(t) = \exp\{-\frac{1}{2} r(r+\alpha-1)\, t\}$, $\alpha = \theta(1-s)$. The pgf of $K_i(t) \equiv N_i(t) + O_i(t)$ is found by replacing the term $1 - f_j(s;\infty)$ on the right of (9.12) by $f_j(s) - f_j(s;\infty)$, where $f_j(s)$ is the pgf of $O_j(0)$ in Section 9.2.

Notice that from (9.8), we have

$$\mathbb{E}N_i(t) = (\theta/2)\, \mathbb{E}\left(\int_0^{\min(t,T)} A_u \, du \mid A_0 = i \right),$$

and, using (9.10), we find that

$$\mathbb{E}N_i(t) = \theta \sum_{j=1}^{i-1} g_{ij}(t) \sum_{l=j}^{i-1} (1/l) \quad (9.13)$$

which is equivalent to Griffiths (1981, Eq. (12)), with the role of the ancestral process $\{A_t\}$ highlighted. Letting $t \to \infty$ in (9.13) recovers the stationary mean given in (9.6).

10. Related Problems

It seems appropriate to give at this stage a brief overview of some other applications of genealogical processes that have not been covered in any detail in this article, and to suggest problems for further research.

It will be clear that the preceeding results apply to essentially haploid genealogies, in which "random mating" obtains, and the effects of selection are ignored. Karlin (1968a) and Engels (1980) explore ancestral processes with particular emphasis on nonrandom mating systems, while Karlin (1968a) and Cockerham and Weir (1977) develop analogous ideas for multiple-locus systems. Some aspects of the theory for diploid models may be found in Karlin (1968a), and the recent book by Cannings and Thompson (1981). The latter book summarizes a variety of other "genealogical" methods; e.g., coefficients of kinship and identity by descent. A detailed analysis of a truly diploid genealogy would be interesting.

The question of selection seems harder to resolve, since the inherent asymmetry of the selection process destroys the simplicity of the genealogy. One possible method of attack might be via branching process models, where the

role of genealogy is clear. Several authors have considered such questions. See in particular Bühler (1967), O'N. Waugh (1981), Joffe and O'N. Waugh (1982), Jagers (1982), which deal with some aspects of the family structure of branching processes. Some problems related to the most recent ancestor of a branching population are discussed in Fleischmann and Prehn (1974), and Fleischmann and Siegmund–Schultze (1977). See also Karlin and McGregor (1967), Kendall (1975), and Jagers and Nerman (1983).

I have restricted attention to some of the simpler population genetics models. Explicit recognition of the genealogical process has played an important role in the analysis of the Ohta–Kimura (1973) ladder model. Some of this is contained in Kesten (1980), Kingman (1980), and Notohara (1981, 1982).

The "dual" approach to diffusion equations, typified by arguments like that leading to (7.25), has been exploited to great effect by Shiga (1982a, b). Dual methods have also been used by Donnelly (1984). He obtains a variety of explicit results for the two-allele Moran model by this approach. An independent derivation of the density of the K-allele mutation models described in Section 8.3 would be of interest. In particular it would be useful to make more explicit the contributions of the relative frequencies of alleles in new lines of descent and the relative frequencies of alleles in original lines of descent to these densities.

I have left aside most mention of the statistical problems related to these genetic models. Much of this, with particular reference to estimating the mutation parameter θ in the infinite alleles model, and an assessment of the effects of selection is discussed by Ewens (1979, Chap. 9).

Griffiths (1980a) has also studied the distribution of the *total* number of lines of descent in a sample, rather than the distribution of original lines of descent. These expressions are somewhat unmanageable, and, insofar as the problem relates to identity by descent, a futher look seems promising. A closer look at mutation in the coalescent process of Section 3 seems important.

Finally, Watterson (1982a, b) has used results about genealogical structure to describe the process of fixation of determining mutations (Kelly, 1979). This has led to further developments in the study of subsampled genealogies which are described in Saunders *et al.* (1984).

