STOCHASTIC PROCESSES IN POPULATION GENETICS

by

Warren John Ewens

A thesis submitted for the degree of
Doctor of Philosophy in the Australian
National University

1963

Preface

This thesis considers the mathematical behaviour of various genetical populations. I was introduced to this subject by my supervisor, Professor P.A.P. Moran, who has suggested most of the problems considered here and who has guided me throughout in their investigation. To him I give my most sincere thanks.

The contents are the original work of the author, except that Chapter 2 is based on a paper (Ewens and Gani, (1961) written jointly with J.M. Gani. It is difficult to sort out the contributions of Gani and myself, and of the part of the paper considered, roughly half is due to each author. Some material due entirely to Gani has been omitted here. Chapter 3 is based on a paper (Ewens (1963a)) of the author, but extra material has been included in this thesis. I should like to record my debt to Dr. G. A. Watterson for discussion on this paper. Chapter 4 is based on a paper (Ewens (1963c)) of the author, while the material of Chapter 5 is being prepared for publication. Chapter 6 is similarly based on published work (Ewens (1963b)), but extra material is here included. Chapters 7 and 8 jointly follow published work (Ewens (1963d)), as also does Chapter 9, (Ewens, (1963e)).

I should like to thank Mrs. Betty Moore for the excellent work she has done in typing this thesis.

W.J. Ewens

CONTENTS

SUMMARY

The various chapters form a consecutive discussion of the problems considered in this thesis. At the same time each chapter is to some extent self-contained and treats a particular facet of the main discussion. Apart from Chapter 1, which is introductory, the chapters may be summarized as follows. Chapter 2 presents a general method for determining survival probabilities for any given allele in various genetical models. If exact results may not be found a method for determining bounds is described. Comparison is then made with the results obtained by other methods. In Chapter 3 two concepts used in some of the remaining chapters are introduced. The first of these is the diffusion pseudo-transient distribution, together with an associated pseudo-transient function. These were derived originally to give a meaning to a function obtained by formal operations on a diffusion equation. (In Chapters 4, 5, and 9 some exact (discrete) pseudo-transient distributions and functions are introduced). The second concept introduced in Chapter 3 is that of an "almost-invariant" function, which is essentially either a martingale or a semi-martingale (cf. Doob, (1953)), and which is used subsequently to derive bounds for exact values for which diffusion methods supply approximations. Chapter 4 extends previous work by Watterson (1961) on the mean time until homozygosity

in the genetical model first introduced by Moran, (1958a). The methods used here differ from those of Watterson. Chapter 5 extends the results of Chapter 4 and provides methods for deriving most functions of interest when the transition matrix of the Markovian variate under consideration is a continuant. This allows a treatment to be made, amongst others, of the case where Moran's model is generalized so that selective advantages may depend themselves on gene frequency. In Chapter 6 numerical values, obtained by using an electronic computor, are compared with diffusion approximations in the genetical model of Wright (1931), for which very few exact results have been found. The result of this chapter is to show how remarkably accurate diffusion approximations can be even for extremely small population sizes. The remaining three chapters treat diploid populations; Chapter 7 covers the case where selective advantages are constant and Chapter 8 the case where they depend on gene frequency. For these two chapters exact results seem very difficult to derive and diffusion approximations only are considered. Chapter 9 is different to the remaining chapters in that it provides an example of a case where diffusion methods should not be used. Here the results obtained by previous authors are discussed and it is argued that diffusion methods have been used when they should not, and that even if diffusion methods were applicable, they have been used in the wrong way. An alternative method of analysis, and indeed an

alternative problem to be discussed, are proposed and some exact results are obtained.

CHAPTER I

INTRODUCTION

This thesis presents a discussion of the stochastic behaviour of various genetical populations. It will be concerned with some characteristic of the individuals in the population under consideration which is controlled by a single locus on a chromosome, so that for example questions of linkage will not be considered. Thus for convenience individuals will be referred to as being (say) AA, meaning that this is the genotype at the locus under consideration. The population will either be haploid, in which one of two possible alleles A and a is allowed at the locus, or else diploid, for which the possible genotypes are AA, Aa, and aa. The case where more than two alleles are allowed at any locus will be considered only for the case of self-sterility populations discussed in Chapter 9. The individuals will throughout be regarded as being monoecious, so that any individual may act as male or female parent. This presents a significant simplification over the case where the individuals are dioecious, that is either male or female, which is discussed later in this chapter.

In all cases the population size will be regarded as remaining effectively fixed (at a constant usually denoted N). Such an assumption will limit the application of the results obtained, but qualitatively the results should hold when the population size is reasonably stable. The above restriction is made in the first instance so that the population will not die out completely, but is also needed, as will be shown later, to ensure that for the populations under consideration a Markovian variate may be found in

terms of which the behaviour of the population may be described.

The restriction to constancy of population size may be relaxed immediately to the case where the population size assumes a cyclic sequence of values, say $N_1\ N_2 \ldots N_k\ N_1\ N_2 \ldots$, as has been observed approximately in the Canadian lynx population. In such populations it will often be the case that a variate may be found which is Markovian in the sense that the values of this variate at successive instances when the population size is (say) N_1 form a Markov chain. In this case the transition matrix would probably be extremely complicated and of no direct use in describing the population behaviour, but the Markovian property by itself will be sufficient for the application of some of the methods considered later. It is, in fact, easily shown that if k is moderate and the N_i are large, then the population behaves effectively as a population of fixed size N, where $kN^{-1} = N_1^{-1} + \ldots + N_k^{-1}$.

The fact that a single Markovian variate can often be found which describes the population behaviour will be used frequently in this thesis, and attention is restricted to the case where such a variate exists. The existence of such a variate will be useful in two ways. In the first instance it may be possible to use directly the transition matrix of the variate to find quantities of interest to the geneticist. Secondly if the transition matrix is too unwieldy for direct use, it will be possible in many cases to use diffusion methods to approximate to the required quantities, and the application of such methods depends on the Markovian nature of the

variate under consideration. The case where no Markovian variate exists but where a "quasi-Markovian" variate may be found has been discussed by Watterson (1960, 1962), who has shown that such variates may be treated by using diffusion methods. However for these variates it would be very difficult to find the bounds on exact values which are derived here for the case in which a Markovian variate can in fact be found.

The restriction to the case where a Markovian variate exists is not so severe as might at first be thought. In Chapters 7 and 8 a single Markovian variate may be found for diploid populations, which allow three possible genotypes. In this case only the Markovian nature of the variate is used to justify the use of diffusion methods and the transition matrix is not considered. In Chapter 2 a situation is discussed when a number of geographically distinct subpopulations exists for which the different subpopulations have different genetic properties (e.g. different selective advantages for a given allele). Nevertheless, given a sufficient amount of migration between the subpopulations it is possible to find a single Markovian variate describing the joint behaviour of the subpopulations. The restriction to monoecious populations mentioned above is made so that a Markovian variate may be found; for dioecious populations this seems impossible under any reasonable population model.

If the population size remains fixed, then in general one or other of the alleles will eventually be lost from the population by random elimination. Here the probability that the allele so

eliminated is A, the mean time until elimination, the variance of this time, the probability that one or other allele has been eliminated by a particular time, the transient behaviour of the population, and other quantities, will be discussed. The effect on these quantities of selection, dominance, mutation, migration, and other influences, is considered. In the case where both alleles may mutate, a stationary distribution may be found for any allele, and the effect of the above influences on this distribution will also be considered. Some discussion is given to finding the latent roots of the transition matrix under consideration, and in particular to the largest non-unit latent root. This is done since by writing the transition matrix and its powers out in spectral form it is clear that these roots, and in particular the largest non-unit root, describe in a sense the rate of approach of the population to homozygosity (i.e. only one allele present), or to the stationary distribution in the case where this will exist. However the usefulness of the latent roots in this respect suffers since in general the corresponding spectral matrices are not known. It may in fact be more useful and even easier to find simply the mean and the variance of the time taken until homozygosity.

It has been mentioned previously that in many cases where exact treatment is too difficult it will be often possible to use diffusion methods to find approximations for the various quantities under consideration. Such diffusion methods are outlined in Chapter 3, but here a slightly different derivation is used which enables bounds to be obtained for the exact value being

approximated. The application of diffusion methods in genetics is very wide, and has the virtue that many of the resulting expressions are given in terms of quite simple functions. For this reason diffusion approximations may even be preferred to exact values, although the latter are known, if the expressions for the exact values are complicated. Needless to say diffusion methods have sometimes been used uncritically, and in Chapter 9 a population is discussed for which this is true. In Chapters 2, 4, and 5, exact values are derived which are shown analytically to be close to their various diffusion approximations, while in Chapter 6 a model is discussed where the diffusion approximations and exact values are compared numerically.

Broadly, the results of the thesis may be summarized as follows.

(i) Diffusion methods provide very close approximations to exact values when they are applicable, even when the population size is small.

(ii) When diffusion methods are inapplicable the results derived formally from them are valueless.

(iii) It is possible to derive a diffusion approximation and in some cases exact values for a distribution which describes in a sense the transient behaviour of genetical populations, and this distribution may be used to derive other functions (e.g. absorption probabilities, mean absorption times, mean occupancy times, variances of absorption times).

(iv) Bounds derived by diffusion methods and similar in form to diffusion approximations may be found within which the value of

a function must lie, when this value cannot be found explicitly.

(v) When individuals in the population die one by one rather than a generation at a time, it is possible to derive exact results for most functions even when the various probabilities in the transition matrix are complicated.

(vi) In some cases it is possible to use diffusion methods to find approximate results in the case where selective advantages are allowed to vary, and in some such cases the selective advantages may be ignored altogether, in other cases they may be treated as being constant.

(vii) The transient behaviour of a population may not be investigated by considering the stationary behaviour of the same population when mutation exists, with the mutation rates being allowed to approach zero. This occurs because of the different allocation of a certain fundamental constant in the two cases.

CHAPTER 2.

ABSORPTION PROBABILITIES

2.1 Introduction

In this chapter a general method is considered for determining survival probabilities or bounds for survival probabilities in haploid populations of fixed size N where selection is allowed. It is supposed that the two types of individual in the population are A and a, corresponding to the two possible alleles at the locus under consideration, and attention is concentrated on the number k of A individuals. The value of k at time t is denoted k_t, and the aim is to find some non-zero constant θ (independent of k_t) solving the equation

$$\mathbb{E}\left[\exp \theta(k_{t+1} - k_t) \mid k_t\right] = 1 \qquad (2.1)$$

If it is not possible to solve this equation independently of k_t, then we try to find bounds for the solution of (2.1) as k_t takes all possible values.

Note that the expectation in (2.1) will henceforth be taken always to mean the expectation conditional on k_t, so that explicit statement of this conditioning is dropped from now on.

Equations analogous to (2.1) are used to derive power and A.S.N. curves in sequential analysis, and in the case $k_t = 1$, (2.1) is equivalent to the equation

$$z = p(z) \qquad (2.2)$$

used to derive survival probabilities in branching processes.

