SOME THEORETICAL ASPECTS OF DIFFUSION THEORY IN POPULATION GENETICS

By G. A. Watterson[1]

Virginia Polytechnic Institute

Summary. This paper considers the problem of approximating a discrete time, discrete states, non-Markov process by a continuous diffusion process. The problem is set in the context of population genetics but may have more general application. In the genetic situation, most authors have treated the stochastic behavior of a gene frequency subject to evolutionary factors by assuming its probability distribution may be approximated by the solution of the Fokker-Planck diffusion equation. It is here shown that under certain sufficient conditions such an approximation is valid, even for genetic models in which the gene frequency does not necessarily form a Markov process.

A summary of some old and new results concerning the asymptotic behavior of gene frequency is given, with special emphasis on the case when mutation is absent so that an absorbing state will ultimately be reached.

0. Introduction. In the theory of population genetics, much use has been made of the Fokker-Planck diffusion equation; outlines of this development, and extensive bibliographies are given in [1], [9]. The justification for such a procedure seems to have had rather less attention, however, and the assumption that the gene frequency of interest forms a Markov process has usually been made. Consider the problem of investigating the evolutionary behavior of a dioecious diploid population with the population size, N, and indeed the sex numbers N_1, N_2 for males and females, being kept constant over time. Then, if only two alleles a and A are of importance at a single chromosome locus, there are six possible types of individuals, and their numbers may be represented, at time t, by six random variables:

	Genotype aa	 Aa	 AA	Totals
Males	k_t	$N_1 - k_t - l_t$	l_t	N_1
Females	r_t	$N_2 - r_t - s_t$	s_t	N_2
				N

of which only four, (k_t, l_t, r_t, s_t) are needed to specify the state of the popu-

---

Received December 1, 1959; revised February 28, 1962.

[1] The main part of this work was completed while the author was a research scholar at the Australian National University. Present address: Monash University, Victoria, Australia.

lation. By evolutionary behavior we mean the manner in which the joint distribution of these variables changes with increasing t. Births and deaths occur at discrete times, and our time scale is chosen so that these times are $t = 1, 2, 3, \cdots$; however, the process could be embedded in a continuous-time process with state changes occurring at random times. One might postulate a population model by defining transition probabilities to the states such that (k_t, l_t, r_t, s_t) forms a Markov chain with a finite number of states.

But the matrix algebra for handling such models is very difficult, and approximate procedures are required. Most authors therefore consider the a-gene relative frequency alone, which in our case is

$$\tfrac{1}{2} + \tfrac{1}{2}N^{-1}(k_t - l_t + r_t - s_t),$$

and assume that this variable has asymptotically the behavior of a diffusion process as $N \to \infty$. Actually, it is more convenient to consider the closely related function.

$$x_t = \tfrac{1}{2} + \tfrac{1}{4}N_1^{-1}(k_t - l_t) + \tfrac{1}{4}N_2^{-1}(r_t - s_t), \tag{0.1}$$

the arithmetic average of the relative frequency of the a-gene in the male and female sub-populations. But in any case, although $\{x_t\}$ has a finite number of states with transitions at discrete times, it is not a Markov process in general, for the transition probability from state (k_t, l_t, r_t, s_t) to one having a particular value for x_{t+1} depends on the k_t, l_t, r_t, s_t explicitly, and not just through the function x_t.

In this paper, we give sufficient conditions under which the diffusion approximation holds for $\{x_t\}$, applicable for the dioecious diploid case described above, but relevant also for other cases such as monoecious diploid or haploid populations. We denote by $F_N(x, t)$ the distribution function of x_t ; that is

$$F_N(x, t) = \Pr\{x_t \leqq x\}, \tag{0.2}$$

and shall show that by a suitable time-scale transformation $t = N^m u$, and under certain sufficient conditions,

$$F_N(x, N^m u) \to F(x, u) \quad \text{say,} \quad \text{as } N \to \infty, \tag{0.3}$$

where $F(x, u)$ is a distribution function uniquely determined by a diffusion equation subject to boundary conditions.

To the author's knowledge, the only rigorous derivation for such a limit is given in [8] for a particular population model, which was, in fact, Markovian with respect to $\{x_t\}$. Moran [15], [16], [17], was perhaps the first author to recognize that the standard Markovian assumptions are not generally applicable in genetics, and for three particular population models he investigated the limit

$$F(x, \infty) = \lim_{\substack{N\to\infty \\ u\to\infty}} F_N(x, N^m u), \tag{0.4}$$

where $F(x, \infty)$ is the asymptotic, stationary distribution function, and found that $F(x, \infty)$ had the same form as would be predicted by the stationary solution of a diffusion equation.

To prove (0.3), two approaches could be made. In one, certain sufficient conditions could be assumed on the 4-variate Markov process so that a 4-variate diffusion approximation would apply. This would extend the univariate argument of Khintchine [11]. Then an additional assumption could be used to collapse the 4-variate distribution so obtained into a degenerate, univariate form. However, we prefer a different approach because we wish our result to apply to essentially bivariate and univariate models also, and because the result can be derived using an extension of Moran's methods [16], [17], where the sufficient conditions can be imposed on the $\{x_t\}$ process directly.

In Section 1, an outline of the proof is given because the complete proof, contained in Sections 2–5, is rather lengthy. In Section 6, some comments are made about applications of the result, and a summary of some old and some new consequences is given, including the stationary distribution, the probability of gene fixation and the time for this to occur, and a concept of entropy.

A second paper is planned in which specific population models of finite size are shown to satisfy the assumptions of this paper.

1. Outline of proof. The assumptions for the proof of (0.3) are contained in Section 2; roughly, they amount to the conditions that the change $x_{t+1} - x_t$ has first and second order moments of the same order of magnitude, say $O(N^{-m})$, in N, and that for sufficiently large t, higher order moments are negligible. The initial value, x_0, is assumed constant at p, independent of N.

In Section 3, the discreteness in $F_N(x, t)$ due to that in x_t and in t is removed by a double transformation using two continuously variable parameters θ and α:

$$(1.1) \qquad \Phi_N(\theta, \alpha) = \int_{0-}^{1} e^{\theta x} \, d_x G_N(x, \alpha), \qquad -B \leqq \theta \leqq B,$$

where

$$(1.2) \qquad G_N(x, \alpha) = (1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} F_N(x, t), \qquad \alpha > 0.$$

Thus $\Phi_N(\theta, \alpha)$ is the moment generating function of $G_N(x, \alpha)$, itself being a mixture of the distributions $F_N(x, t)$. As a consequence of the assumptions, a third-order differential equation is found for $\Phi_N(\theta, \alpha)$, (3.5).

In Section 4, it is shown that $\Phi_N(\theta, \alpha)$ and $G_N(x, \alpha)$ converge to unique limits $\Phi(\theta, \alpha)$, $G(x, \alpha)$ as $N \rightarrow \infty$, satisfying differential equations (4.1), (4.9) respectively. Section 5 shows that the convergence $G_N(x, \alpha) \rightarrow G(x, \alpha)$ implies $F_N(x, N^m u) \rightarrow F(x, u)$ through the inversion of (1.2), where $F(x, u)$ is a distribution function whose derivative satisfies the diffusion equation (5.1).