APPENDIX I

Here we record some of the details that lead to explicit results for the process B_t of Section 5. For our purposes, the easiest way to obtain the transition probabilities $h_{ij}(t)$ of (5.2) and (5.3) is via the spectral expansion of the generator $Q = (q_{ij})$ defined at (4.7):

$$\begin{aligned} q_{jj} &= -j(j+\theta-1)/2, \qquad j = 0,\ldots, i, \\ q_{jj-1} &= j(j+\theta-1)/2, \qquad j = 1,\ldots, i, \\ q_{jk} &= 0, \qquad \text{otherwise.} \end{aligned} \tag{A1}$$

So we seek to identify the right eigenvectors $\mathbf{r}^{(k)}$, and left eigenvectors $\mathbf{l}^{(k)}$, corresponding to the eigenvalue $\lambda_k = q_{kk}$, $k = 0, 1,\ldots, i$. This is straightforward, giving for $k \geqslant 1$,

$$\begin{aligned} l_j^{(k)} &= 0, \qquad j > k, \\ &= \binom{k}{j}(-1)^{k-j}\frac{(j+\theta)_{(k-1)}}{(k+\theta)_{(k-1)}}, \qquad j \leqslant k, \end{aligned} \tag{A2}$$

while $l_j^{(0)} = \delta_{0j}$, $j = 0, 1,\ldots, i$. Similarly, $r_j^{(0)} = 1$, $\forall j$ and

$$\begin{aligned} r_j^{(k)} &= 0, \qquad j < k, \\ &= \binom{j}{k}\frac{(k+\theta)_{(k)}}{(j+\theta)_{(k)}}, \qquad j \geqslant k. \end{aligned} \tag{A3}$$

The transition functions are then given by

$$h_{ij}(t) = \sum_{k=j}^{i} \rho_k(t)\, r_i^{(k)} l_j^{(k)}, \qquad 0 \leqslant j \leqslant i,$$

with $\rho_k(t) = \exp\{\lambda_k t\}$, which reduces to the results of (5.2) and (5.3).

Bounds on $\mathbb{P}[B_t \geqslant m \mid B_0 = i]$

For $B_0 = i$, $t \geqslant 0$, $1 \leqslant m \leqslant i$, define the process

$$Z_m(t) = \frac{\rho_m^{-1}(t)(B_t)_{(m)}}{(B_t+\theta)_{(m)}}.$$

An elementary martingale argument based on the eigenvectors in (A3) (cf. Karlin and Taylor, 1975, Chap. 6, p. 241) shows that $\{Z_m(t),\ t \geqslant 0\}$ is a martingale relative to $\{B_t, t \geqslant 0\}$. In particular

$$\mathbb{E}Z_m(t) = Z_m(0) = \frac{i_{[m]}}{(i+\theta)_{(m)}}.$$

But for $m \leqslant l \leqslant i$,

$$\frac{m_{[m]}}{(m+\theta)_{(m)}} \leqslant \frac{l_{[m]}}{(l+\theta)_{(m)}} \leqslant \frac{i_{[m]}}{(i+\theta)_{(m)}},$$

and since

$$\frac{\rho_m(t)\, i_{[m]}}{(i+\theta)_{(m)}} = \rho_m(t)\, \mathbb{E}Z_m(t) = \sum_{l=m}^{i} \frac{l_{[m]}}{(l+\theta)_{(m)}} h_{il}(t),$$

we get

$$\frac{m_{[m]}}{(m+\theta)_{(m)}} \mathbb{P}[B_t \geqslant m \mid B_0 = i] \leqslant \frac{\rho_m(t)\, i_{[m]}}{(i+\theta)_{(m)}}$$

$$\leqslant \frac{i_{[m]}}{(i+\theta)_{(m)}} \mathbb{P}[B_t \geqslant m \mid B_0 = i]. \qquad \text{(A4)}$$

Rearranging (A4) then gives (5.8).

Moments of B_t

We define, via (5.1),

$$F(a, b; c; z) = \sum_n \frac{(a)_n (b)_{(n)}}{c_{(n)}} \frac{z^n}{n!}.$$

Using (5.2), we get

$$\sum_{j=1}^{i} h_{ij}(t)\, s^j = \sum_{l=1}^{i} \rho_l(t)(-1)^l \frac{(2l+\theta-1)\, i_{[l]}\, \theta_{(l-1)}}{(i+\theta)_{(l)}\, l!}$$

$$\times \{F(-l, \theta+l-1; \theta; s) - 1\}, \qquad |s| \leqslant 1.$$

It follows that

$$\hat{F}_i(s) = \mathbb{E}(s^{B_t} \mid B_0 = i)$$

$$= 1 + \sum_{l=1}^{i} \rho_l(t) \frac{(-1)^l (2l+\theta-1)\, i_{[l]}\, \theta_{(l-1)}}{(i+\theta)_{(l)}\, l!} F(-l, \theta+l-1; \theta; s). \qquad \text{(A5)}$$

Differentiation of (A5) n times with respect to s, and letting $s \uparrow 1$ gives, in conjunction with formulas for the hypergeometric function (Abramowitz and Stegun, 556, (15.1.20), and 255 (6.1.7)), the factorial moments of B_t given in Section 5.4.