2.2 Description of Method

It is supposed that k_t is a Markovian variate, so that with it may be associated a $(N+1)x(N+1)$ transition matrix $P = \{p_{ij}\}$, where

$$p_{ij} = \text{Prob}\left\{ k_{t+1} = j \mid k_t = i \right\} \qquad i,j = 0,1,\ldots,N$$

If the state E_i corresponds to the event "number of A individuals = i" $(i = 0,1,\ldots,N)$ then if there is no mutation the states E_0 and E_N will be absorbing, all other states will be transient, and P may be written in the partitioned form

$$P = \left(\begin{array}{c:c:c} 1 & \underline{0}' & 0 \\ \hdashline R_0 & Q & \underline{R}_N \\ \hdashline 0 & \underline{0}' & 1 \end{array}\right)$$

where $\underline{R}_0$, $\underline{R}_N$ are column vectors having elements p_{i0} p_{iN} $(i = 1,2,\ldots,N-1)$ resp. It is well-known that the column vector

$$\underline{\xi} = \left[I - Q \right]^{-1} \underline{R}_N$$

gives the set of survival probabilities for A individuals for $k_0 = 1,2,\ldots,N-1$, but this result is not useful for our purposes and we obtain $\underline{\xi}$, or bounds for $\underline{\xi}$, as follows. Let the matrix $P(\theta)$ be defined by

$$P(\theta) = \left\{ p_{ij} \exp (j-i)\theta \right\}$$

so that $P(0) \equiv P$. Then following Bartlett (1955), section 2.22, the moment-generating function $M^{(t)}(\theta|k_0)$ of the variate $k_t - k_0$ may be written as

$$M^{(t)}(\theta|k_0) = \underset{\sim}{r}'(k_0)\, P^t(\theta)\, \underset{\sim}{\psi} \qquad (2.3)$$

where $\underset{\sim}{r}'(k_0)$ is the row vector $(0\ldots010\ldots0)$, where the unity occurs in the position k_0, and $\underset{\sim}{\psi}$ is a column vector each of whose elements is unity. It follows, since P is the transition matrix of a finite Markov chain with absorbing states, that

$$\lim_{t\to\infty} P^t(\theta) = \begin{pmatrix} 1 & \vdots & \underset{\sim}{0}' & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_0(\theta) & \vdots & 0 & \vdots & S_N(\theta) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \vdots & \underline{0}' & \vdots & 1 \end{pmatrix}$$

where $S_0(\theta)$, $S_N(\theta)$ are column vectors, each having N-1 elements which are

$$P(0,i)\exp(-i\theta), \quad P(N,i)\exp(N-i)\theta \qquad (i=1,2,..,N-1)$$

respectively. The $P(N,i)$ are survival probabilities for A individuals for $k_0 = i$ $(i = 1,2,\ldots,N-1)$, and $P(0,i) = 1-P(N,i)$. Thus from (2.3) it follows that

$$\lim_{t\to\infty} M^{(t)}(\theta|k_0) = \left\{1-P(N,k_0)\right\}\exp(-k_0\theta) + P(N,k_0)\exp(N-k_0)\theta$$

$$\ldots \qquad (2.4)$$

The right-hand side in (2.4) is unity for $\theta=0$, and in general there

will exist a unique non-zero value of θ for which the right-hand side in (2.4) is again unity. If such a value of θ is known it is easy to solve for $P(N,k_0)$. It is now shown how such a value of θ may be found or approximated by relating the required value to the non-zero solution for θ of the equation

$$M(\theta|k_0) \equiv M^{(1)}(\theta|k_0) = 1 \qquad (2.5)$$

To do this we establish

Lemma 2.1

If for all i $(i = 1,2,..,N-1)$ $M(\theta|i) \to \infty$ as $\theta \to \pm \infty$, then the same is true of $M^{(t)}(\theta|i)$ $(t > 1)$.

Proof

Since $M(\theta|i) \to \infty$ as $\theta \to \pm \infty$, there will exist two values of θ, namely $\theta_u > 0$ and $\theta_\ell < 0$, for which

$$M(\theta|i) > 1 \quad \text{for } \theta \geq \theta_u \qquad (i = 1,2,\ldots,N-1)$$

and

$$M(\theta|i) > 1 \quad \text{for } \theta \leq \theta_\ell \qquad (i = 1,2,\ldots,N-1)$$

If $\underline{M}^{(t)}(\theta)$ is the vector whose components are the $M^{(t)}(\theta|i)$ $(i = 0,1,2,\ldots,N-1)$, then

$$\underline{M}^{(t)}(\theta) = P^t(\theta)\underline{\psi} = P^{t-1}(\theta)\, P(\theta)\underline{\psi}$$

$$= P^{t-1}(\theta)\, \underline{M}^{(1)}(\theta)$$

and for values of θ greater than θ_u or less than θ_ℓ, this may be

written

$$\underline{M}^{(t)}(\theta) \geq P^{t-1}(\theta)\underline{\psi}$$

$$= \underline{M}^{(t-1)}(\theta)$$

where one vector is defined as being greater than another if its elements are greater than the corresponding elements in the second. By induction it follows that $M^{(t)}(\theta|i) \to \infty$ as $\theta \to \pm \infty$, so that the lemma is proved.

It is also clear that if m_i is defined by

$$m_i = \left[\frac{d}{d\theta} M(\theta|i)\right]_{\theta=0}$$

and also if

$$m_i^{(t)} = \left[\frac{d}{d\theta} M^{(t)}(\theta|i)\right]_{\theta=0}$$

then $m_i > 0$ $(i = 1,2,\ldots,N-1)$, implies $m_i^{(t)} > 0$ $(i = 1,2,\ldots,N-1)$ and that a similar statement holds, replacing $>$ by $<$. Also, since

$$\frac{d^2}{d\theta^2} M^{(t)}(\theta|i) = E\left\{(k_t - i)^2 \exp \theta(k_t - i)\right\} \geq 0,$$

for all θ, each $M^{(t)}(\theta|i)$ is a convex function. It follows that under the conditions of the lemma, and if the m_i are all positive (or all negative), then for each $M^{(t)}(\theta|i)$ there exists a unique non-zero value of θ, denoted $\theta_i^{(t)}$, for which

$$M^{(t)}(\theta_i^{(t)}|i) = 1.$$

Theorem 2.1

If the conditions of lemma 2.1 hold, and if $\theta_1^{(1)} = \theta_2^{(1)} = \ldots = \theta_{N-1}^{(1)} = \theta^*$ say, then $M^{(t)}(\theta^*|i) = 1$ for all i and for all $t \geq 1$.

Proof

For any $t \geq 1$,

$$\underline{M}^{(t)}(\theta^*) = \begin{pmatrix} 1 \\ M^{(t)}(\theta^*|1) \\ \vdots \\ M^{(t)}(\theta^*|N-1) \\ 1 \end{pmatrix}$$

$$= P^t(\theta^*)\underline{\Psi}$$

$$= P^{t-1}(\theta^*) \begin{pmatrix} 1 \\ M(\theta^*|1) \\ \vdots \\ M(\theta^*|N-1) \\ 1 \end{pmatrix}$$

$$= P^{t-1}(\theta^*)\underline{\Psi}$$

$$= \underline{M}^{(t-1)}(\theta^*)$$

Then since $\underline{M}(\theta^*) = \underline{\Psi}$ it follows by iteration that $M^{(t)}(\theta^*) = \underline{\Psi}$ for all $t \geq 1$, so that Theorem 2.1 is proved.

Theorem 2.2

If the conditions of lemma 2.1 hold, and also if $m_i < 0$ $(i = 1,2,\ldots,N-1)$ and $0 < a \leq \theta_i^{(1)} \leq b$ $(i = 1,2,\ldots,N-1)$ then

$$M^{(t)}(a|i) \leq 1 \leq M^{(t)}(b|i) \qquad (i = 1,2,\ldots,N-1)$$

Proof

Since each $M(\theta|i)$ is convex, then under the conditions of the theorem it follows that

$$M(a|i) \leq 1 \leq M(b|i) \qquad (i = 1,2,\ldots,N-1)$$

We may therefore write

$$\underline{M}(a) \leq \underline{\psi} \leq \underline{M}(b)$$

Thus
$$\begin{aligned} M^{(t)}(a) &= P^t(a)\ \underline{\psi} \\ &= P^{t-1}(a)\ \underline{M}(a) \\ &\leq P^{t-1}(a)\ \underline{\psi} \\ &= \underline{M}^{(t-1)}(a) \end{aligned}$$

Hence by iteration

$$\underline{M}^{(t)}(a) \leq \underline{\psi} \quad \text{for } t \geq 1.$$

It follows by a similar argument that

$$\underline{M}^{(t)}(b) \geq \underline{\psi} \quad \text{for } t \geq 1,$$

so that Theorem 2.2 is established.

Using (2.4) and Theorem 2.1, it follows that for any transition matrix P for which the associated set of moment-generating functions satisfy the conditions of Theorem 2.1,

$$\left\{1-P(N,i)\right\} \exp(-i\theta^*) + P(N,i) \exp(N-i)\theta^* = 1$$

$$(i = 1,2,\ldots,N-1)$$

so that

$$P(N,i) = \frac{\exp(i\theta^*) - 1}{\exp(N\theta^*) - 1} \qquad (i = 1,2,\ldots,N-1)$$

Further, for any transition matrix for which the associated set of moment-generating functions satisfy the conditions of Theorem 2.2, we have

$$\left\{1-P(N,i)\right\} \exp(-ia) + P(N,i) \exp(N-i)a \leq 1$$

$$\leq \left\{1-P(N,i)\right\} \exp(-ib) + P(N,i) \exp(N-i)b$$

$$(i = 1,2,\ldots,N-1)$$

so that

$$\frac{\exp(ib)-1}{\exp(Nb)-1} \leq P(N,i) \leq \frac{\exp(ia)-1}{\exp(Na)-1} \qquad (i = 1,2,\ldots,N-1)$$

Finally, it follows by arguments analogous to those used above that

if $m_i > 0$ $(i = 1,2,\ldots,N-1)$, so that

$$a \leq \theta_i^{(1)} \leq b < 0$$

then the same inequality will hold also. The results of Theorems 2.1 and 2.2 may be used to find exact values or bounds for the probability of survival of A individuals, given that initially the number of such individuals is i $(i = 1,2,\ldots,N-1)$, for various genetic models.

2.3 Applications

Case 1

We consider first the overlapping generation model introduced by Moran (1958a). Here individuals die one by one at random and are replaced by new individuals which are A with probability proportional to the product of the number of A individuals before the birth-death event and the selective advantage of A individuals. If immediately before the birth-death event the number of A individuals is i, it will subsequently be i-1, i, or i+1 with respective probabilities

$$p_{i,i-1} = \mu_2\, i(N-i)\left[N\left\{\mu_1 i + \mu_2(N-i)\right\}\right]^{-1}$$

$$p_{i,i} = 1 - p_{i,i-1} - p_{i,i-1}$$

$$p_{i,i+1} = \mu_1\, i(N-i)\left[N\left\{\mu_1 i + \mu_2(N-i)\right\}\right]^{-1}$$

where μ_1 is the selective advantage of A individuals and μ_2 is that of the a individuals. It follows that

$$M(\theta|i) = e^{-\theta} p_{i,i-1} + p_{i,i} + e^{\theta} p_{i,i+1}$$

Clearly $M(\theta|i)$ satisfies the conditions of lemma 2.1, and also if $\mu_1 \neq \mu_2$, then $m_i \neq 0$. (The case $\mu_1 = \mu_2$ may easily be treated separately.) Thus there exists a unique non-zero solution $\theta_i^{(1)}$ of the equation $M(\theta|i) = 1$, given by

$$\theta_i^{(1)} = \ell n(p_{i,i-1}/p_{i,i+1})$$

$$= \ell n(\mu_2/\mu_1)$$

This solution is independent of i and we may therefore write

$$\theta^* = \ell n\,(\mu_2/\mu_1)$$

Thus the conditions of Theorem 2.1 hold, and it follows immediately that the probability $P(N,k)$ of eventual survival of A individuals, given initially k such individuals, is given by

$$P(N,k) = \frac{(\mu_2/\mu_1)^k - 1}{(\mu_2/\mu_1)^N - 1} \qquad (k = 1,2,\ldots,N-1)$$

This result was established by Moran (1958a) but by a different method than that used above.