2. Assumptions. In genetics, perhaps the most frequently considered population models have the properties that the expected drifts in gene frequency are due to migration, mutation, and selection, while second moments are influenced by sampling variances of binomial type. Consider, then, a given state (k_t, l_t, r_t, s_t) of the population at time t. Using the symbol E_t to denote an expectation conditional on this state, we assume that

$$\begin{aligned} E_t[e^{\theta(x_{t+1}-x_t)} - 1] = N^{-m}\theta\{b - (b + c)x_t + sx_t(1 - x_t) \\ \cdot[\tfrac{1}{2}(d + 1) - dx_t] + \tfrac{1}{4}\theta a x_t(1 - x_t) + W_N(\theta, t)\} \end{aligned} \tag{2.1}$$

where

(i) $m = 2$ or 1, depending on whether births (and deaths) occur one at a time at $t = 1, 2, 3, \cdots$, or the whole generation dies at these times.

(ii) a, b, c, s, d are constants independent of N, with the interpretations

a $(a > 0)$ is the ratio of effective to actual population size,

b, c $(\geqq 0)$ are mutation and/or migration rates,

s, d are selection and dominance coefficients.

(iii) $W_N(\theta, t)$ is a function of $N, \theta, t, k_t, l_t, r_t, s_t$, uniformly bounded with respect to t and N, such that for θ in any bounded interval $[-B, B]$, and for some n, $0 \leqq n < m$, its unconditional absolute moment satisfies

$$E|W_N(\theta, N^n u)| \to 0 \tag{2.2}$$

uniformly with respect to u in any interval (ϵ, ∞) with $\epsilon > 0$, as $N \to \infty$.

We further assume

$$x_0 = p, \tag{2.3}$$

a constant independent of N.

From (2.1), (2.2), we see these assumptions imply that at time $t = N^n u$,

$$E_t(x_{t+1} - x_t) \doteqdot N^{-m}\{b - (b + c)x_t + sx_t(1 - x_t)[\tfrac{1}{2}(d + 1) - dx_t]\}, \tag{2.4}$$

$$E_t(x_{t+1} - x_t)^2 \doteqdot N^{-m}\cdot\tfrac{1}{2}ax_t(1 - x_t), \tag{2.5}$$

and that higher moments are $o(N^{-m})$. Processes of this type have been considered, for example, by Wright [22], Feller [3], Kimura [13], and Crow and Kimura [2] from the diffusion point of view without strictly proving the validity of the approximation. For definiteness we use the specific moments above, although other examples are known (c.f. [12]) in which (2.5) is replaced by

$$E_t(x_{t+1} - x_t)^2 \doteqdot N^{-m}\tfrac{1}{2}ax_t^2(1 - x_t)^2,$$

or by

$$E_t(x_{t+1} - x_t)^2 \doteqdot N^{-m}\tfrac{1}{2}a(x_t - \xi)^2.$$

In (2.4), (2.5), we see that a time-scale transformation $t = N^m u$ is necessary to accumulate a change in x, from the initial value p, which is non-degenerate as $N \to \infty$. However, (2.2) implies that the approximations (2.4), (2.5) are valid for $t = N^n u$, of smaller order of magnitude than $N^m u$.

The formulation of the problem in terms of moment generating functions, implicit in (2.1), is convenient from the point of view of considering specific examples of population models, and also because x_t lies in the closed interval [0, 1] and its moments are similarly bounded, so that recourse to characteristic functions is unnecessary.

3. Transformation of the discrete distributions. The function $G_N(x, \alpha)$ introduced in (1.2) is a distribution on [0, 1] because each function $F_N(x, t)$ is, and because the mixture weights $(1 - e^{-\alpha N^{-m}})e^{-\alpha N^{-m} t}$ are positive and sum to unity with respect to t. Therefore, the moment generating function (1.1) exists, and we have

$$\begin{aligned}
\Phi_N(\theta, \alpha) &= \int_{0-}^{1} e^{\theta x}\, d_x G_N(x, \alpha) \\
&= \int_{0-}^{1} e^{\theta x}\, d_x\Big[(1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} F_N(x, t)\Big] \\
&= (1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} \int_{0-}^{1} e^{\theta x}\, d_x F_N(x, t),
\end{aligned}$$

the interchange of the order of integration and summation being justified by a limit theorem such as in [10] p. 104. Thus we have the moment generating function of the mixture as the mixture of moment generating functions of the $F_N(x, t)$:

$$\Phi_N(\theta, \alpha) = (1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} \phi_N(\theta, t) \quad \text{say,} \tag{3.1}$$

where

$$\phi_N(\theta, t) = \int_{0-}^{1} e^{\theta x}\, d_x F_N(x, t) = E(e^{\theta x_t}).$$

We now proceed to incorporate the assumption (2.1) into (3.1). Clearly

$$\begin{aligned}
\phi_N(\theta, t + 1) - \phi_N(\theta, t) &= E(e^{\theta x_{t+1}} - e^{\theta x_t}) \\
&= E\{e^{\theta x_t} E_t[e^{\theta(x_{t+1} - x_t)} - 1]\},
\end{aligned}$$

which by (2.1) is

$$\begin{aligned}
&\phi_N(\theta, t + 1) - \phi_N(\theta, t) \\
&\quad = N^{-m}\theta E(e^{\theta x_t}\{b - (b + c)x_t + s x_t(1 - x_t)[\tfrac{1}{2}(d + 1) - d x_t] \\
&\qquad\qquad + \tfrac{1}{4}\theta a x_t(1 - x_t) + W_N(\theta, t)\})
\end{aligned}$$

$$\begin{aligned}
&= N^{-m}\theta\{b\phi_N(\theta, t) - [b + c - \tfrac{1}{2}s(d + 1) - \tfrac{1}{4}\theta a][\partial\phi_N(\theta, t)/\partial\theta] \\
&\qquad - [\tfrac{1}{2}s(d + 1) + sd + \tfrac{1}{4}\theta a][\partial^2\phi_N(\theta, t)/\partial\theta^2] \\
&\qquad + sd[\partial^3\phi_N(\theta, t)/\partial\theta^3] + E[e^{\theta x_t} W_N(\theta, t)]\},
\end{aligned} \tag{3.2}$$

using $E(x_t^j e^{\theta x_t}) = [\partial^j \phi_N(\theta, t)/\partial\theta^j]$, $j = 1, 2, 3$. Multiplying both sides of (3.2) by $(1 - e^{-\alpha N^{-m}})e^{-\alpha N^{-m} t}$ and adding over all t, we get from the left hand side

$$\begin{aligned}(1 - e^{-\alpha N^{-m}}) &\sum_{t=0}^{\infty} e^{-\alpha N^{-m}t}[\phi_N(\theta, t + 1) - \phi_N(\theta, t)] \\ &= (1 - e^{-\alpha N^{-m}})e^{\alpha N^{-m}} \sum_{t=0}^{\infty} e^{-\alpha N^{-m}(t+1)}\phi_N(\theta, t + 1) \\ &\quad - (1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m}t}\phi_N(\theta, t) \\ &= e^{\alpha N^{-m}}[\Phi(\theta, \alpha) - (1 - e^{-\alpha N^{-m}})\phi_N(\theta, 0)] - \Phi_N(\theta, \alpha)\end{aligned}$$

by (3.1). Further, by (2.3), $\phi_N(\theta, 0) = E(e^{\theta x_0}) = e^{\theta p}$ independent of N, so the left of (3.2) has been transformed to

$$e^{\alpha N^{-m}}(1 - e^{-\alpha N^{-m}})[\Phi_N(\theta, \alpha) - e^{\theta p}]. \tag{3.3}$$