The Ancestral Chain A_t

Most of the results of the appendix apply to the process A_t by limiting arguments, letting $\theta \downarrow 0$. The analogous eigenvectors for the Q-matrix (3.8) become

$$l_j^{(k)} = \binom{k}{j}(-1)^{k-j}\frac{j_{(k-1)}}{k_{(k-1)}}, \qquad j = 1, 2, \ldots, k,$$
$$= 0, \qquad j = k+1, \ldots, i,$$
$$r_j^{(k)}\binom{j}{k}\frac{k_{(k)}}{j_{(k)}}, \qquad j = k, k+1, \ldots, i,$$
$$= 0, \qquad j = 1, \ldots, k-1,$$

and $\lambda_k = -k(k-1)/2$ are the corresponding eigenvalues. Compare Gladstien (1978). The same type of martingale argument that leads to (A4) then gives the bounds described in (6.5).

A limiting relationship for the hypergeometric function (Abramowitz and Stegun, 1972, p. 556, (15.1.2)) applied to (A5) as $\theta \downarrow 0$ gives

$$\begin{aligned}\hat{F}_i^0(s) &= \mathbb{E}(s^{A_t} \mid A_0 = i)\\ &= s + \sum_{l=1}^{i-1}(2l+1)(-1)^l \frac{i_{[l+1]}}{i_{(l+1)}}\rho_{l+1}^0(t)\, sF(-l, 1+l; 2; s)\\ &= s + \sum_{l=1}^{i-1}(2l+1)(-1)^l \frac{i_{[l+1]}}{i_{(l+1)}}\rho_{l+1}^0(t)\, s(1-s)\\ &\quad \times F(l+2, 1-l; 2; s) \qquad \text{(A7)}\end{aligned}$$

for $i \geqslant 1$, $|s| \leqslant 1$.

APPENDIX II. Derivation of a Transition Density

The aim of this appendix is to give a simple derivation of the transition density (7.23) using the sampling relationship (7.15) and (7.21).

From (7.21), we have for $n \geqslant 0$,

$$m_n = \mathbb{E}(Y^n(t) \mid Y(0) = p) = \sum_{j=0}^{n} h_{nj}(t)\, p^j. \qquad \text{(B1)}$$

Using (7.15), and rearranging gives

$$\begin{aligned}m_n &= \sum_{r=0}^{\infty} h_r(t) \sum_{j=0}^{\min(r,n)} \frac{\binom{r}{j}\binom{n+\theta-1}{n-j}}{\binom{n+r+\theta-1}{n}} p^j \qquad \text{(B2)}\\ &= \sum_{r=0}^{\infty} h_r(t)\frac{\Gamma(r+\theta)\,\Gamma(n+\theta)}{\Gamma(\theta)\,\Gamma(r+n+\theta)} F(-r, -n; \theta; p)\end{aligned}$$

$$= \sum_{r=0}^{\infty} h_r(t) \frac{\Gamma(r+\theta)\,\Gamma(n+\theta)}{\Gamma(\theta)\,\Gamma(r+n+\theta)} (1-p)^r F(-r, \theta+n; \theta; p/(p-1))$$

$$= \sum_{r=0}^{\infty} h_r(t) \frac{\Gamma(r+\theta)\,\Gamma(n+\theta)}{\Gamma(\theta)\,\Gamma(r+n+\theta)} \sum_{j=0}^{r} \binom{r}{j} p^j (1-p)^{r-j} \frac{(\theta+n)_{(j)}}{\theta_{(j)}}$$

$$= \sum_{r=1}^{\infty} h_r(t)\, p^r + \sum_{r=1}^{\infty} h_r(t) \sum_{j=0}^{r-1} \binom{r}{j} p^j (1-p)^{r-j}$$

$$\times \left\{ \frac{\Gamma(r+\theta)\,\Gamma(n+\theta)}{\Gamma(\theta)\,\Gamma(r+n+\theta)} \frac{(\theta+n)_{(j)}}{(\theta)_{(j)}} \right\}. \tag{B3}$$