Case 2

We consider now the non-overlapping generation model

(c.f. Moran (1960)), where the number of A individuals in any generation is a Markovian variate with transition matrix

$$P = \left\{ p_{ij} \right\} = \left\{ \binom{N}{j} \Pi_i^j \, (1-\Pi_i)^{N-j} \right\}$$

and

$$\Pi_i = \frac{1 - \exp(-2\phi i N^{-1})}{1 - \exp(-2\phi)} \qquad (\phi > 0)$$

In this case

$$M(\theta|i) = e^{-i\theta} \left[\Pi_i \, e^{\theta} + 1 - \Pi_i \right]^N$$

which satisfies the conditions of lemma 2.1 and also the condition $m_i \neq 0$. The solution $\theta_i^{(1)}$ of the equation $M(\theta|i) = 1$ is given by

$$\theta_i^{(1)} = -2\phi$$

which is independent of i, so that we may write

$$\theta^* = -2\phi \ .$$

Hence, using Theorem 2.1,

$$P(N,k) = \frac{\exp(-2\phi k) - 1}{\exp(-2\phi N) - 1}$$

a result which has again been established by Moran (1960), but by using a different method than that used by him in establishing the result of case 1.

Case 3

In the non-overlapping generation model due to Wright (1931) the number of A individuals in any generation is a Markovian variate with transition matrix

$$P = \left\{ p_{ij} \right\} = \left\{ \binom{N}{j} p_i^j \, (1-p_i)^{N-j} \right\}$$

where $p_i = (1+s)i \, (N+si)^{-1}$.

This matrix is obtained by supposing that each individual produces offspring in a Poisson distribution, the parameter for each A parent being $\lambda(1+s)$ and that for a parents being λ, conditioned by total population size in the next generation being N. In this case

$$M(\theta|i) = e^{-i\theta} \, (p_i \, e^{\theta} + 1-p_i)^N \qquad (2.6)$$

These moment-generating functions satisfy the conditions of lemma 1, and since

$$\left[\frac{d}{d\theta} M(\theta|i) \right]_{\theta=0} = si(N-i)(N+si)^{-1}$$

the m_i are either all positive or all negative for non-zero s. The case s = 0 is again easily treated separately. The solutions $\theta_i^{(1)}$ of the equations $M(\theta|i) = 1$, in the case $s \neq 0$, are not identical, so that it is necessary to find bounds a and b for the solutions $\theta_i^{(1)}$ and to apply the result of Theorem 2.2 It is assumed for the moment that $s > 0$; a similar treatment holds

for $s < 0$.

There are several ways of finding bounds (not necessarily as sharp as possible) for the solutions $\theta_i^{(1)}$. A lower bound has been provided (c.f. Ewens and Gani (1961)) by noting that the value of θ for which $M(\theta|i)$ reaches its minimum is

$$\theta = -\ell n(1+s)$$

so that it would be sufficient to put $b = -\ell n(1+s)$. Also, considering the value θ' defined by

$$\exp \theta' = (1-s)(1+s)^{-1}$$

we have

$$M(\theta'|i) = \left[\left(\frac{1+s}{1-s} \right)^x \left(\frac{1-sx}{1+sx} \right) \right]^N$$

where $x = iN^{-1}$. It is readily shown (c.f. Ewens and Gani (1961)) that $M(\theta'|i) > 1$, so that a sufficient value for a is given by

$$a = \ell n \left\{ (1-s)(1+s)^{-1} \right\}$$

Therefore it follows, from Theorem 2.2 that

$$\frac{\left(\frac{1}{1+s} \right)^k - 1}{\left(\frac{1}{1+s} \right)^N - 1} \leq P(N,k) \leq \frac{\left(\frac{1-s}{1+s} \right)^k - 1}{\left(\frac{1-s}{1+s} \right)^N - 1} \qquad (2.7)$$

this result being given incorrectly (with reversed inequalities)

in Ewens and Gani (1961). By using a different method Moran (1960) has found the sharper bounds

$$\frac{\exp(-2sk) - 1}{\exp(-2sN) - 1} \geq P(N,k) \geq \frac{\exp\{-2sk(1+s)^{-1}\} - 1}{\exp\{-2sN(1+s)^{-1}\} - 1}$$

This set of inequalities possesses an interesting property of symmetry which may be used to improve the bounds in (2.7). Suppose that the upper bound

$$\frac{\exp(-2sk) - 1}{\exp(-2sN) - 1} \geq P(N,k) \tag{2.8}$$

has been established, but not the lower bound. Then the lower bound may be found by considering a individuals and deriving an upper bound similar to (2.8), which gives a lower bound for the probability of fixation of A individuals. The fact that the selective advantage of A individuals is 1+s means that when the selective advantage of a individuals is put equal to unity, that of A individuals becomes 1+s. Thus when the selective advantage of A individuals is put equal to unity, that of a individuals becomes not 1-s but $(1+s)^{-1}$, since in the model under consideration, at least, it is in their ratios rather than in their differences that selective advantages operate. Now

$$s = \frac{\text{selective advantage of A individuals}}{\text{selective advantage of a individuals}} - 1$$

so defining

$$s' = \frac{\text{selective advantage of a individuals}}{\text{selective advantage of A individuals}} - 1$$

we get $s' = (1+s)^{-1} - 1 = -s(1+s)^{-1}$.

Then replacing k by N-k and s by s′ in (2.8) it follows that

$$\text{Prob}\left\{ \text{a individuals become fixed} \right\}$$

$$\leq \frac{\exp\left\{ 2s(N-k)(1+s)^{-1} \right\} - 1}{\exp\left\{ 2s\ N(1+s)^{-1} \right\} - 1}$$

$$\text{Thus } P(N,k) \geq 1 - \frac{\exp\left\{ 2s(N-k)(1+s)^{-1} \right\} - 1}{\exp\left\{ 2s\ N(1+s)^{-1} \right\} - 1}$$

$$= \frac{\exp\left\{ -2sk(1+s)^{-1} \right\} - 1}{\exp\left\{ -2sN(1+s)^{-1} \right\} - 1}$$

which is the lower bound given by Moran. Therefore there is no reason to state that one or other of these bounds is sharper than the other, and in particular it is not true that the upper bound, which turns out to be the diffusion approximation for $P(N,k)$ is any better than the lower bound.

An analogous symmetry relation may be used to improve the bounds (2.7). Clearly the lower bound may be improved by considering an upper bound for the survival probability of a individuals. Replacing k by N-k and s by $-s(1+s)^{-1}$ in the upper bound in (2.7) we get

Prob (a individuals become fixed)

$$\leq \frac{\left\{\frac{1+s(1+s)^{-1}}{1+s(1+s)^{-1}}\right\}^{N-k} - 1}{\left\{\frac{1+s(1+s)^{-1}}{1-s(1+s)^{-1}}\right\}^{N} - 1}$$

$$= \frac{(1+2s)^{N-k} - 1}{(1+2s)^{N} - 1}$$

Thus $P(N,k) \geq 1 - \frac{(1+2s)^{N-k} - 1}{(1+2s)^{N} - 1}$

$$= \frac{\left(\frac{1}{1+2s}\right)^{k} - 1}{\left(\frac{1}{1+2s}\right)^{N} - 1} \qquad (2.9)$$

If the lower bound in (2.7) is replaced by (2.9), the two bounds then provided differ from those of Moran by terms of order s^3, and consequently the two sets of bounds are extremely close for small s. Also, the upper and lower bounds in each set will be close for small s, and the arithmetic mean of the bounds should give a very close approximation to the true probability.

Case 4

The methods outlined above may be used to derive exact values or bounds for survival probabilities in infinite populations subject to immigration. If we denote by $f(\theta)$

the moment-generation function of the offspring distribution of any individual, then in the case of no immigration the equation $M(\theta|i) = 1$ reduces to

$$e^{-i\theta} f^i(\theta) = 1$$

$$\text{or } f(\theta) = e^{\theta}$$

the well-known branching process equation. In the case where immigration is present a more complicated equation is derived. Details are given in section 5 in Ewens and Gani (1961); since this section is entirely due to Gani, further discussion is omitted here.

Case 5

As mentioned in Chapter 1, it is possible in some cases to derive a Markovian variate for processes for which the entire population is divided into subpopulations, with different selective advantages of A individuals in each, provided that a sufficient amount of migration takes place. As the simplest possible example we consider the case where the total population, of size Nn, is divided into n distinct subpopulations each of size N. It is supposed that the selective advantage of A individuals in the i^{th} subpopulation is $s_i (i = 1,2,\ldots,n)$ and that the proportion of individuals in the i^{th} subpopulation that are A at any given time is x_i. Now suppose that a (large) intermigration takes place between subpopulations, having the effect that the Nx_i A individuals in the i^{th} subpopulation are equally distributed

among the n subpopulations and that the same happens for the $N(1-x_i)$ a individuals. A new generation is now formed within each subpopulation as described in case 3 above, and then a new migration takes place, and so on. We consider the population immediately after the birth of two consecutive generations, before migration occurs. The original number of A individuals is $N(x_1 + \ldots + x_n)$. For the next generation the number of A individuals in the i^{th} subpopulation has moment-generating function

$$\left[p_i e^{\theta} + 1-p_i \right]^N$$

where

$$p_i = (x_1 + \ldots + x_n)(1+s_i) \left[n+s_i(x_1 + \ldots + x_n) \right]^{-1}$$

If $z = (x_1 + \ldots + x_n)n^{-1}$ then p_i becomes $z(1+s_i)(1+zs_i)^{-1}$ and the M.G.F. of the increase in the number of A individuals in the entire population from one generation to the next is

$$\exp(-Nnz\theta) \prod_{i=1}^{n} \left[\frac{z(1+s_i)e^{\theta}}{1+s_i z} + \frac{1-z}{1+s_i z} \right]^N$$

$= M(\theta|z)$ say, from which it is clear that z is a Markovian variate. From the fact that

$$M'(0|z) = Nz(1-z) \sum_{i=1}^{n} s_i(1+zs_i)^{-1}$$

it follows that if all the s_i are negative (or all positive)

and are less than unity in absolute value, then $M'(0|z)$ is always negative (or always positive) for all z. Since also the conditions of lemma 2.1 are satisfied bounds can be found for the probability of eventual survival of A individuals by using Theorem 2.2. Clearly bounds may be obtained quickly by using the maximum and minimum s_i and the methods of case 3; however if the s_i differ to any great extent it should be possible to find sharper bounds by using $M(\theta|z)$directly.

CHAPTER 3.

DIFFUSION METHODS

3.1 Introduction

In many cases in genetics it is difficult to find exact formulae for quantities of interest. On the other hand it is often possible and simple to use diffusion methods to derive approximations for these quantities, provided that a single Markovian variate exists in terms of which the population behaviour may be described. For convenience we shall for the moment speak loosely of the value of a variate at the t^{th} generation, or perhaps at the t^{th} birth-death event, as the value at time t. This implies that changes in the population structure can only take place at times 1,2,3,... .

Application of diffusion methods rests on the supposition that there exists a Markovian variate x whose value x_t at time t is such that $E(x_{t+1}-x_t|x_t)$ and $E\left\{(x_{t+1}-x_t)^2|x_t\right\}$ are both $O(N^{-\alpha})$, where N is the population size and α a positive constant, and that higher moments are $o(N^{-\alpha})$. It is sufficient for our purposes to assume $0 \leq x \leq 1$ and that $x = 0$ and $x = 1$ are absorbing barriers, so that once x reaches 0 or 1 it remains fixed. This corresponds genetically to the fixation of an allele in a population where mutation is absent. We sketch the derivation of the forward and backward Kolmogorov equations for comparison with methods used later. To do this the time axis must be rescaled so that unit time corresponds to N^{α} of the previous time units.