The same transformation on the right hand side of (3.2) produces

$$\begin{aligned} N^{-m}\theta(1 - e^{-\alpha N^{-m}}) &\sum_{t=0}^{\infty} e^{-\alpha N^{-m}t} \\ &\cdot\{b\phi_N(\theta, t) - [b + c - \tfrac{1}{2}s(d + 1) - \tfrac{1}{4}\theta a][\partial\phi_N(\theta, t)/\partial\theta] \\ &- [\tfrac{1}{2}s(d + 1) + sd + \tfrac{1}{4}\theta a][\partial^2\phi_N(\theta, t)/\partial\theta^2] \\ &+ sd[\partial^3\phi_N(\theta, t)/\partial\theta^3] + E[e^{\theta x_t}W_N(\theta, t)]\} \\ = N^{-m}&\theta\{b\Phi_N(\theta, \alpha) - [b + c - \tfrac{1}{2}s(d + 1) - \tfrac{1}{4}\theta a] \\ &\cdot[\partial\Phi_N(\theta, \alpha)/\partial\theta] - [\tfrac{1}{2}s(d + 1) + sd + \tfrac{1}{4}\theta a][\partial^2\Phi_N(\theta, \alpha)/\partial\theta^2] \\ &+ sd[\partial^3\Phi_N(\theta, \alpha)/\partial\theta^3] + (1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m}t}E[e^{\theta x_t}W_N(\theta, t)]\},\end{aligned} \tag{3.4}$$

the interchange of the order of summation and differentiation being justified by the fact that the series

$$(1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m}t}[\partial^j\phi_N(\theta, t)/\partial\theta^j], \qquad j = 1, 2, 3,$$

is composed of continuous functions and is uniformly convergent with respect to θ in $-B \leqq \theta \leqq B$.

Equating (3.3), (3.4), and after a little rearrangement, we find

$$\begin{aligned} \theta sd[\partial^3\Phi_N(\theta, \alpha)/\partial\theta^3] &- \theta[\tfrac{1}{2}s(d + 1) + sd + \tfrac{1}{4}\theta a][\partial^2\Phi_N(\theta, \alpha)/\partial\theta^2] \\ &- \theta[b + c - \tfrac{1}{2}s(d + 1) - \tfrac{1}{4}\theta a][\partial\Phi_N(\theta, \alpha)/\partial\theta] \\ &+ [\theta b - N^m e^{\alpha N^{-m}}(1 - e^{-\alpha N^{-m}})]\Phi_N(\theta, \alpha) \\ = -\theta(1 - e^{-\alpha N^{-m}})&\sum_{t=0}^{\infty} e^{-\alpha N^{-m}t}E[e^{\theta x_t}W_N(\theta, t)] \\ &- N^m e^{\alpha N^{-m}}(1 - e^{-\alpha N^{-m}})e^{\theta p}.\end{aligned} \tag{3.5}$$

4. Passage to the limit. For a fixed value of α ($\alpha > 0$), (3.5) may be considered as an ordinary differential equation with non-homogeneous terms. In this section we show that its solution $\Phi_N(\theta, \alpha)$ converges to a unique limit as $N \to \infty$.

Consider the distributions $\{G_N(x, \alpha)\}$; as N increases, either this sequence converges, or there exist two or more subsequences which converge to different limit functions. By the Montel-Helly Theorem [10] p. 103, we can in either case pick out a subsequence $\{G_{N^*}(x, \alpha)\}$ converging for all x to a monotonic limit $G(x, \alpha)$ say, which will itself be a distribution on [0, 1] because every $G_N(x, \alpha)$ is of this form. We shall subsequently prove that $G(x, \alpha)$ is uniquely determined from (3.5) and this requirement, and it follows that the entire sequence $\{G_N(x, \alpha)\}$ converges to this limit.

By the first limit theorem ([10] p. 104—note that continuity of the limit is not necessary) we see that the subsequence of generating functions $\{\Phi_{N^*}(\theta, \alpha)\}$ also converges, to $\Phi(\theta, \alpha)$ say, where $\Phi(\theta, \alpha)$ is the generating function corresponding to $G(x, \alpha)$, and this convergence is uniform for θ in any bounded interval $[-B, B]$. Furthermore, derivatives of $\Phi_{N^*}(\theta, \alpha)$ will converge to corresponding derivatives of $\Phi(\theta, \alpha)$. Before proceeding to the limit of (3.5), we make two observations concerning quantities therein:

(i) $N^m e^{\alpha N^{-m}}(1 - e^{-\alpha N^{-m}}) \to \alpha$ as $N \to \infty$

(ii) Because x_t is bounded in [0, 1] for all t, for θ in $[-B, B]$ we have

$$\begin{aligned}
&|\theta(1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} E[e^{\theta x_t} W_N(\theta, t)]| \\
&\qquad \leqq Be^B(1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} E|W_N(\theta, t)| \\
&\qquad = Be^B(1 - e^{-\alpha N^{-m}}) \left(\sum_{t=0}^{N^n u - 1} + \sum_{t=N^n u}^{\infty} \right) e^{-\alpha N^{-m} t} E|W_N(\theta, t)|,
\end{aligned}$$

where $N^n u$ is an integer, $u \geqq \epsilon > 0, 0 \leqq n < m$,

$$\begin{aligned}
&\leqq Be^B(1 - e^{-\alpha N^{-m+n}}) \operatorname{Max}_{0 \leqq t \leqq N^n u - 1} E|W_N(\theta, t)| \\
&\quad Be^B e^{-\alpha N^{-m+n} u} \operatorname{Max}_{t \geqq N^n u} E|W_N(\theta, t)|
\end{aligned}$$

replacing $E|W_N(\theta, t)|$ by its maximum in the appropriate ranges and summing the resulting geometric series. By the assumptions referring to (2.2), $W_N(\theta, t)$ is uniformly bounded with respect to t for θ in $[-B, B]$, so that the first term approaches zero as $N \to \infty$ because $-m + n < 0$ and therefore $1 - e^{-\alpha N^{-m+n}} \to 0$. The second term approaches zero by (2.2).

In view of the above limiting results, letting $N \to \infty$ through the sequence $\{N^*\}$, we have from (3.5)

$$\begin{aligned}
&\theta sd[\partial^3\Phi(\theta, \alpha)/\partial\theta^3] - \theta[\tfrac{1}{2}s(d + 1) + sd + \tfrac{1}{4}\theta a][\partial^2\Phi(\theta, \alpha)/\partial\theta^2] \\
&\quad - \theta[b + c - \tfrac{1}{2}s(d + 1) - \tfrac{1}{4}\theta a][\partial\Phi(\theta, \alpha)/\partial\theta] + (\theta b - \alpha)\Phi(\theta, \alpha) = -\alpha e^{\theta p}.
\end{aligned} \tag{4.1}$$

The third order equation (4.1) has three, linearly independent, solutions.