Now for $\gamma > 0$ form the Laplace transform

$$\tilde{f}(t, p; \gamma) = \sum_{n=0}^{\infty} \frac{(-\gamma)^n}{n!} m_n.$$

From (B3), this is

$$e^{-\gamma} \sum_{r=0}^{\infty} h_r(t)\, p^r + \sum_{r=1}^{\infty} h_r(t) \sum_{j=0}^{r-1} \binom{r}{j} p^j (1-p)^{r-j}$$

$$\times \left\{ \sum_{n=0}^{\infty} \frac{(-\gamma)^n}{n!} \frac{\Gamma(r+\theta)\,\Gamma(n+\theta)}{\Gamma(\theta)\,\Gamma(r+n+\theta)} \frac{(\theta+n)_{(j)}}{(\theta)_{(j)}} \right\}.$$

But it is elementary to verify that the term in $\{\ \}$ above is just

$$\int_0^1 e^{-\gamma y} \frac{\Gamma(r+\theta)}{\Gamma(j+\theta)\,\Gamma(r-j)} y^{j+\theta-1} (1-y)^{r-j-1}\, dy.$$

Hence $\mathbb{P}[Y(t) = 1] = \sum_{r=0}^{\infty} h_r(t)\, p^r$, and the density $f_a(t, p; y)$ of $Y(t)$, for $0 < y < 1$, is

$$f_a(t, p; y) = \sum_{r=1}^{\infty} h_r(t) \sum_{j=0}^{r-1} \binom{r}{j} p^j (1-p)^{r-j} \frac{\Gamma(r+\theta)\, y^{j+\theta-1} (1-y)^{r-j-1}}{\Gamma(j+\theta)\,\Gamma(r-j)}.$$

This is precisely (7.23), as was to be shown.

The density (7.18) for the pure random drift case follows immediately by letting $\theta \downarrow 0$, or by computations analogous to the above using (7.17).

ACKNOWLEDGMENTS

I would like to express my thanks to Dr. R. C. Griffiths for stimulating my interest in line of descent processes and their relatives, and for helpful correspondence. Dr. I. W. Saunders made several helpful comments on an early draft of this paper, and the penetrating comments of Dr. G. A. Watterson and Professor W. J. Ewens removed several obscurities in the text. The referees comments and the support of a Biomedical Research Support Grant to Colorado State University are acknowledged.

REFERENCES

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

BÜHLER, W. J. 1967. The distribution of generations and other aspects of the family structure of branching processes, "Proceedings of the Sixth Berkely Symp. Prob. and Math. Statist.," pp. 463–480, Univ. of California Press, Berkely.

BURROWS, P. M., AND COCKERHAM, C. C. 1974. The distribution of the time to fixation of neutral genes, *Theor. Pop. Biol.* **5**, 192–207.

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.

CANNINGS, C., AND THOMPSON, E. A. 1981. "Genealogical and Genetic Structure," Cambridge Studies in Mathematical Biology No. 3, Cambridge Univ. Press, London.

CHAKRABORTY, R. 1977. Distribution of nucleotide site differences between two randomly chosen cistrons in a population of variable size, *Theor. Pop. Biol.* **11**, 11–22.

COCKERHAM, C. C., AND WEIR, B. S. 1977. Digenic descent measures for finite populations, *Genetical Res. (Cambridge)* **30**, 121–147.

CROW, J. F., AND KIMURA, M. 1956. Some genetic problems in natural populations, "Proceedings of Third Berkeley Symp. Math. Statist. and Prob.," Vol. IV, pp. 1–22, Univ. of California Press, Berkeley.

CROW, J. F., AND KIMURA, M. 1970. "An Introduction to Population Genetics theory," Harper & Row, New York.

DONNELLY, P. 1984. The transient behavior of the Moran model in population genetics, *Proc. Cambridge Phil. Soc.*, to appear.

ENGELS, W. R. 1980. Loss of selectively neutral alleles in small populations and regular mating systems, *Theor. Pop. Biol.* **17**, 345–364.

EWENS, W. J. 1964. The pseudo-transient distribution, and its use in genetics, *J. Appl. Prob.* **1**, 141–156.

EWENS, W. J. 1972. The sampling theory of selectively neutral alleles, *Theor. Pop. Biol.* **3**, 87–112.

EWENS, W. J. 1974. A note on the sampling theory of infinite alleles and infinite sites models, *Theor. Pop. Biol.* **6**, 143–148.

EWENS, W. J. 1979. "Mathematical Population Genetics," Springer, New York.