We let $\phi(x;t)$ be the density function of x at time t, and $\psi(u;x)$ be the probability that x next changes to x+u, so that

$\psi(u;x)$ is the distribution of the jump u, given x. Then as a particular case of the Chapman-Kolmogorov equation we have

$$\phi(x;t+h) = \int \phi(x-u;t)\ \psi(u;x-u)\ du$$

(c.f. Moran (1962) p.75). By expanding both sides in Taylor series and taking the leading terms, we obtain eventually

$$\frac{\partial \phi(x;t)}{\partial t} = -\frac{\partial}{\partial x}\left\{ m(x)\ \phi(x;t) \right\} + \tfrac{1}{2}\frac{\partial^2}{\partial x^2}\left\{ v(x)\ \phi(x;t) \right\} \qquad (3.1)$$

as the (forward) Kolmogorov (or Fokker-Planck) equation asymptotically satisfied by the distribution of x at time t. Here $m(x)$ and $v(x)$ are the first two moments of the distribution of the jump u, given x.

The backward equation may be derived similarly (c.f. Barucha - Reid (1960) p.130), by considering a Chapman-Kolmogorov equation similar to the above, except that we consider a small increment of time immediately after the process starts. Thus if initially x = p, and if

$$\Theta\ (x;t) = \int_{0-}^{x} \phi(y;t)\ dy$$

the backward equation becomes

$$\frac{\partial}{\partial t}\left\{\Theta(x;t|p)\right\} = m(p)\ \frac{\partial}{\partial p}\left\{\Theta(x;t|p)\right\} + \frac{v(p)}{2}\frac{\partial^2}{\partial p^2}\left\{\Theta(x;t|p)\right\} \qquad (3.2)$$

This equation will usually be more useful to us than the equivalent equation

$$\frac{\partial}{\partial t}\left\{\phi(x;t|p)\right\} = m(p)\,\frac{\partial}{\partial p}\left\{\phi(x;t|p)\right\} + \frac{v(p)}{2}\,\frac{\partial^2}{\partial p^2}\left\{\phi(x;t|p)\right\} \qquad (3.3)$$

which is adjoint to (3.1).

In particular it follows from (3.2), by letting x take in turn the values 0_+ and 1_-, that the probability $G(t;p)$ that the process is absorbed at 0 (at 1) before time t, given that initially $x = p$, is the solution of the equation

$$\frac{\partial}{\partial t}\,G(t;p) = m(p)\,\frac{\partial}{\partial p}\,G(t;p) + \frac{v(p)}{2}\,\frac{\partial^2}{\partial p^2}\,G(t;p) \qquad (3.4)$$

subject to appropriate boundary conditions in each case. By putting the left-hand in (3.4) equal to zero and replacing $G(t;p)$ by $G(p)$, the probability that eventually $x = 0$ may be found when appropriate boundary conditions are imposed.

It also follows from (3.4) after some manipulation (c.f. Feller (1954)) that the mean time $U(p)$ until one or other boundary is reached, given that initially $x = p$, is the solution of the equation

$$-1 = m(p)\,\frac{d\,U(p)}{dp} + \frac{v(p)}{2}\,\frac{d^2U(p)}{dp^2}\,, \qquad (3.5)$$

again with appropriate boundary conditions, and that the mean value of $S(p)$ of the square of the time taken is the solution of

$$-2U(p) = m(p)\frac{dS(p)}{dp} + \frac{v(p)}{2}\frac{d^2S(p)}{dp^2}, \qquad (3.6)$$

again subject to the obvious boundary conditions.

3.2 Pseudo-transient distributions and functions

The concept of a "pseudo-transient" distribution was developed (c.f. Ewens (1963a)) to give a meaning to a function derived from formal operations on equation (3.1). The interpretation of this formal solution is found later in this chapter, and the function so derived will be useful in subsequent chapters.

We consider for the moment equation (3.1) in the particular case where $0 \leq x \leq 1$ and

$$\left.\begin{aligned} m(x) &= x(1-x)\,\xi(x) \\ v(x) &= ax(1-x) \end{aligned}\right\} \qquad (3.7)$$

where $\xi(x)$ is an arbitrary polynomial which is $O(1)$ (possibly constant or zero) and a is a constant, also $O(1)$. The drift and diffusion coefficients in genetical applications are usually of this form, and to be definite it is supposed from now on, unless otherwise stated, that $m(x)$ and $v(x)$ are of the form (3.7). This implies, in the terminology of Feller, (1952, 1954) that $x = 0$, $x = 1$ are "exit" boundaries; in other words the probability is unity that one or other boundary is reached in finite time, and once a boundary is reached the variate x remains fixed at that boundary.

It is supposed that initially $x = p$. Then the probability $P_0(p)$ that the boundary $x = 0$ is reached before the boundary $x = 1$ is given by the solution of

$$m(p)\frac{dP_0(p)}{dp} + \frac{v(p)}{2}\frac{d^2P_0(p)}{dp^2} = 0 \qquad (3.8)$$

which is (3.4) with the left-hand side put equal to zero, subject to the boundary conditions $P_0(0) = 1$, $P_0(1) = 0$. If $P_1(p)$ is the probability that the boundary $x = 1$ is reached before the boundary $x = 0$, then using (3.8),

$$P_0(p) = 1 - P_1(p) = \frac{\int_p^1 \psi(x)\,dx}{\int_0^1 \psi(x)\,dx} \qquad (3.9)$$

where

$$\psi(x) = \exp\left[-2\int^x m(y)/v(y)\ dy \right] \qquad (3.10)$$

It has been shown by Watterson (1962) that the rate of flux of probability into the "exit" $x = 0$ is given by

$$\frac{dP_0(p;t)}{dt} = \lim_{x\to 0+}\left[\tfrac{1}{2}\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t) \right] \qquad (3.11)$$

where $P_0(p;t)$ is the probability that the process has been absorbed at $x = 0$ by time t. Similarly

$$\frac{dP_1(p;t)}{dt} = -\lim_{x\to 1-}\left[\tfrac{1}{2}\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t) \right] \qquad (3.12)$$

is the rate of flux of probability into the "exit" x = 1.

The derivation of (3.11) and (3.12) is suggested by writing (3.1) formally as

$$\frac{\partial F}{\partial t}(x;t) = \tfrac{1}{2}\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t) \tag{3.13}$$

We also note that the solution of (3.5) subject to the boundary conditions $U(0) = U(1) = 0$, is

$$U(p) = \int_0^p f_0(x)\,dx + \int_p^1 f_1(x)\,dx \tag{3.14}$$

where

$$f_0(x) = \frac{2P_0(p)}{v(x)}\left[\psi(x)\right]^{-1}\int_0^x \psi(y)\,dy$$

and $$f_1(x) = \frac{2P_1(p)}{v(x)}\left[\psi(x)\right]^{-1}\int_x^1 \psi(y)\,dy$$

One of the purposes of deriving the pseudo-transient distribution is to provide a meaning for the two components on the right-hand side in(3.14).

The formal operation used to derive the pseudo-transient distribution is to solve the "stationary" equation

$$\tfrac{1}{2}\frac{d^2}{dx^2}\left\{v(x)f(x)\right\} - \frac{d}{dx}\left\{m(x)f(x)\right\} = 0 \tag{3.15}$$

obtained by equating the left-hand side in (3.1) to zero and by replacing $f(x;t)$ by $f(x)$. In the case under consideration such a

formal solution can be obtained easily but some controversy (see the discussion in Watterson (1962)) has been attached to its interpretation. This follows from the fact that the only true stationary distribution of the process is the trivial one where x takes the value 0 with probability $P_0(p)$ and the value 1 with probability $P_1(p)$, while the formal solution gives a distribution on the interval (0,1).

By proceeding formally we obtain from (3.15) successively

$$\tfrac{1}{2}\frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) = C_1 \qquad (3.16)$$

and

$$f(x) = \frac{C_2\left[\psi(x)\right]^{-1}}{v(x)} + \frac{2C_1\left[\psi(x)\right]^{-1}}{v(x)}\int^x \psi(y)\,dy \qquad (3.17)$$

which may be written alternatively (if $C_1 \neq 0$)

$$f(x) = \frac{2\,C_1\left[\psi(x)\right]^{-1}}{v(x)}\int_c^x \psi(y)\,dy \qquad (3.18)$$

where the constant C is related to C_2. Clearly the formal solution (3.18) is not a stationary distribution of the diffusion process, since with the choice (3.7) of m(x) and v(x) no non-trivial stationary distribution can exist, and it remains to find a meaningful interpretation of (3.18) when the constants are allocated suitably.

To find such an interpretation it is useful to consider an associated "return" process, which is the same as the process considered above (i.e. the "non-return" process) except that once

one of the boundaries x = 0 or x = 1 is reached the process is immediately restarted with x = p, the original value. For the new process probability does not accumulate at x = 0 or x = 1 and a non-trivial stationary distribution will exist. It will be by finding this stationary distribution that an interpretation for (3.18) will be found in the non-return process. Return processes of the type discussed above have been considered by Feller (1954), who has given equations analogous to (3.1) and (3.2) satisfied by f(x;t) in the more general case where whenever a boundary is reached the process is restarted at a point y with distribution $h_0(y)$ or $h_1(y)$, depending on the boundary just attained. In our case these are point distributions at y = p. Feller shows that the backward equation (3.2) continues to hold, but that the forward equation must be replaced, in our case, by

$$\frac{\partial}{\partial t}\int_{\Omega} f(x;t)\,dx = \int_{\Omega} \frac{\partial}{\partial x}\left[\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t)\right]dx \qquad (3.19)$$

if $p \notin \Omega$ (Ω any interval in (0,1))

and

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Omega} f(x;t)\,dx = {} & \int_{\Omega} \frac{\partial}{\partial x}\left[\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t)\right]dx \\ & + \lim_{x\to 0}\left[\tfrac{1}{2}\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t)\right] \\ & - \lim_{x\to 1}\left[\tfrac{1}{2}\frac{\partial}{\partial x}\left\{v(x)f(x;t)\right\} - m(x)f(x;t)\right] \qquad (3.20)\end{aligned}$$

if $p \in \Omega$.

The interpretation to be attached to (3.20) is that the rate of change of probability mass in any interval covering p is equal to the net rate of change of probability mass due to flux through the interval, plus the rate of increase in probability allowing for the possibility that one or other boundary is reached and the process restarted at p. If the interval Ω does not cover p the latter terms must be eliminated, so that (3.19) is obtained for this case.

The return process admits a stationary distribution f(x) which satisfies, using (3.19)

$$\int_{\Omega} \frac{d}{dx}\left[\tfrac{1}{2} \frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) \right] = 0 \qquad (3.21)$$

where $p \notin \Omega$.

Now suppose that Ω is any interval of the form $(0,\ell)$, where $\ell < p$. Since ℓ may be chosen arbitrarily in $(0,p)$ we must have

$$\tfrac{1}{2} \frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) = D_1 \qquad (3.22)$$

for $(0 < x < p)$, where D_1 is a suitable constant. Similarly, by considering intervals of the type $(k,1)$, where $k > p$, we obtain

$$\tfrac{1}{2} \frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) = -D_2 \qquad (3.23)$$

for $(p < x < 1)$, where D_2 is another suitable constant.