However, we seek a solution which is a moment generating function of a random variable on [0, 1], and this solution will be proved to be unique. Our solution must have the form

$$\Phi(\theta, \alpha) = \int_{0-}^{1} e^{\theta x} \, d_x G(x, \alpha),$$

where, for each $\alpha > 0$, $G(x, \alpha)$ is a distribution function on [0, 1]. Substituting into (4.1), and differentiating under the integral signs, we get

$$\begin{aligned} &\theta \int_{0-}^{1} e^{\theta x}\{b - (b + c)x + sx(1 - x)[\tfrac{1}{2}(d + 1) - dx]\} \, d_x G(x,\alpha) \\ &\qquad + \tfrac{1}{4}\theta^2 a \int_{0-}^{1} e^{\theta x} x(1 - x) \, d_x G(x, \alpha) = \alpha \int_{0-}^{1} e^{\theta x} \, d_x G(x, \alpha) - \alpha e^{\theta p}. \end{aligned} \tag{4.2}$$

Writing $e^{\theta p}$ as $\int_{0-}^{1} e^{\theta x} \, d_x F(x, 0)$ where $F(x, 0)$ is the distribution function of the constant $x_0 = p$, and integrating the two terms on the right of (4.2) by parts, gives

$$\begin{aligned} \alpha \int_{0-}^{1} e^{\theta x} \, d_x G(x, \alpha) &- \alpha \int_{0-}^{1} e^{\theta x} \, d_x F(x, 0) \\ &= \alpha[e^{\theta x}(G(x, \alpha) - F(x, 0))]_{0-}^{1} - \alpha\theta \int_{0-}^{1} e^{\theta x}[G(x, \alpha) - F(x, 0)] \, dx \\ &= -\alpha\theta \int_{0-}^{1} e^{\theta x}[G(x, \alpha) - F(x, 0)] \, dx, \end{aligned}$$

since $G(1, \alpha) = F(1,0) = 1$, $G(0-, \alpha) = F(0-, 0) = 0$. Substituting into (4.2), and cancelling θ, we obtain

$$\begin{aligned} &\int_{0-}^{1} e^{\theta x}\{b - (b + c)x + sx(1 - x)[\tfrac{1}{2}(d + 1) - dx]\} \, d_x G(x, \alpha) \\ &\qquad + \tfrac{1}{4}\theta a \int_{0-}^{1} e^{\theta x} x(1 - x) \, d_x G(x, \alpha) = -\alpha \int_{0-}^{1} e^{\theta x}[G(x, \alpha) - F(x, 0)] \, dx \end{aligned} \tag{4.3}$$

which holds at $\theta = 0$ by continuity, at which point we find

$$\begin{aligned} &\int_{0-}^{1} \{b - (b + c)x + sx(1 - x)[\tfrac{1}{2}(d + 1) - dx]\} \, d_x G(x, \alpha) \\ &\qquad = -\alpha \int_{0-}^{1} [G(x, \alpha) - F(x, 0)] \, dx. \end{aligned} \tag{4.4}$$

Define $H(x, \alpha)$ by

$$\begin{aligned} H(x, \alpha) = &\int_{0-}^{x} \{b - (b + c)x + sx(1 - x) \\ &\cdot[\tfrac{1}{2}(d + 1) - dx]\} \, d_x G(x, \alpha) + \alpha \int_{0-}^{x} [G(x, \alpha) - F(x, 0)] \, dx. \end{aligned} \tag{4.5}$$

Then clearly, $H(0-, \alpha) = 0$, and by (4.4), $H(1, \alpha) = 0$. Integrating the first and last elements of (4.3) by parts we obtain

$$\begin{aligned}\tfrac{1}{4}\theta a \int_{0-}^{1} e^{\theta x} x(1-x)\, d_x G(x, \alpha) &= -[e^{\theta x} H(x, \alpha)]_{0-}^{1} \\ &\quad + \theta \int_{0-}^{1} e^{\theta x} H(x, \alpha)\, dx = \theta \int_{0-}^{1} e^{\theta x} H(x, \alpha)\, dx,\end{aligned} \tag{4.6}$$

from which θ may be cancelled and the result will hold also at $\theta = 0$ by continuity.

By the uniqueness theorem for Laplace-Stietjes transforms, [21] p. 63, the equality (4.6) ensures that

$$\tfrac{1}{4}a \int_{0-}^{x} x(1-x)\, d_x G(x, \alpha) = \int_{0-}^{x} H(x, \alpha)\, dx, \tag{4.7}$$

at least for almost all x. The right hand side is a continuous function, however, while the left is monotonic since $x(1-x) \geqq 0$ and $G(x, \alpha)$ is monotonic in the range of integration, so that (4.7) holds for all x in $[0, 1]$. Further, since the right of (4.7) is differentiable with respect to x, so too is the left, and so too is $G(x, \alpha)$ except possibly where $x(1-x) = 0$. Hence we have

$$\tfrac{1}{4}ax(1-x)[\partial G(x, \alpha)/\partial x] = H(x, \alpha), \qquad 0 < x < 1. \tag{4.8}$$

But within the same interval, the right of (4.8) is differentiable, and by (4.5) we get

$$\begin{aligned}\tfrac{1}{4}ax(1-x)[\partial^2 G(x, \alpha)/\partial x^2] &+ \tfrac{1}{4}a(1-2x)[\partial G(x, \alpha)/\partial x] \\ &= \{b - (b+c)x + sx(1-x)[\tfrac{1}{2}(d+1) - dx]\}[\partial G(x, \alpha)/\partial x] \\ &\qquad + \alpha G(x, \alpha) - \alpha F(x, 0),\end{aligned}$$

that is,

$$\begin{aligned}&\tfrac{1}{4}ax(1-x)[\partial^2 G(x, \alpha)/\partial x^2] \\ &+ \{\tfrac{1}{4}a(1-2x) - b + (b+c)x - sx(1-x)[\tfrac{1}{2}(d+1) - dx]\} \\ &\cdot[\partial G(x, \alpha)/\partial x] - \alpha G(x, \alpha) = -\alpha F(x, 0), \qquad 0 < x < 1.\end{aligned} \tag{4.9}$$

We have thus reduced the problem to solving a second-order differential equation for $G(x, \alpha)$, of a type studied by Hille [7] and Feller [4], [5]. We need to show that this equation has only one solution which is a probability distribution on $[0, 1]$. Usually, a second-order equation requires two boundary conditions for a unique solution, whilst here we are imposing a restriction on the general form of the solution. However, except for special cases, we can impose conventional boundary conditions. Consider equations (4.5), (4.8). Letting $x \to 0$ from above, from (4.5) we see that

$$\lim_{x \to 0+} H(x, \alpha) = b[G(0+, \alpha) - G(0-, \alpha)].$$

But on the left of (4.8) we find

$$\lim_{x\to 0+} \tfrac{1}{4}ax(1 - x)[\partial G(x, \alpha)/\partial x] = 0,$$

because $\partial G/\partial x$ is integrable and cannot have a pole of order unity or higher as $x \to 0+$. Thus, from (4.8),

$$b[G(0+, \alpha) - G(0-, \alpha)] = 0, \tag{4.10}$$

and if $b > 0$, then $G(x, \alpha)$ has no jump at $x = 0$, and $G(0, \alpha) = 0$ is one boundary condition on (4.9). Similarly, if $c > 0$, then $G(x, \alpha)$ has no jump at $x = 1$, and $G(1, \alpha) = 1$ is another boundary condition on (4.9). The conclusion, then, is that if both (mutation rates) b, c are strictly positive, there can be determined, in view of the two boundary conditions, a unique solution of (4.9). We will subsequently show that this solution is a distribution.