FELLER, W. 1968. "An Introduction to Probability Theory and Its Applications," 3rd ed., Wiley, New York.

FELSENSTEIN, J. 1971. The rate of loss of multiple alleles in finite haploid populations, *Theor. Pop. Biol.* **2**, 391–403.

FLEISCHMANN, K., AND PREHN, U. 1974. Ein Grenzwertsatz für subkritische Verzweigungsprozesse mit endlich vielen Typen von Teilchen, *Math. Nachr.* **64**, 357–362.

FLEISCHMANN, K., AND SIEGMUND-SCHULTZE R. 1977. The structure of reduced critical Galton–Watson processes, *Math. Nachr.* **79**, 233–241.

GLADSTIEN, K. 1976. Loss of alleles in a haploid population with varying environment, *Theor. Pop. Biol.* **10**, 383–394.

GLADSTIEN, K. 1977. Haploid populations subject to varying environment: The characteristic values and the rate of loss of alleles, *SIAM J. Appl. Math.* **32**, 778–783.

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

GOLDING, G. B., AND STROBECK, C. 1982. The distribution of nucleotide differences between two finite sequences, *Theor. Pop. Biol.* **22**, 96–107.

GRIFFITHS, R. C. 1979a. On the distribution of allele frequencies in a diffusion model, *Theor. Pop. Biol.* **15**, 140–158.

GRIFFITHS, R. C. 1979b. Exact sampling distributions from the infinite neutral alleles model, *Adv. Appl. Prob.* **11**, 326–354.

GRIFFITHS, R. C. 1979c. A transition density expansion for a multiallele diffusion model, *Adv. Appl. Prob.* **11**, 310–325.

GRIFFITHS, R. C. 1980a. Lines of descent in the diffusion approximation of neutral Wright–Fisher models, *Theor. Pop. Biol.* **17**, 37–50.

GRIFFITHS, R. C. 1980b. Genetic identity between populations when mutation rates vary within and across loci, *J. Math. Biol.* **10** 195–204.

GRIFFITHS, R. C. 1981. Transient distribution of the number of segregating sites in a neutral infinite-sites model with no recombination, *J. Appl. Prob.* **18**, 42–51.

GRIFFITHS, R. C. 1982. The number of alleles and segregating sites in a sample from the infinite alleles model, *Adv. Appl. Prob.* **14**, 225–239.

GRIFFITHS, R. C., AND LI, W. H. 1983. Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure, *Theor. Pop. Biol.* **32**, 19–33.

HOLGATE, P. 1979. Notes on information in population genetics, *J. Appl. Prob.* **16**, 473–481.

JAGERS, P. 1982. How probable is it to be first born? and other branching process applications to kinship problems, *Math. Biosci.* **59**, 1–16.

JAGERS, P., AND NERMAN, O. 1983. "The Growth and Composition of Branching Populations," Tech. Rep., Univ. of Göteborg, Sweden.

JOFFE, A., AND WAUGH, W. A. O'N. 1982. Exact distributions of kin numbers in Galton–Watson process, *J. Appl. Prob.* **19**, 767–775.

KARLIN, S. 1968a. "Equilibrium Behavior of Population Genetic Models with Non-Random Mating," Gordon & Breach, New York.

KARLIN, S. 1968b. "Total Positivity," Vol. I, Stanford Univ. Press, Stanford, Calif.

KARLIN, S., AND MCGREGOR, J. 1964. On some stochastic models in genetics, *in* "Stochastic Models in Medicine and Biology" (J. Gurland, Ed.), Univ. of Wisconsin Press, Madison.

KARLIN, S., AND MCGREGOR, J. 1967. The number of mutant forms maintained in a population, "Proceedings of Fifth Berkely Symposium on Math. Statist. and Prob.," Vol. IV, pp. 415–438, Univ. of California Press, Berkeley.

KARLIN, S., AND MCGREGOR, J. 1972. Addendum to a paper of Ewens, *Theor. Pop. Biol.* **3**, 113–116.

KARLIN, S., AND TAYLOR, H. M. 1975. "A First Course in Stochastic Processes," Academic Press. New York.

KARLIN, S., AND TAYLOR, H. M. 1980. "A Second Course in Stochastic Processes," Academic Press, New York.

KELLY, F. P. 1976. On stochastic population models in genetics, *J. Appl. Prob.* **13**, 127–131.

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.