The constants D_1 and D_2 will later be identified with flux rates (c.f. equations (3.11) and (3.12)). Using (3.22) and (3.23) we obtain

$$f(x) = \frac{2D_1}{v(x)} \left[\psi(x) \right]^{-1} \int_A^x \psi(y)\, dy \qquad 0 < x < p$$

$$f(x) = \frac{2D_2}{v(x)} \left[\psi(x) \right]^{-1} \int_x^B \psi(y)\, dy \qquad p < x < 1 \tag{3.24}$$

Equation (3.24) specifies the stationary distribution of the return process, and it remains to allocate the four constants A, B, D_1 and D_2. The constants A and B are readily evaluated by noting that since f(x) is a density function, it is L - integrable on (0,p) and (p,1). Now with the choice (3.7) of m(x) and v(x),

$$\psi(x) = \exp \left[\text{polynomial in } x \right]$$

so that since $v(x) = ax(1-x)$, it is necessary to put A = 0, B = 1 to satisfy the integrability condition. This leaves us with

$$f(x) = \frac{2D_1}{v(x)} \left[\psi(x) \right]^{-1} \int_0^x \psi(y)\, dy \qquad 0 < x < p$$

$$f(x) = \frac{2D_2}{v(x)} \left[\psi(x) \right]^{-1} \int_x^1 \psi(y)\, dy \qquad p < x < 1 \tag{3.25}$$

involving now only the two constants D_1 and D_2. One relation between D_1 and D_2 may be obtained immediately from the fact that f(x) is a density function, so that its integral over (0,1) is unity. A second and independent relation is obtained by noting

from (3.25) that

$$\frac{\lim_{x\to 0} f(x)}{\lim_{x\to 1} f(x)} = \frac{D_1}{D_2}$$

This implies that whenever x is in one or other of the two intervals $(0,\varepsilon)$ and $(1-\varepsilon,1)$, (ε arbitrarily small), the probability that it is in $(0,\varepsilon)$ is asymptotically

$$\frac{D_1}{D_1+D_2} + O(\varepsilon) \quad .$$

Now $P_0(p)$ is differentiable at $p = 0$, so that given that x is in $(0,\varepsilon)$, the probability that the process next enters $x = 0$ rather than $x = 1$ can be made as close to unity as desired, by letting $\varepsilon\to 0$. This implies that

$$\frac{D_1}{D_1+D_2} = P_0(p)$$

which provides a second relation between D_1 and D_2. Solving the two simultaneous equations for D_1 and D_2, it is found that

$$\left.\begin{aligned} D_1 &= P_0(p)/U(p) \\ D_2 &= P_1(p)/U(p) \end{aligned}\right\} \qquad (3.26)$$

It may be noted that this implies, using (3.11) and (3.12), that the rate of flux of probability into $x = 0$ in the return process is asymptotically $P_0(p)/U(p)$, which might have been anticipated. An analogous result holds for $x = 1$. Using (3.25) and (3.26) we obtain finally for the stationary distribution of the return process

$$f(x) = \frac{2P_0(p)\left[\psi(x)\right]^{-1}}{U(p)\ v(x)} \int_0^x \psi(y)\ dy \qquad 0 < x < p$$

$$f(x) = \frac{2P_1(p)\left[\psi(x)\right]^{-1}}{U(p)\ v(x)} \int_x^1 \psi(y)\ dy \qquad p < x < 1 \tag{3.27}$$

By using equation (3.9) it is immediately verified that $f(x)$ is continuous at $x = p$, so that an alternative way of deriving the second relation between D_1 and D_2 would have been to prove $f(x)$ continuous. However $f(x)$ is not differentiable at $x = p$, but is differentiable elsewhere.

Our main interest in the return process stationary distribution is for the information it provides about the original process. It follows at once that the interpretation of $f(x)$ in the original non-return process is that

$$U(p) \int_{x_1}^{x_2} f(x)\ dx \qquad 0 \leq x_1 < x_2 \leq 1 \tag{3.28}$$

is the mean time the process spends in the range (x_1, x_2) before absorption at 0 or 1. It may be noted by comparing (3.27) with (3.18), that the formal solution (3.18) of the "stationary" equation (3.15) will therefore admit this interpretation if the constants are allocated suitably.

Thus so far as the original process is concerned, (3.27) is not strictly the density function of any variate; however since it describes in a sense the behaviour of the process before

absorption it is named the "pseudo-transient distribution" of x. Similarly U(p)f(x) will be called the "pseudo-transient function".

In the case $m(x) \equiv 0$, $v(x) = x(1-x)$, it follows from (3.27) that

$$f(x) = \frac{2(1-p)}{U(p)(1-x)} \qquad 0 < x < p$$

$$f(x) = \frac{2p}{U(p)x} \qquad p < x < 1 \tag{3.29}$$

where U(p) is given in this case by

$$U(p) = -2\left[p\ell np + (1-p)\,\ell n(1-p)\right] .$$

It follows immediately that $-2p\ell np$ is the mean time the process spends in (0,p) before absorption and that $-2(1-p)\ell n(1-p)$ is the mean time spent in (p,1), thus providing meanings for the two components on the right-hand side of (3.14). Also, it may be noted that although

$$U(p) = P_0(p)\int_0^p \frac{2}{v(x)}\left[\psi(x)\right]^{-1}\int_0^x \psi(y)\,dy\,dx$$

$$+ P_1(p)\int_p^1 \frac{2}{v(x)}\left[\psi(x)\right]^{-1}\int_x^1 \psi(y)\,dy\,dx \tag{3.30}$$

it is not necessarily true that

$$\int_0^p \frac{2}{v(x)}\left[\psi(x)\right]^{-1}\int_0^x \psi(y)\,dy\,dx$$

is the mean time before absorption conditional on absorption at x = 0.

In fact in the next chapter a discrete process is discussed for which one may write equally well

$$U(p) = P_0(p)\ \xi_1(p) + P_1(p)\ \theta_1(p)$$

$$\text{or} \quad U(p) = P_0(p)\ \xi_2(p) + P_1(p)\ \theta_2(p) \tag{3.31}$$

where $\xi_1(p) \neq \xi_2(p)$ and $\theta_1(p) \neq \theta_2(p)$. Thus $\xi_1(p)$, say, is not necessarily the mean absorption time given absorption at $x = 0$, and the reason for the possibility of the alternative forms in (3.31) is that a functional relation exists between $P_0(p)$, $\xi_1(p)$, and $\xi_2(p)$.

3.3 Comments on the Pseudo-transient distribution

Before discussing other interpretations of formal solutions of (3.15), two comments may be made about the pseudo-transient distribution (3.27). Firstly, it is not the same as the limiting ($t \to \infty$) distribution of x, given x is in the open interval (0, 1). In fact it has been shown by Kimura (1955a) that in the case $m(x) = 0$, $v(x) = x(1-x)$, the solution of the equation

$$\frac{\partial \phi(x;t)}{\partial t} = \tfrac{1}{2} \frac{\partial^2}{\partial x^2} \left\{ v(x)\ \phi(x;t) \right\}$$

is

$$\phi(x;t) = \sum_{i=1}^{\infty} \frac{4(2i+1)p(1-p)}{i(i+1)}\ T^1_{i-1}(1-2p) T^1_{i-1}(1-2x) \exp\left\{ -\tfrac{1}{2} i(i+1)t \right\}$$

where $T^1_{i-1}(z)$ is a Gegenbauer polynomial which is defined in terms of the hypergeometric function. The leading term in this

expression for large t is

$$6p(1-p)\ \exp(-t)$$

which is independent of x, so that the limiting distribution of x, conditional on no absorption, is rectangular, and not of the form (3.29).

It should also be noted that (3.27) is not the same as the limiting distribution of x, given x is in (0, 1), in the case where mutation in both directions is allowed, as the mutation rates tend to zero. The asymptotic conditional distribution in the case $m(x) = 0$, $v(x) = x(1-x)$, is

$$f(x) = \frac{\text{const}}{x(1-x)} \qquad \frac{1}{N} \leqq x \leqq \frac{N-1}{N} \tag{3.32}$$

where N is the population size. This again is not of the form (3.29) although it may be noted that (3.32) is formally a solution of (3.15). This asymptotic conditional distribution has been discussed by Moran (1962), p.129, and will be returned to in the next section.

The second comment about (3.27) is that it could be used for the purposes of making inferences about the initial value of x. A population under observation may well have been in existence for a long but unknown time before observation started, so that although it may be possible to make good estimates of the selective advantages, and to observe, say, that mutation is absent in the population, it is impossible to use the conditional distribution of x, $(0 < x < 1)$, for any initial value p and time to to make

inferences about p, since the value of t will be unknown. In such cases it seems reasonable to use the pseudo-transient distribution to estimate p. This could be done with one or several observations. If it is known that $m(x) = 0$, $v(x) = x(1-x)$, then in the case of one observation x the maximum-likelihood estimator $\hat{p}$ of p is simply x. When $m(x) \neq 0$ it may no longer be true that $\hat{p} = x$, but here it would be more difficult to find $\hat{p}$, due to the complexity of the pseudo-transient distribution and to the fact that it takes different functional forms in different ranges.

3.4 Other Interpretations

We now turn to other formal solutions of (3.15). Here it is useful to divide these solutions into two groups; those which put $C_1 = 0$ in (3.16) and(3.17) and those which allow C_1 to remain arbitrary. The left-hand side in (3.16) has the interpretation of a probability flux (c.f. equation (3.13)), so that the restriction $C_1 = 0$ implies no asymptotic probability flux and may only be applied to those solutions of (3.15) which admit a true stationary distribution, (in genetical terms, those cases where a two-way mutation exists). Conversely if a true stationary distribution does exist, then C_1 must be put equal to zero, and by doing this we obtain (c.f. equation (3.17))

$$f(x) = \frac{\text{const}}{v(x)} \exp\left[2\int^{x} m(y)/v(y)\ dy\right]$$

which is Wright's well-known equation for stationary distributions.

It is therefore of prime importance to recognise, when formally solving (3.15), whether the solution obtained is a true stationary distribution or a pseudo-transient distribution since the allocation of C_1 will depend on which of the two distributions is relevant. This point has frequently been overlooked and incorrect results obtained by an incorrect allocation of C_1. It is pertinent to examine the discussion in Watterson (1962), section 6, on this point. Here equation (3.17) has been obtained for a particular value of $m(x)$, (Watterson's equation (6.2)). In the case under discussion mutation is absent so that no true stationary distribution can exist, and by a correct allocation of constants the pseudo-transient distribution can be derived from Watterson's equation (6.2). However Watterson is not here interested in pseudo-transient distributions and notes only that if the constants in his equation (6.2) are non-zero and do not take different values in different ranges, then his $f(x)$ is not integrable on $(0, 1)$ so that the true stationary distribution is purely discrete at $x = 0$ and $x = 1$.

We discuss three other interpretations of formal solutions of (3.15). Feller (1951) in an earlier investigation obtained the solution

$$f(x) = \frac{\text{const}}{x(1-x)} \tag{3.33}$$

in the case $m(x) = 0$, and stated that this solution is meaningless. However (3.33) has been obtained by putting $C_1 = 0$, which is not allowable, since no non-trivial stationary distribution exists for this choice of $m(x)$. The more general ($C_1 \neq 0$) solution is

$$f(x) = \frac{A}{x} + \frac{B}{1-x} \qquad (3.34)$$

and this does admit a meaningful interpretation when A and B are suitably allocated, as has been shown above (equation 3.29).