Consider now the case when $b = c = 0$. Hille [7], see also [4] pp. 487, 488, has shown that the homogeneous equation corresponding to (4.9) has two linearly independent solutions whose behavior at $x = 0$, $x = 1$ depends on the coefficients of $\partial^2 G/\partial x^2$ and $\partial G/\partial x$ in (4.9). In the present case, these coefficients are $\frac{1}{4}ax(1 - x)$, and

$$\begin{aligned} &\tfrac{1}{4}a(1 - 2x) - sx(1 - x)[\tfrac{1}{2}(d + 1) - dx] \\ &\qquad = (d/dx)[\tfrac{1}{4}ax(1 - x)] - sx(1 - x)[\tfrac{1}{2}(d + 1) - dx], \end{aligned}$$

and the criterion of interest is the integrability, or otherwise, of the function

$$\frac{4}{ax(1 - x)} \exp\left\{\int^x \frac{sx(1 - x)[\frac{1}{2}(d + 1) - dx]}{\frac{1}{4}ax(1 - x)}\, dx\right\}$$

at $x = 0$ and $x = 1$. Clearly its integral does not converge at these points, which implies that the two linearly independent solutions mentioned above are such that one is unbounded at $x = 0$, the other at $x = 1$. Hence (4.9), the non-homogeneous equation, can have at most one bounded solution, that is, the particular solution, and therefore at most one solution which is a distribution function.

The cases $b = 0$, $c > 0$ and $b > 0$, $c = 0$ can easily be treated by a combination of the two previous arguments, so that if there exists a distribution on [0, 1] satisfying (4.9) and an appropriate (single) boundary condition, it will be unique. The actual existence of a distribution function is clear from the derivation of (4.9), or can be implied from the following section.

This completes the proof that $\{\Phi_N(\theta, \alpha)\}$ and $\{G_N(x, \alpha)\}$ converge to uniquely determined limits $\Phi(\theta, \alpha)$ and $G(x, \alpha)$. In the following section we investigate how these may be interpreted in connection with the original distribution $F_N(x, t)$.

5. Interpretation of the limit. Having assured that the problem has a unique solution, we may now verify, rather than deduce, the nature of that solution.

Consider the function $f(x, u)$ defined as follows:

(i) $f(x, u)$ satisfies the forward diffusion equation

$$(\partial f/\partial u) = \tfrac{1}{4}a(\partial^2/\partial x^2)[x(1-x)f] \tag{5.1}$$
$$-(\partial/\partial x)[\{b-(b+c)x+sx(1-x)[\tfrac{1}{2}(d+1)-dx]\}f],$$
$$0 < x < 1, u > 0,$$

(ii) subject to the boundary condition

$$\int_{0+}^{x} f(x, u)\,dx \to \begin{cases} 0 & x < p \\ 1 & x \geqq p \end{cases} \qquad \text{as } u \to 0,$$

(iii) and to the lateral conditions, when applicable: if $b > 0$, then

$$\lim_{x\to 0+}\{\tfrac{1}{4}a(\partial/\partial x)[x(1-x)f] \tag{5.2}$$
$$-(b-(b+c)x+sx(1-x)[\tfrac{1}{2}(d+1)-dx])f\} = 0, \qquad u > 0,$$

and/or if $c > 0$, then

$$\lim_{x\to 1-}\{\tfrac{1}{4}a(\partial/\partial x)[x(1-x)f] \tag{5.3}$$
$$-(b-(b+c)x+sx(1-x)[\tfrac{1}{2}(d+1)-dx])f\} = 0, \qquad u > 0.$$

Feller [4] has shown that these conditions are sufficient to determine a unique solution of (5.1), one for which $f(x, u) \geqq 0$ in $0 < x < 1$, $u > 0$ and

$$\int_{0+}^{1-} f(x, u)\,dx \leqq 1,$$

with equality when both b and c are positive. When $b > 0$, $x = 0$ is either an "entrance" or a "regular" boundary in Feller's terminology, depending on whether $b \geqq \frac{1}{4}a$ or $0 < b < \frac{1}{4}a$ apply, and in either case the lateral condition (5.2) may be imposed. However, if $b = 0$, (5.2) is not appropriate; its left side may be interpreted as the flux of probability into the "exit" or absorbing state $x = 0$. Thus the random variable x is continuous within $(0, 1)$ but has a discrete probability, $P_0(u)$ say, of being in the state 0 at time u, given by the solution of

$$dP_0(u)/du = \lim_{x\to 0+}\{\tfrac{1}{4}a(\partial/\partial x)[x(1-x)f(x, u)]$$
$$-(b-(b+c)x+sx(1-x)[\tfrac{1}{2}(d+1)-dx])f(x, u)\}$$

which in our case reduces to $dP_0(u)/du = \frac{1}{4}af(0+, u)$. Thus

$$P_0(u) = P_0(0) + \tfrac{1}{4}a\int_0^u f(0+, u)\,du \tag{5.4}$$

where

$$P_0(0) = \begin{cases} 0 & \text{if } x \neq 0 \text{ at time } u = 0, \\ 1 & \text{if } x = 0 \text{ at time } u = 0. \end{cases}$$

Similarly, if $c > 0$, the boundary $x = 1$ is either an entrance or a regular boundary (when $c \geqq \frac{1}{4}a$ or $c < \frac{1}{4}a$) and (5.3) may be applied. Otherwise when $c = 0$, the flux into the state $x = 1$ is minus the left hand side of (5.3) and writing $P_1(u)$ as the probability of $x = 1$ at time u,

$$P_1(u) = P_1(0) + \tfrac{1}{4}a \int_0^u f(1-, u)\, du \tag{5.5}$$

where

$$P_1(0) = \begin{cases} 0 & \text{if } x \neq 1 \text{ at time } u = 0, \\ 1 & \text{if } x = 1 \text{ at time } u = 0. \end{cases}$$

In summary, then, the solution $f(x, u)$ is a density function of a random variable, x, in $(0, 1)$ but with discrete probabilities $P_0(u)$ and $P_1(u)$ at $x = 0$, $x = 1$ which are non-zero only when $b = 0$ and $c = 0$ respectively.