KEMPTHORNE, O. 1967. The concept of identity of genes by descent, *in* "Proceedings of the Fifth Berkely Symp. Math. Statist. and Prob.," Vol. IV, pp. 333–348, Univ. of California Press, Berkeley.

KENDALL, D. G. 1975. Some problems in mathematical geneaology, *in* "Perspectives in Probability and Statistics" (J. Gani, Ed.), Distributed by Academic Press, London, for the Applied Probability Trust. pp. 325–345.

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

KIMURA, M. 1955a. Solution of a process of random genetic drift with a continuous model, *Proc. Natl. Acad. Sci. USA* **41**, 144–150.

KIMURA, M. 1955b. Random genetic drift in a multi-allele locus, *Evol.* **9**, 419–435.

KIMURA, M. 1956. Random genetic drift in a tri-allelic locus; Exact solution with a continuous model, *Biometrics* **12**, 57–66.

KIMURA, M. 1970. The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population, *Genet. Res. (Cambridge)* **15**, 131–134.

KIMURA, M., AND OHTA, T. 1969. The average number of generations until fixation of a mutant gene in a finite population, *Genetics* **61**, 763–771.

KIMURA, M., AND OHTA, T. 1973. The age of a neutral mutant persisting in a finite population, *Genetics* **75**, 199–212.

KINGMAN, J. F. C. 1980. "The Mathematics of Genetic Diversity," CBMS–NSF Regional Conf. Ser. in Appl. Math., Vol. 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, *Stochastic. Process. Appl.* **13**, 235–248.

KINGMAN, J. F. C. 1982c. Exchangeability and the evolution of large populations, *in* "Exchangeability in Probability and Statistics" (G. Koch and F. Spizzichino, Eds.), pp. 97–112, North-Holland, Amsterdam.

LEWIS, J. W., AND POLLAK, E. 1982. Genetic identity in subdivided populations, 1. Two equal-sized subpopulations, *Theor. Pop. Biol.* **22**, 218–240.

LI, W.-H. 1977. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population, *Genetics* **85**, 331–337.

LI, W.-H., AND NEI, M. 1977. Persistence of common alleles in two related populations or species, *Genetics* **86**, 901–914.

LITTLER, R. A. 1975. Loss of variability at one locus in a finite population, *Math. Biosci.* **25**, 151–163.

LITTLER, R. A., AND FACKERELL, E. D. 1975. Transition densities for neutral multi-allele diffusion models, *Biometrics* **31**, 117–123.

LITTLER, R. A., AND GOOD, A. J. 1978. Ages, extinction times and first passage probabilities for a multiallele diffusion model with irreversible mutation, *Theor. Pop. Biol.* **13**, 214–225.

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

NAGYLAKI, T. 1980. The strong-migration limit in geographically structured populations, *J. Math. Biol.* **9**, 101–114.

NOTOHARA, M. 1981. Eigenanalysis for the Kolmogorov backward equation for neutral multi-allelic model, *J. Math. Biol.* **11**, 235–244.

NOTOHARA, M. 1982. The lattice models of neutral multi-alleles in population genetics theory, *J. Math. Biol.* **15**, 79–92.

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

SAUNDERS, I. W., TAVARÉ, S., AND WATTERSON, G. A. 1984. On the genealogy of nested subsamples from a haploid population, *J. Appl. Prob.* in press.

SAWYER, S. 1977. On the past history of an allele now known to have frequency p, *J. Appl. Prob.* **14**, 439–450.

SHIGA, T. 1982a. Wandering phenomena in infinite-allelic diffusion models, *Adv. Appl. Prob.* **14**, 457–483.

SHIGA, T. 1982b. Diffusion processes in papulation genetics, *J. Math. Kyoto Univ.* in press.

STEWART, F. M. 1976. Variability in the amount of heterozygosity maintained in neutral populations, *Theor. Pop. Biol.* **9**, 188–201.

WATTERSON, G. A. 1975. On the number of segregating sites in generical models without recombination, *Theor. Pop. Biol.* **7**, 256–276.

WATTERSON, G. A. 1976a. Reversibility and the age of allele. I. Moran's infinitely many neutral alleles model, *Theor. Pop. 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. 1982a. Substitution times for mutant nucleotides, *J. Appl. Prob.* **19**A, 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? *Theor. Pop. Biol.* **11**, 141–160.

WAUGH, W. A. O'N. 1981. Application of the Galton–Watson process to the kin number problem, *Adv. Appl. Prob.* **13**, 631–649.