A second interpretation has been given by Kolmogorov (1959), who obtains (3.34) as the formal solution of (3.15) when $m(x) = 0$. Here Kolmogorov does not allow A and B to take different values in $(0,p)$ and $(p, 1)$ so that his interpretation of (3.34) cannot be the same as the interpretation of (3.29). The abstract in which Kolmogorov's interpretation of (3.34) is given is extremely concise and very difficult to follow; however if his interpretation is similar to that given below by Moran, then A and B in equation (3.34) should be put equal, since Moran's interpretation considers the possibility of mutation, so that a true stationary distribution does exist and must be of the form (3.33) rather than (3.34).

The third interpretation of formal solutions of (3.15) has been given by Moran (1962), p.129. If we consider the distribution of the frequency of a given gene in a population for which mutation without selection is allowed, a stationary distribution of the form

$$f(x) = \text{const } x^{\beta_1 - 1} (1-x)^{\beta_2 - 1}$$

where β_1 and β_2 are related to the mutation rates, is obtained. If β_1 and β_2 approach zero, this distribution will concentrate in the tail-ends of the interval $[0,1]$, indicating that for very small mutation rates there is a large probability that at any given time

one or other gene is not present in the population. The distribution

$$f(x) = \frac{\text{const}}{x(1-x)} \qquad \left(\frac{1}{2N} \leq x \leq \frac{2N-1}{2N}\right)$$

(where 2N is the population size if the population is haploid and N is the population size if the population is diploid) is then a limiting distribution approached, though never attained, by the distribution of the frequency of a given gene, conditional on $x \neq 0$ or 1, for arbitrarily small mutation rates. This interpretation satisfies the criterion discussed above, that since a true stationary distribution does exist (since mutation is allowed), the solution has been derived by putting $C_1 = 0$ in (3.16) and (3.17).

3.5 Mutation in one direction

It is possible to derive a meaningful pseudo-transient distribution in the case where equation (3.7) no longer holds. As an example which will be useful subsequently we consider the situation where the drift and diffusion coefficients are given by

$$\left.\begin{aligned} m(x) &= -cx \\ v(x) &= ax(1-x) \end{aligned}\right\} \qquad (3.35)$$

These correspond genetically to the case where if x is the proportio of the gene (say A) under consideration, and is Markovian, then mutation takes place (at rate c) from A to a but not in the reverse

direction. The constant c is supposed of the same order as a. With the coefficients (3.35) no selection occurs; however selection could be allowed by replacing m(x) by

$$m(x) = sx(1-x) - cx \tag{3.36}$$

This does not essentially alter the type of results obtained and for clarity the coefficients (3.35) are retained.

With the above choice of coefficients the process will eventually become "absorbed" with x = 0, and we wish to determine the mean absorption time, given that initially x = p. This may be done by modifying slightly the methods of the previous sections. The pseudo-transient distribution f(x) will satisfy equation (3.15) and hence

$$\tfrac{1}{2}\frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) = \text{const.} \tag{3.37}$$

The allocation of the constant in (3.37) will depend on the value of x, since the left-hand side has the interpretation of a probability flux and consequently its value depends on whether $x > p$ or $x < p$. Using the arguments of the previous sections it follows that the constant must be put equal to zero for $p < x < 1$ and equal to $[U(p)]^{-1}$ for $0 < x < p$, where $U(p)$ is the mean time for absorption at x = 0.

Considering first the case $0<x<p$ it follows, using (3.35) and (3.37), that f(x) satisfies

$$\frac{d}{dx}\left\{x(1-x)f(x)\right\} + \frac{2cx}{a}f(x) = \frac{2}{aU(p)}$$

$$\text{or} \quad \frac{df(x)}{dx} + \left[\frac{1-2x+2ca^{-1}x}{x(1-x)} \right] f(x) = \frac{2}{ax(1-x)\ U(p)}$$

This equation is solved easily by introducing the integrating factor $x(1-x)^{1-2ca^{-1}} = x(1-x)^{d}$ say. The solution is

$$f(x) = -2a^{-1}x^{-1}\left[U(p)\right]^{-1} d^{-1} + k\, x^{-1}(1-x)^{-d} \tag{3.38}$$

where k is a suitable constant, which may be derived from the requirement that $f(x)$ be integrable near $x = 0$. From this it follows that $f(x)$ is given by

$$f(x) = 2a^{-1}\, x^{-1}\left[U(p)\right]^{-1} d^{-1}\left[(1-x)^{-d} - 1\right] \tag{3.39}$$

$$\text{for } 0 < x < p.$$

For $p < x < 1$ the constant in (3.37) must be put equal to zero, from which it follows eventually that

$$f(x) = \text{const}\ \ x^{-1}(1-x)^{-d} \qquad (p < x < 1) \tag{3.40}$$

The constant in (3.40) may be found by noting that the integral of $f(x)$ over $(0,1)$ must be unity, but it is simpler to assume the continuity of $f(x)$ at $x = p$, thus fixing the constant, and then show that with this choice the integral of $f(x)$ over $(0,1)$ is in fact unity. By doing this we obtain the expression

$$2a^{-1}\left[U(p)\right]^{-1} d^{-1}\left[1-(1-p)^{d}\right]$$

for the constant in (3.40), so that the pseudo-transient distribution is given by

$$f(x) = 2a^{-1}\left[U(p)\right]^{-1} d^{-1}x^{-1}\left[(1-x)^{-d} - 1\right] \qquad 0 < x < p \qquad (3.40)$$

$$f(x) = 2a^{-1}\left[U(p)\right]^{-1} d^{-1}x^{-1}\left[1-(1-p)^{d}\right]\left[(1-x)^{-d}\right] \qquad p < x < 1 \qquad (3.41)$$

and the pseudo-transient function is simply $U(p)f(x)$. (3.42)

It follows that

$$U(p) = 2a^{-1}d^{-1}\int_0^p x^{-1}\left[(1-x)^{-d} - 1\right] dx$$

$$+ 2a^{-1}d^{-1}\left[1-(1-p)^{d}\right]\int_p^1 (1-x)^{-d}\, dx \,. \qquad (3.43)$$

All the above has assumed that $2ca^{-1} - 1 \neq 0$. However in this particular case the above expressions simplify and proceeding from (3.37) we find eventually

$$f(x) = -2a^{-1}\left[U(p)\right]^{-1} x^{-1}\ell n(1-x) \qquad 0 < x < p$$

$$f(x) = -2a^{-1}\left[U(p)\right]^{-1} x^{-1}\ell n(1-p) \qquad p < x < 1 \qquad (3.44)$$

and
$$U(p) = -2a^{-1}\int_0^p x^{-1}\ell n(1-x)\, dx$$

$$-2a^{-1}\ell n(1-p)\int_p^1 x^{-1}\, dx \qquad (3.45)$$

3.6 Almost-invariant Functions

The second concept with which this chapter is concerned is that of an "almost-invariant" function. By using this concept it

will be possible to obtain not only approximations (which will usually be identical to diffusion approximations) of quantities of genetical interest, but also strict bounds for the true values being approximated.

The methods used in deriving these functions are similar to those used at the beginning of this chapter. It is supposed that there exists a Markovian variate x_t $(t=0,1,2,\ldots)$ $(0 \leq x_t \leq 1)$ for which

$$m(x_t) \equiv E(x_{t+1}-x_t) \text{ is } O(N^{-\alpha})$$

$$v(x_t) \equiv E(x_{t+1}-x_t)^2 \text{ is also } O(N^{-\alpha})$$

but higher moments are $o(N^{-\alpha})$. Here N is the population size and α is a positive constant. Consider firstly the case of absorption probabilities. Suppose that there exists a function $\phi(x)$ satisfying

$$E\left[\ \phi(x_{t+1}) - \phi(x_t) \,|\, x_t\ \right] = 0 \qquad (3.46)$$

Note that finding a function $\phi(x)$ satisfying (3.46) is equivalent to finding some function (which is itself a random variable) of the random variable x_{t+1} which possesses the martingale property, (c.f. Doob, 1953). Thus equations such as (3.49) below could be derived strictly by using theorems on expectations of martingales with optional stopping.

By expanding $\phi(x_{t+1})$ in a Taylor series about x_t and ignoring terms which are $o(N^{-\alpha})$ we obtain

$$m(x_t)\,\phi'(x_t) + \tfrac{1}{2}v(x_t)\,\phi''(x_t) = 0 \qquad (3.47)$$

This differential equation is to be true for all x_t, $(0 < x_t < 1)$, and it may be solved readily for $\phi(x_t)$. The solution is

$$\phi(x) = A + B\int^{x} \exp\left[-2\int^{y} m(z)/v(z)\,dz\right] dy \qquad (3.48)$$

where A and B are arbitrary constants. By iteration in (3.46) it follows if the initial value of x is p, that since x_t is eventually either 0 (with probability $P_0(p)$ or 1 (with probability $P_1(p)$), that

$$P_0(p)\,\phi(0) + P_1(p)\,\phi(1) - \phi(p) = 0 \qquad (3.49)$$

or
$$P_1(p) = 1 - P_0(p) = \frac{\phi(p) - \phi(0)}{\phi(1) - \phi(0)}$$

$$= \frac{\displaystyle\int_0^{p} \exp\left[-2\int^{y} m(z)/v(z)\,d\,z\right] dy}{\displaystyle\int_0^{1} \exp\left[-2\int^{y} m(z)/v(z)\,dz\right]} \qquad (3.50)$$

It is clear that this solution is identical to that which is obtained by solving (3.8). However (3.50) is in fact only a close approximation to the true value of $P_1(p)$, since terms have been ignored in passing from (3.46) to (3.47). On the other hand, if it is possible to find an increasing function $\phi^*(x)$, (which in

practice will be equal to $\phi(x)$ with a slight modifying term) for which the inequality

$$E\left[\phi^*(x_{t+1}) - \phi^*(x_t) | x_t \right] \leq 0$$

holds, then by carrying out an iteration as before we obtain

$$P_0(p)\ \phi^*(0) + P_1(p)\ \phi^*(1) - \phi^*(p) \leq 0$$

so that $$P_1(p) \leq \frac{\phi^*(p) - \phi^*(0)}{\phi^*(1) \quad \phi^*(0)} \tag{3.51}$$

since $\phi^*(1) - \phi^*(0)$ is positive.

Once more it may be noted that $\phi^*(x_{t+1})$, as defined above, is a semi-martingale, so that (3.51) can be derived more strictly by using optional stopping theorems for semi-martingales. The inequality (3.51) provides an upper bound for the exact value of $P_1(p)$, approximated by (3.50). Similarly if there exists an increasing function $\phi^{**}(x)$ for which

$$E\left[\phi^{**}(x_{t+1}) - \phi^{**}(x_t) | x_t \right] \geq 0$$

then

$$P_1(p) \geq \frac{\phi^{**}(p) - \phi^{**}(0)}{\phi^{**}(1) - \phi^{**}(0)} \tag{3.52}$$

thus providing a lower bound for $P_1(p)$. In subsequent chapters bounds for $P_1(p)$ will be derived where exact evaluation is very difficult, and these bounds may be compared with those obtained by the methods of Chapter 2. It may be noted in this connection that if $m(x)$ and $v(x)$ are constant, then $\phi(x)$ will be of the form

const exp(cx), and that this is identical to the "almost invariant (moment-generating) function" discussed in Chapter 2.