Consider then the corresponding distribution function $F(x, u)$ defined by

$$F(x, u) = \begin{cases} 0, & x < 0, \\ P_0(u) + \int_{0+}^x f(x, u)\, dx, & 0 \leqq x < 1, \\ 1, & x \geqq 1, \end{cases} \tag{5.6}$$

with possible jumps of $P_0(u)$, $P_1(u)$ at $x = 0$ and $x = 1$. It is easy to verify that the function

$$G(x, \alpha) = \alpha \int_0^\infty e^{-\alpha u} F(x, u)\, du, \qquad \alpha > 0,$$

is a solution of (4.9) subject to (4.10) and the corresponding condition at $x = 1$. By the uniqueness already established in Section 4, we have therefore shown that $G_N(x, \alpha) \to G(x, \alpha)$ as $N \to \infty$, that is, from (1.2),

$$(1 - e^{-\alpha N^{-m}}) \sum_{t=0}^{\infty} e^{-\alpha N^{-m} t} F_N(x, t) \to \alpha \int_0^\infty e^{-\alpha u} F(x, u)\, du$$

for all $\alpha > 0$. But the left hand side can be written as the integral expression

$$(1 - e^{-\alpha N^{-m}}) N^m \int_0^\infty e^{-\alpha [N^m u] N^{-m}} F_N(x, [N^m u])\, du$$

where $[N^m u]$ here denotes the integral part of $N^m u$. As N increases,

$$(1 - e^{-\alpha N^{-m}}) N^m e^{-\alpha [N^m u] N^{-m}} \to \alpha e^{-\alpha u},$$

so we may write

$$\lim_{N \to \infty} \int_0^\infty e^{-\alpha u} F_N(x, [N^m u])\, du = \int_0^\infty e^{-\alpha u} F(x, u)\, du, \qquad \alpha > 0.$$

By the uniqueness theorem for Laplace transforms, this implies, at least for all u where $F(x, u)$ is continuous, that

$$F_N(x, [N^m u]) \to F(x, u) \tag{5.7}$$

where $F(x, u)$ is defined in (5.6), arising from the diffusion equation (5.1); the differentiability with respect to u implicit in (5.1), (5.4), (5.5), (5.6) ensures the convergence holds for all $u > 0$. Thus for large N, and the time-scale $t = N^m u$, $F(x, u)$ is the diffusion approximation to the distribution of the variate $x_{[N^m u]}$, and of $x_{N^m u}$ if we use the t time scale as being continuous with $x_t = x_{[t]}$.

This completes the main aim of the paper, namely, the proof of (0.3).

6. Applications of the diffusion approximation. In a second paper, it is planned to give specific examples of population models having a finite (but large) number of individuals, and showing that for them the above theory is applicable. For the present, we follow the example of many authors and use (5.1) as a starting point.

We first point out some of the misuses that have been made of diffusion theory. It is unimportant whether the diffusion equation (5.1) is written in terms of the u or t time-scales, provided the order of magnitudes of the first and second moments (2.4), (2.5) are the same. However in some published versions of (5.1), a factor N^{-1} has been applied to the term $\frac{1}{4}a(\partial^2/\partial x^2)[x(1-x)f]$ but not to the first derivative term. This implies that mutation and selection are of higher order of magnitude in N than is the sampling variance, and the correct approach would be to use asymptotically deterministic, rather than diffusion, theory. On occasion, the solution of (5.1) has been obtained by ignoring the lateral conditions (5.2), (5.3) when applicable, or by using them when inappropriate.

Perhaps the most controversial point is the interpretation of the "stationary distribution," $\lim_{u\to\infty} F(x, u)$. Equating $\partial f/\partial u$ to zero in (5.1), and dropping the dependence on u, the first integral becomes

$$\tfrac{1}{4}a(d/dx)[x(1-x)f(x)] \\ - \{b - (b+c)x + sx(1-x)[\tfrac{1}{2}(d+1) - dx]\}f(x) = C_1,$$

say, where C_1 is a constant. Provided one or other (or both) of b and c are positive, (5.2) or (5.3) imply $C_1 = 0$. A second integration results in

$$(6.1)\quad f(x) = C_2 \exp\{2sa^{-1}[(d+1)x - dx^2]\}x^{4ba^{-1}-1}(1-x)^{4ca^{-1}-1}.$$

The constant C_2 must be chosen so that $\int_0^1 f(x)\,dx \leqq 1$ with equality when $b > 0$, $c > 0$ both hold. Thus if either b or c is zero, C_2 must also be zero. In genetic terms, if mutation exists in both directions, the density function $f(x, u)$ approaches the stationary density (6.1); if one or other mutation rate is zero, $f(x, u) \to 0$ as $u \to \infty$ and the population is ultimately absorbed in either state $x = 0$ or $x = 1$ where all individuals are of the same homozygous genotype. When $b = c = 0$, the constant C_1 cannot be determined by (5.2) or (5.3), and the general stationary solution of (5.1) has the form

$$(6.2)\quad \begin{aligned} f(x) &= \exp\{2sa^{-1}[(d+1)x - dx^2]\} \\ &\quad \cdot \left[C_2 + C_1 \int^x e^{-2sa^{-1}[(d+1)x-dx^2]}\,dx\right] x^{-1}(1-x)^{-1}, \end{aligned}$$

but again this function is not integrable on [0, 1] unless $C_1 = C_2 = 0$, that is, $f(x) \equiv 0$. As before, the stationary distribution is purely discrete at $x = 0$ and $x = 1$.

There has been some debate as to what significance, if any, should be attached to non-zero solutions (6.1), (6.2) when mutation is absent in one or both directions. Fisher [6] held the view that diffusion theory was inapplicable near the boundaries $x = 0, x = 1$; and used techniques for branching process models to investigate the discrete probabilities at values of x of the form $x = r/2N$, r fixed. Diffusion theory would collapse all such states into the single limit $x = 0$. However, in my opinion, the only valid objections to diffusion theory are that the original assumptions may not hold in a specific model, or that the continuous approximation is not adequate for small populations. But aside from these objections, there seems no reason to discard the solution $f(x, u)$, nor its limiting value of zero, even in the tails of the distribution, provided one also recognizes the role $P_0(u)$ and $P_1(u)$ play. There have been attempts to find solutions to (5.1) which have constant flux into the boundaries; this surely cannot mean solutions for which $(d/du)[P_0(u) + P_1(u)]$ is constant, for then probabilities greater than unity will be encountered. But in any case, although the proportional flux $[1 - P_0(u) - P_1(u)]^{-1}(d/du)[P_0(u) + P_1(u)]$ is asymptotically constant as $u \to \infty$, such a condition cannot be imposed arbitrarily for finite times because of the definiteness of (5.4) and (5.5), nor in my mind could it be imposed on an actual population except by artificial means outside of, and invalidating, the stochastic model.