As a second use of such functions we consider the problem of finding an approximation, and also bounds, for the mean time until either x = 0 or x = 1 is reached. In this case it is supposed that the function $\phi(x)$ satisfies the boundary conditions $\phi(0) = \phi(1) = 0$ and also the relation

$$E\left[\, \phi(x_{t+1}) - \phi(x_t) \,|\, x_t \,\right] = -N^{-\alpha} \tag{3.53}$$

By expanding $\phi(x_{t+1})$ about x_t as before and ignoring terms which are $o(N^{-\alpha})$ it follows that

$$m(x)\,\phi'(x) + \tfrac{1}{2}v(x)\,\phi''(x) = -1 \tag{3.54}$$

which is the same as (3.5). The solution of (3.54), subject to $\phi(0) = \phi(1) = 0$, is

$$\phi(x) = 2C\int_0^x \psi(y)\,dy - 2\int_0^x \psi(y) \int^y \left[v(z)\psi(z)\right]^{-1} dz\, dy$$

where

$$C = \left[\int_0^1 \psi(y) \int^y \left[v(z)\psi(z)\right]^{-1} dz\, dy\right] \div \left[\int^1 \psi(y)\,dy\right] \tag{3.55}$$

and $\psi(x)$ has been defined in (3.10).

By iterating in (3.53) until x = 0 or x = 1 is reached, we obtain, since $\phi(0) = \phi(1) = 0$,

$$U(p) = \phi(p)N^{\alpha} \tag{3.56}$$

where $U(p)$ is the mean time until absorption at x = 0 or x = 1 and

and p is the initial value of x. It may be verified, when allowance is made for the different time scales, that (3.56) is identical to (3.14).

The solution (3.56) for U(p) is again only approximate, because of the terms which are ignored in passing from (3.53) to (3.54). However bounds may again be found for U(p) by deriving functions $\phi^*(x)$ and $\phi^{**}(x)$ for which

$$E\left[\ \phi^*(x_{t+1}) - \phi^*(x_t) \mid x_t\ \right] \leqq -N^{-\alpha}$$

and

$$E\left[\ \phi^{**}(x_{t+1}) - \phi^*(x_t) \mid x_t\ \right] \geqq -N^{-\alpha}$$

It is found readily that

$$N^{\alpha}\ \phi^{**}(p) \leqq U(p) \leqq N^{\alpha}\ \phi^*(p),$$

and thus bounds are given for the true mean absorption time. Once more these bounds will be closely related functionally to the diffusion approximation $N^{\alpha}\ \phi(p)$.

CHAPTER 4

THE MEAN ABSORPTION TIME

IN A GENETICAL MODEL

41. Introduction

In this chapter explicit expressions are obtained for both the mean time until homozygosity is reached and for the variance of this time in the genetical model introduced by Moran (1958a). This model describes the behaviour of a haploid population of fixed size N with two possible alleles A and a. Attention is fixed on the number j of A individuals. Then in this model, j is a Markovian variate with transition matrix

$$p_{j,j-1} = \mu_2 j(N-j)\left[N\left\{\mu_1 j + \mu_2(N-j)\right\}\right]^{-1} \equiv \Pi_j$$

$$p_{j,j+1} = \mu_1 j(N-j)\left[N\left\{\mu_1 j + \mu_2(N-j)\right\}\right]^{-1} \equiv \eta_j \qquad (4.1)$$

$$p_{j,j} = 1 - p_{j,j-1} - p_{j,j+1} \equiv 1-\Pi_j-\eta_j.$$

This model is considered in Chapter 2 (as case 1), and the result of Chapter 2 will be needed later. Watterson (1961) has obtained the mean time and the variance in the case of no selection (i.e. $\mu_1 = \mu_2$); thus the present results include his as a particular case. It should be noted that the methods given here are different from those of Watterson; the present argument proceeds by analogy with that of the previous chapter in considering pseudo-transient distributions. In fact we shall find an exact (discrete) pseudo-transient distribution (which may be compared with the diffusion approximation pseudo-transient distribution), the sum of whose terms gives the required mean time. Finally the exact results may be compared with those derived from equations (3.5)

and (3.6).

4.2 A Lemma

It is useful to establish the following lemma. Consider a finite Markov chain with two absorbing states, the remaining states being transient, the states being labelled so that the transition matrix appears in the form

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \multicolumn{2}{c}{R} & \multicolumn{3}{c}{Q} \end{pmatrix}$$

Then (c.f. Kemeny and Snell (1960)) if the process starts in the k^{th} transient state, the mean time until an absorbing state is entered is the k^{th} element in the column vector

$$\underset{\sim}{m} = \left[1 - Q \right]^{-1} \underline{\psi}$$

where $\underline{\psi}$ is a column vector of unities. If the process is amended so that whenever an absorbing state is entered, the process is restarted again in the initial state, then P must be amended to

$$P^* = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \multicolumn{2}{c}{R} & \multicolumn{3}{c}{Q} \end{pmatrix}$$

where the unities appear in the position corresponding to the initial state. All the states in the new process are persistent, and the process admits a stationary vector $\underset{\sim}{\lambda}$ satisfying

$$\underset{\sim}{\lambda}' \left[I - P^* \right] = \underset{\sim}{0}'$$

or

$$\underset{\sim}{\lambda}' \begin{pmatrix} 1 & 0 & 0 & \cdots & -1 & \cdots & 0 \\ 0 & 1 & 0 & & -1 & & 0 \\ \multicolumn{3}{c}{-R} & & \multicolumn{4}{c}{I - Q} \end{pmatrix} = \underset{\sim}{0}'$$

If $\underset{\sim}{\lambda}$ is normalized so that the sum of its first two elements is unity, and the normalized vector is written $(\lambda_1 \ \lambda_2 \ \underset{\sim}{\lambda}^*)'$, then

$$\underset{\sim}{\lambda}^* \left[I - Q \right] = \underset{\sim}{\xi}'$$

where $\underset{\sim}{\xi}'$ is a row vector all of whose elements are zero, except the k^{th}, which is unity. Hence

$$\underset{\sim}{\lambda}^{*\prime} = \underset{\sim}{\xi}' \left[I - Q \right]^{-1}$$

and therefore

$$\underset{\sim}{\lambda}^{*\prime} \underset{\sim}{\psi} = \underset{\sim}{\xi}' \left[I - Q \right]^{-1} \underset{\sim}{\psi} \quad . \tag{4.2}$$

Equation (4.2) states that the sum of the elements in $\underset{\sim}{\lambda}^*$ is equal to the sum of the elements in the k^{th} row of $[I - Q]^{-1}$; i.e. is equal to the k^{th} element in $\underset{\sim}{m}$. The elements in $\underset{\sim}{m}$ will in fact be derived for the particular Markov chain under consideration by finding first the elements of $\underset{\sim}{\lambda}^*$.

4.3 Mean Absorption Time

If the state E_i in the genetical model (4.1) is defined as "number of A individuals = i, i = 0,1,2,...,N", then E_0 and E_N are absorbing and the transition matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \Pi_1 & 1-\Pi_1-\eta_1 & \eta_1 & \cdots & 0 \\ 0 & \Pi_2 & 1-\Pi_2-\eta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Suppose now that initially there were k A individuals. Then the genetic model, and hence also P, is amended by putting $P_{0k} = p_{Nk} = 1$. If the amended transition matrix is P*, then

$$I - P^* = \begin{pmatrix} 1 & 0 & 0 & \cdots & -1 & \cdots & 0 \\ -\Pi_1 & \Pi_2+\eta_1 & -\eta_1 & \cdots & 0 & \cdots & 0 \\ 0 & -\Pi_2 & \Pi_2+\eta_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & \cdots & 0 \end{pmatrix}$$

Denote the stationary vector of the new process by $\underset{\sim}{\lambda}' = (\lambda_0\ \lambda_1\ \ldots\ \lambda_N)$ and put $\lambda_0 = P_0$ (a constant) for the moment. Then since $\underset{\sim}{\lambda}'\,[1-P^*] = \underset{\sim}{0}'$, consideration of the first two equations

$$\lambda_0 - \lambda_1\Pi_1 = 0$$

$$\lambda_1(\Pi_1 + \eta_1) - \lambda_2\Pi_2 = 0$$

gives

$$\left.\begin{aligned} \lambda_1 &= P_0/\Pi_1 \\ \lambda_2 &= P_0(1+\alpha)/\Pi_2 \end{aligned}\right\} \qquad (4.3)$$

where α μ_1/μ_2. Also λ_i $(1 \leq i \leq k-3)$ obeys the recurrence relation

$$-\eta_i\lambda_i + (\eta_{i+1} + \Pi_{i+1})\,\lambda_{i+1} - \Pi_{i+2}\lambda_{i+2} = 0 \qquad (4.4)$$

This relation is simplified by putting $\Pi_i\lambda_i = \xi_i$ and by noting that $\eta_i/\Pi_i = \alpha$ = constant. Thus (4.4) becomes

$$-\alpha\,\xi_i + (\alpha+1)\xi_{i+1} - \xi_{i+2} = 0 \qquad (4.5)$$

the solution of which is

$$\xi_i = A + B\alpha^i \quad (A,B, \text{ arbitrary constants}).$$

Using the boundary conditions (4.3), which are written more conveniently

$$\left.\begin{aligned} \xi_1 &= P_0 \\ \xi_2 &= P_0(1+\alpha) \end{aligned}\right\} \qquad (4.6)$$

we obtain

$$A = -P_0/(\alpha-1)\ , \quad B = P_0/(\alpha-1)$$

so that

$$\xi_i = P_0(\alpha^i-1)/(\alpha-1)$$

and hence

$$\lambda_i = P_0(\alpha^i-1)/\left[\,\Pi_i(\alpha-1)\right] \quad (i=1,2,\ldots,k-1) \qquad (4.7)$$

The elements $\lambda_{k+1} \ldots \lambda_N$ are found as follows. Putting for the moment $\lambda_N = P_N$ (P_N a constant) we obtain

$$-\eta_{N-1}\,\lambda_{N-1} + P_N = 0$$

$$-\eta_{N-2}\,\lambda_{N-2} + (\Pi_{N-1} + \eta_{N-1})\,\lambda_{N-1} = 0$$

Hence

$$\left.\begin{aligned} \lambda_{N-1} &= P_N/\eta_{N-1} \\ \lambda_{N-2} &= P_M(1+\alpha^{-1})/\eta_{N-2} \end{aligned}\right\} \qquad (4.8)$$

If ξ_i is now defined by $\xi_i = \eta_i\lambda_i$, then from the recurrence relation (4.4) we obtain

$$-\xi_i + (1+\alpha^{-1})\xi_{i+1} - \alpha^{-1}\,\xi_{i+2} = 0$$

or

$$-\alpha\,\xi_i + (\alpha+1)\,\xi_{i+1} - \xi_{i+2} = 0$$

which is the same as (4.5). Thus the general solution is

$$\xi_i = C + D\alpha^i \quad (C,\ D,\ \text{arbitrary constants})$$

subject to the boundary conditions (4.8), which are more conveniently written

$$\xi_{N-1} = P_N$$

$$\xi_{N-2} = P_N(1+\alpha^{-1}) \ .$$

These conditions fix C and D as

$$C = P_N\,\alpha^N/(\alpha^N-\alpha^{N-1}) \ , \quad D = -P_N/(\alpha^N-\alpha^{N-1})$$

Therefore

$$\xi_i = P_N(\alpha^N-\alpha^i)/(\alpha^N-\alpha^{N-1})$$

and

$$\lambda_i = P_N(\alpha^N-\alpha^i)/\left[\ \eta_i(\alpha^N-\alpha^{N-1})\right]$$

$$(i = k+1,\ldots,N-1) \qquad (4.9)$$

It remains to evaluate λ_k. We have

$$-\lambda_0 - \eta_{k-1}\lambda_{k-1} + (\eta_k + \Pi_k)\lambda_k - \Pi_{k+1}\lambda_{k+1} - \lambda_N = 0$$

$$\text{or } \lambda_k = \left\{P_0 + P_N + P_0\left(\frac{\alpha^k - 1}{\alpha - 1}\right) + P_N\left(\frac{\alpha^{N-1} - \alpha^k}{\alpha^N - \alpha^{N-1}}\right)\right\} \div \left\{\Pi_k + \eta_k\right\}$$