To further investigate the roles that (6.1) and (6.2) do *not* play in the solution of diffusion problems without mutation, we consider the special case when selection is also absent; thus, with $b = c = s = d = 0$, (5.1) becomes simply

$$\partial f(x, u)/\partial u = \tfrac{1}{4}a(\partial^2/\partial x^2)[x(1 - x)f(x, u)], \qquad 0 < x < 1, u > 0, \tag{6.3}$$

and the integrable solution consistent with a fixed initial value $x_0 = p$ has been found by Kimura [13] p. 890:

$$\begin{aligned} f(x, u) = p(1 - p) \sum_{i=1}^{\infty} (2i + 1)i(i + 1) \\ \cdot F(1 - i, i + 2, 2; p)F(1 - i, i + 2, 2; x)e^{-\frac{1}{4}i(i+1)au}. \end{aligned} \tag{6.4}$$

For large u, we have from the first term in this series

$$f(x, u) \doteqdot 6p(1 - p)e^{-\frac{1}{2}au}. \tag{6.5}$$

But as the stationary solution of (6.3) is

$$f(x) = [C_2/x(1 - x)] + [C_1/(1 - x)], \tag{6.6}$$

we see that this has little if any connection with the transient solution (6.4), nor even with the stationary density conditional on no absorption, which from (6.5) is clearly uniform for x on (0, 1).

There are at least two other interpretations of (6.2). With a suitable choice of constants, (6.2) is an approximation to (6.1) when the mutation rates b and c are small but not zero, at least for most of the range $0 < x < 1$ but not near the endpoints. This does not appear to be a very interesting interpretation, for (6.1) is no more complicated than (6.2) and an approximation seems unnecessary. Kolmogorov has given an interpretation of (6.6) in an abstract, [14]. However, the abstract is so brief that I am unable to grasp the import of his interpretation.

Turning to less controversial aspects of the case when there are no mutations, one is interested in the rate of approach to homozygosity, that is, the rate at which fixation or loss of the "a" gene takes place. As mentioned earlier, with $b = c = 0$ the boundaries $x = 0$, $x = 1$ are exits, and according to Feller [5] p. 11, an exit boundary is accessible, that is, there is a positive probability it will be reached from the interior of (0, 1) in a finite time. Feller gives several theorems concerning first passage time problems, some of which will be described below.

THEOREM (Feller [5]): *Let* $x_0 = p$ *be the initial value, and* $P_1(p, u)$ *the probability of absorption at the boundary* $x = 1$ *before time* u. *Then for the diffusion process* (5.1), *the Laplace transform*

$$z(p, \lambda) = \int_0^\infty e^{-\lambda u} P_1(p, u)\, du \tag{6.7}$$

satisfies

$$\tfrac{1}{4}ap(1 - p)(\partial^2 z/\partial p^2) + sp(1 - p)[\tfrac{1}{2}(d + 1) - dp](\partial z/\partial p) - \lambda z = 0,$$

with boundary conditions $z(0, p) = 0$, $z(1, p) = \lambda^{-1}$.

More directly, Kimura [13] p. 896 uses the equivalent result that

$$\begin{aligned} &\tfrac{1}{4}ap(1 - p)(\partial^2 P_1/\partial p^2) \\ &\qquad + sp(1 - p)[\tfrac{1}{2}(d + 1) - dp](\partial P_1/\partial p) = (\partial P_1/\partial u), \end{aligned} \tag{6.8}$$

with boundary conditions $P_1(0, u) = 0$, $P_1(1, u) = 1$. Note that these methods are alternatives to the application of (5.5); however, as (6.8) is the backward equation to (5.1), the two methods are equivalent and have the same complexities for solution. In the particular case with no mutation or selection, the solution, by either direct solving of (6.8) or by use of (6.4), (5.5), is

$$\begin{aligned} P_1(p, u) = p + p(1 - p)\sum_{i=1}^{\infty} (2i + 1)(-1)^i \\ \cdot F(1 - i, i + 2, 2; p)e^{-\frac{1}{4}i(i+1)au}. \end{aligned} \tag{6.9}$$

As well as this, Kimura gives methods for solution in cases with selection, in particular when the dominance coefficient takes the values $d = -1, 0$, or $+1$. No general solution of (6.8) has been found explicitly, but Kimura [13] p. 896 has obtained its smallest eigenvalue, λ_0 say, as a power series expansion in the selection coefficients s and d. Thus

$$\lambda_0 = \tfrac{1}{2}a[1 + K_1 sa^{-1} + K_2(sa^{-1})^2 + K_3(sa^{-1})^3 + K_4(sa^{-1})^4 + \cdots] \tag{6.10}$$

where

$$K_1 = -\frac{1}{5}d, \qquad K_2 = \frac{1}{2\cdot 5} + \frac{2^2\cdot 3}{5^3\cdot 7}d^2, \qquad K_3 = \frac{1}{2\cdot 5^3\cdot 7}d - \frac{2^2}{5^5\cdot 7}d^3,$$

$$K_4 = -\frac{1}{2^3\cdot 5^3\cdot 7} - \frac{7^3}{2\cdot 3^3\cdot 5^5}d^2 - \frac{2^2\cdot 3^5}{5^6\cdot 7^3\cdot 11}d^4.$$

The leading transient term in the eigenfunction expansion of $P_1(p, u)$ has a factor $e^{-\lambda_0 u}$, and thus λ_0 governs the asymptotic rate of approach of $P_1(p, u)$ to its limit, and also, of course, of $P_0(p, u)$ which satisfies (6.8) with the boundary conditions reversed.

The limiting value of $P_1(p, u)$ as $u \to \infty$ is the probability of ultimate absorption in state $x = 1$, that is, when all genes are of the type, "a". This limit, say $P_1(p, \infty)$, satisfies the stationary equation

$$\tag{6.11} \tfrac{1}{4}ap(1-p)(dP_1/dp^2) + sp(1-p)[\tfrac{1}{2}(d+1) - dp](dP_1/dp) = 0,$$

with boundary conditions $P_1(0, \infty) = 0$, $P_1(1, \infty) = 1$; for a theoretical discussion see [5]. The solution is ([13] p. 896)

$$\tag{6.12} P_1(p, \infty) = \int_0^p e^{-2sa^{-1}[(d+1)x-dx^2]}\,dx \Big/ \int_0^1 e^{-2sa^{-1}[(d+1)x-dx^2]}\,dx.$$

In particular, with no selection, this reduces to

$$\tag{6.13} P_1(p, \infty) = p,$$

which agrees with the limit of (6.9). Similarly, $P_0(p, \infty) = 1 - P_1(p, \infty)$ since absorption is certain to occur in one or other state. Equation (6.12) has been used to find the probability of survival of a single mutant in a population by taking $p = \frac{1}{2}N^{-1}$; this initial state does not satisfy our assumption (2.3), nor is it appropriate for dioecious models with x_0 defined by (0.1) unless the sex numbers N_1 and N_2 happen to be equal. The initial value x_0 would depend on which sex the mutant occurred in:

$$\tag{6.14} x_0 = \begin{cases} 1/4N_1 & \text{if male,} \\ 1/4N_2 & \text{if female.} \end{cases}$$

However, it has been verified [19] for some finite Markov chain models that the result (6.13), obtained without selection, is an exact result, and the substitution of (6.14) into (6.13) is therefore valid.

The function $P_0(p, u) + P_1(p, u)$ is the probability of absorption, in either state, by time u, and is thus the distribution function for the first absorption time. The evaluation of the moments of this distribution, and in particular, the expected time required for the population to become homozygous, seems not to have been made by other authors. The theoretical result has been mentioned by Feller [5], and in our case gives the following: Let

$$U(p) = \int_0^\infty u\, d_u[P_0(p, u) + P_1(p, u)];$$

then $U(p)$ is the solution of

$$\tfrac{1}{4}ap(1-p)(d^2U/dp^2) + sp(1-p)[\tfrac{1}{2}(d+1) - dp](dU/dp) = -1 \tag{6.15}$$

with boundary conditions $U(0) = U(1) = 0$. The solution is

$$\begin{aligned} U(p) = {} & 4a^{-1}p \int_0^1 [g(x)]^{-1} \left[\int_{\frac{1}{2}}^x g(y)y^{-1}(1-y)^{-1}\, dy \right] dx \\ & - 4a^{-1} \int_0^p [g(x)]^{-1} \left[\int_{\frac{1}{2}}^x g(y)y^{-1}(1-y)^{-1}\, dy \right] dx, \end{aligned} \tag{6.16}$$

where $g(y) = \exp\{4sa^{-1}[\tfrac{1}{2}(d+1) - dy]\}$. In particular with no selection, $s = d = 0$, (6.16) simplifies to

$$U(p) = -4a^{-1}[p \log p + (1-p)\log(1-p)]. \tag{6.17}$$

It may be verified that (6.17) is the mean of $P_0(p, u) + P_1(p, u)$ where $P_0(p, u) = P_1(1-p, u)$ given by (6.9); but the general expression (6.16) can only be obtained in the above fashion for lack of a general solution for $P_0(p, u) + P_1(p, u)$. Higher order moments can also be found by solving ordinary differential equations analogous to (6.15); see [20] for an example of this.

The result (6.17) strongly suggests that the function $U(p)$ can be utilized to define an "entropy" for diffusion processes with absorbing states. Moran [18] has given definitions of entropy for completely regular Markov processes which would apply in our situation when mutation is present. With mutation absent, a new definition is needed. Entropy can be defined either for a particular time in a process, or for a particular state. Thus, we might define the entropy of state $x = p$ as being given by (6.15), and the ultimate value of the entropy is zero since then $x = 0$ or $x = 1$. And we would have the (trivial) result that the expected time to reach an absorbing state is equal to the initial entropy. However, there is no certainty that the entropy so defined would be monotonic decreasing with time. Alternatively, we might define the entropy at time u as

$$H(u) = \int_0^1 U(x)\, d_x F(x, u) = \int_{0+}^{1-} U(x) f(x, u)\, dx. \tag{6.18}$$

Then, the initial and ultimate values are $U(p)$ and 0 respectively with the previous interpretation, but in addition, $H(u)$ is a decreasing function of u. For, from (6.18),

$$\begin{aligned} (dH(u)/du) &= \int_{0+}^{1-} U(x)(\partial f(x, u)/\partial u)\, dx \\ &= \int_{0+}^{1-} \{\tfrac{1}{4}ax(1-x)(d^2U(x)/dx^2) \\ &\qquad + sx(1-x)[\tfrac{1}{2}(d+1) - dx](dU(x)/dx)\} f(x, u)\, dx, \end{aligned}$$

using (5.1) and integrating by parts twice, $= -\int_{0+}^{1-} f(x, u)\,dx$ from (6.15), $\leqq 0$ as required. The result allows other formulations for $H(u)$, namely,

$$\begin{aligned} H(u) &= H(0) + \int_0^u [F(0+, u) - F(1-, u)]\,du \\ &= U(p) + \int_0^u [F(0+, u) - F(1-, u)\,du \\ &= U(p) - u + \int_0^u [P_0(p, u) + P_1(p, u)]\,du. \end{aligned}$$

Whether such a definition of entropy will serve any useful purpose is, of course, another question.

In the results of Section 6, we have used throughout the time-scale of u, not t. One u-unit corresponds to N generations, or to N^2 birth-death events, in the finite population, and thus if the generation is taken as time unit, the eigenvalue (6.10) should be scaled as $N^{-1}\lambda_0$, the expected time for absorption to occur becomes $NU(p)$, and the transient solutions (6.4), (6.5), (6.9) should be similarly adjusted.

Acknowledgments. The author thanks Professor P. A. Moran, Mr. J. E. Moyal and a referee for valuable suggestions and discussion.

REFERENCES

[1] Bharucha-Reid, A. T. (1960). *Elements of the Theory of Markov Processes and their Applications.* McGraw-Hill, New York.

[2] Crow, J., and Kimura, M. (1956). Some genetic problems in natural populations **4** 1–22. *Proc. Third Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[3] Feller, W. (1951). Diffusion processes in genetics, 227–246. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[4] Feller, W. (1952). The parabolic differential equation and the associated semi-group of transformations. *Ann. Math.* **55** 468–519.

[5] Feller, W. (1954). Diffusion processes in one dimension. *Trans. Amer. Math. Soc.* **77** 1–31.

[6] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford Univ. Press.

[7] Hille, E. (1949). On the integration problem for Fokker-Planck's equation in the theory of stochastic processes, 183–194. *Onzième Congr. des Math. Scand.*

[8] Karlin, Samuel, and McGregor, James. On a genetics model of Moran. *Proc. Cambridge Philos. Soc.* **58** 299–311.

[9] Kempthorne, Oscar. (1956). *An Introduction to Genetic Statistics.* Wiley, New York.

[10] Kendall, M. G., and Stuart, A. (1958). *The Advanced Theory of Statistics*, **1**. Griffin, London.

[11] Khintchine, A. (1948). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Chelsea, New York.

[12] Kimura, M. (1955). Stochastic processes and distribution of gene frequencies under natural selection, **20** 33–53. *Cold Spring Harbor Symp. on Quantitative Biology.* Biological Lab. Cold Spring Harbor., L. I., New York.

[13] Kimura, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28** 882–901.

[14] Kolmogorov, A. N. (1959). Transition of branching processes into diffusion processes

and some problems in genetics (in Russian). *Teor. Veroyatnost. i Primenen.* **4** 233–236.

[15] Moran, P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71.

[16] Moran, P. A. P. (1958). A general theory of the distribution of gene frequencies. I. Overlapping generations. *Proc. Roy. Soc. London, Ser. B* **149** 102–112.

[17] Moran, P. A. P. (1958). A general theory of the distribution of gene frequencies. II. Non-overlapping generations. *Proc. Roy. Soc. London, Ser. B* **149** 113–116.

[18] Moran, P. A. P. (1961). Entropy, Markov processes and Boltzmann's H-theorem. *Proc. Cambridge Philos. Soc.* **57** 833–842.

[19] Watterson, G. A. (1959). Non-random mating, and its effect on the rate of approach to homozygosity. *Ann. Human Genetics* **23** 204–220.

[20] Watterson, G. A. (1961). Markov chains with absorbing states: a genetic example. *Ann. Math. Statist.* **32** 716–729.

[21] Widder, D. V. (1946). *The Laplace Transform.* Princeton Univ. Press.

[22] Wright, S. (1945). The differential equation of the distribution of gene frequencies. *Proc. Nat. Acad. Sci.* **31** 382–389.