... (4.10)

Thus all elements in the stationary distribution have been evaluated in terms of P_0 and P_N. These elements are now normalized so that $P_0 + P_N = 1$. Also it is clear that

$$\frac{P_0}{P_N} = \frac{\text{Prob \{absorption at 0 in absorbing case\}}}{\text{Prob \{absorption at N in absorbing case\}}}$$

$$= \frac{\alpha^{N-k} - 1}{\alpha^N - 1} \div \frac{\alpha^N - \alpha^{N-k}}{\alpha^N - 1}$$

using the result of Chapter 2. Since $P_0 + P_N = 1$ we have

$$P_0 = \frac{\alpha^{N-k} - 1}{\alpha^N - 1}$$

$$P_N = \frac{\alpha^N - \alpha^{N-k}}{\alpha^N - 1}$$

Thus with these values of P_0 and P_N, all elements in the normalized vector $\underset{\sim}{\lambda}'$ are given explicitly by (4.7), (4.9), and (4.10). It follows, by using the above values for P_0 and P_N, that (4.10) simplifies to

$$\lambda_k = \frac{\alpha^k - 1}{\alpha - 1} \cdot \frac{\alpha^{N-k} - 1}{\Pi_k(\alpha^N - 1)}$$

which is what would have been obtained by putting i = k in (4.7) or alternatively by putting i = k in (4.9), (although the functional form of $\lambda_1 \ldots \lambda_{k-1}$ differs from that of $\lambda_{k+1} \ldots \lambda_{N-1}$). Thus the two sets of λ's "dovetail" at λ_k . By using the results at the beginning of the chapter it follows that the mean time $\mu(k)$ until one or other gene is eliminated, given initially k A individuals, is the sum of mean number of times the process is in each of the various transient states, i.e.

$$\mu(k) = \frac{\alpha^{N-k}-1}{\alpha^N-1} \sum_{i=1}^{k} \frac{\alpha^i-1}{\Pi_i(\alpha-1)} + \frac{\alpha^N-\alpha^{N-k}}{\alpha^N-1} \sum_{i=k+1}^{N} \frac{\alpha^N-\alpha^i}{\eta_i(\alpha^N-\alpha^{N-1})} \quad \ldots \quad (4.11)$$

which is the required result. The various elements in(4.11) constitute the exact pseudo-transient function of the process.

An alternative method for finding (4.11) is to solve the difference equation (c.f. Feller (1957)),

$$\mu(k) = \Pi_k\mu(k-1) + (1-\Pi_k-\eta_k)\,\mu(k) + \eta_k\mu(k+1) + 1 \qquad (4.12)$$

However it is difficult to proceed directly from (4.12), and our interest in this equation is that it provides a check on the solution (4.11), and after some algebra it may be shown that the solution (4.11) does in fact solve (4.12), as well as the boundary conditions $\mu(0) = \mu(N) = 0$.

As a second (partial) check (of 4.11) we proceed as follows. The method of deriving $\mu(k)$ above indicates that the

mean number h_i of times the process is in state i before absorption is given by

$$h_i = \frac{\alpha^{N-k} - 1}{\alpha^N - 1} \cdot \frac{\alpha^i - 1}{\Pi_i(\alpha-1)} \qquad i \leq k$$

$$h_i = \frac{\alpha^N - \alpha^{N-k}}{\alpha^N - 1} \cdot \frac{\alpha^N - \alpha^i}{\eta_i(\alpha^N - \alpha^{N-1})} \qquad i > k$$

Now given that the process is in state E_i, it will leave E_i after the next birth-death event with probability $\Pi_i + \eta_i$. If each h_i is multiplied by $\Pi_i + \eta_i$, the product is the mean number of times the process is in E_i and then leaves E_i at the next birth-death event. Thus the mean number of transitions is

$$\sum_{i=1}^{N-1} (\Pi_i + \eta_i) h_i \tag{4.13}$$

But for processes where transitions of the type $E_i \to E_i$ are ignored the transition probabilities become

$$\left.\begin{aligned} P(E_i \to E_{i+1}) &= \Pi_i(\Pi_i + \eta_i)^{-1} = (1+\alpha)^{-1} \\ P(E_i \to E_{i+1}) &= \eta_i(\Pi_i + \eta_i)^{-1} = \alpha(1+\alpha)^{-1} \end{aligned}\right\} \tag{4.14}$$

These probabilities are those of a homogeneous asymmetric random walk, and the mean number of steps before one or other boundary is reached is given by (Feller (1957))

$$k\frac{(1+\alpha)}{1-\alpha} - \frac{1+\alpha}{1-\alpha} \cdot \frac{N}{\alpha^N - 1}(\alpha^N - \alpha^{N-k}) \tag{4.15}$$

Thus (4.13) should equal (4.15), and it is a matter of algebra to prove that this is so. In fact an alternative method of finding (4.11) is to start from the fact that in the random walk on [0,N] with initial point k, and with $p_{i,i-1} = (1+\alpha)^{-1}$, $p_{i,i+1} = \alpha(1+\alpha)^{-1}$, then the mean number of times n_i that the process is at i before reaching either 0 or N is given by

$$\left.\begin{aligned} n_i &= \frac{\alpha^{N-k} - 1}{\alpha^N - 1} \cdot \frac{\alpha+1}{\alpha-1} \cdot (\alpha^i - 1) && i \leq k \\ &= \frac{\alpha^N - \alpha^{N-k}}{\alpha^N - 1} \cdot \frac{\alpha+1}{\alpha^{N+1} - \alpha^N} \cdot (\alpha^N - \alpha^i) && i > k \end{aligned}\right\} \qquad (4.16)$$

Multiplying each n_i by $(\Pi_i + \eta_i)^{-1}$ and adding over i = 1,2,...,N-1, we re-obtain (4.11).

In the particular case where there is no selection, then $\alpha = 1$, and the above methods are inappropriate. However by letting $\alpha \to 1$ in (4.11), it follows that

$$\mu(k) = (N-k) \sum_{i=1}^{k} N(N-i)^{-1} + k \sum_{i=1}^{N-k-1} N(N-i)^{-1} \qquad (4.17)$$

which is the result obtained by Watterson (1961). This result could be derived more strictly by considering difference equations similar to (4.4). In fact when $\alpha=1$, (4.4) becomes (putting $\Pi_i \lambda_i = \xi_i$ as before)

$$-\xi_i + 2\xi_{i+1} - \xi_{i+2} = 0$$

the general solution of which is

$$\xi_i = A + Bi \ .$$

Proceeding as before, (4.17) eventually follows. In this case it is possible once more to verify that (4.17) satisfies (4.12).

A partial check similar to that given previously is to note that in this case $\Pi_i + \eta_i = 2i(N-i)N^{-2}$, so that the mean number of times the process is in E_i and then leaves is

$$\left.\begin{array}{ll} \frac{N(N-k)}{N-i} \cdot \frac{2i(N-i)}{N^2} & i \leq k \\ \frac{kN}{i} \cdot \frac{2i(N-i)}{N^2} & i > k \end{array}\right\} \qquad (4.18)$$

Summing the terms in (4.18) gives the total mean number of transitions as $k(N-k)$, the well-known result for the symmetrical random walk.

Once more, (4.17) could have been obtained by arguing in the reverse direction. We obtain firstly that for the symmetrical random walk on $[0,N]$, the mean number n_i of times the process is at i before reaching 0 or N is given by

$$n_i = 2i(N-k)/N \qquad i \leq k$$

$$n_i = 2k(N-i)/N \qquad i > k$$

Multiplying each n_i by $\left[2i(N-i)N^{-2}\right]^{-1}$, and adding over $i = 1,2,..N-1$ gives (4.17).

4.4 The Variance of the Absorption Time

In order to find an exact expression for the variance it may be noted that the elements in the k^{th} row of $[I-Q]^{-1}$ are the

various elements constituting the sums in (4.11). Specifically, if

$$\theta_{ij} = \frac{\alpha^{N-i} - 1}{\alpha^N - 1} \cdot \frac{\alpha^j - 1}{\Pi_j(\alpha-1)}$$

$$\psi_{ij} = \frac{\alpha^N - \alpha^{N-i}}{\alpha^N - 1} \cdot \frac{\alpha^N - \alpha^j}{\eta_j(\alpha^N - \alpha^{N-1})}$$

then

$$\left[I-Q\right]^{-1} = \begin{pmatrix} \theta_{11} & \psi_{12} & \psi_{13} & \psi_{14} & .. & \psi_{1,N-1} \\ \theta_{21} & \theta_{22} & \psi_{23} & \psi_{24} & .. & \psi_{2,N-1} \\ \theta_{31} & \theta_{32} & \theta_{33} & \psi_{34} & .. & \psi_{3,N-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \theta_{N-1,1} & \theta_{N-1,2} & \theta_{N-1,3} & \theta_{N-1,4} & & \theta_{N-1,N-1} \end{pmatrix}$$

Thus the variance of the absorption time (c.f. Kemeny and Snell (1960)) is the k^{th} element in the column vector

$$\left[2(I - Q)^{-1} - I\right](I - Q)^{-1} \underset{\sim}{\xi} - \underset{\sim}{\mu}^2$$

where $\underset{\sim}{\xi}$ is a column vector of unities and $\underset{\sim}{\mu}^2$ is a column vector of the $\mu^2(k)$'s. It follows that

$$\sigma_k^2 = 2 \sum_{i=1}^{k} \theta_{ki} \left\{ \sum_{j=1}^{i} \theta_{ij} + \sum_{j=i+1}^{n-1} \psi_{ij} \right\}$$

$$+ 2 \sum_{i=k+1}^{n-1} \psi_{ki} \left\{ \sum_{j=1}^{i} \theta_{ij} + \sum_{j=i+1}^{n-1} \psi_{ij} \right\} - \mu(k) - \mu^2(k) \qquad (4.19)$$

Letting $\alpha \to 1$,

$$\theta_{ij} \to N(N-i)/(N-j)$$

$$\psi_{ij} \to Ni/j$$

and it is easily checked that in this case (4.19) agrees with the formula found by Watterson (1961).

4.5 Diffusion Approximations

It is possible now to compare the formulae given above with those given by diffusion methods, and thus to test the adequacy of the latter. Before doing so it is useful to approximate (4.11) by a formula involving integrations rather than summations. If the proportion of individuals in the population at time t that are A is denoted x_t, then diffusion methods may only be used when $E(x_{t+1} - x_t)$ is not of higher order of magnitude than $E(x_{t+1} - x_t)^2$. For this to occur it is necessary that $\mu_1 - \mu_2$ be $O(N^{-1})$ or $o(N^{-1})$. Therefore putting $\alpha = \mu_1/\mu_2 = 1 + hN^{-1}$, we have

$$\alpha^N \approx \exp(h) \qquad (4.20)$$

and if $y = iN^{-1}$, $\alpha^i \approx \exp(hy)$.

Thus putting $kN^{-1} = p$, it follows, using (4.11), that $\mu(k)$ may be approximately

$$\frac{N^2}{h}\left[\frac{e^{h(1-p)}-1}{e^h-1}\int_0^p (e^{hy}-1)k_1(y)\,dy\right.$$

$$\left.+\frac{e^h-e^{h(1-p)}}{e^h-1}\int_p^1 \frac{e^h-e^{hy}}{e^h}\,k_2(y)\,dy\right] \qquad (4.21)$$

where $k_1(y)$ and $k_2(y)$ are the "continuity" analogues of Π_i^{-1} and η_i^{-1} and are

$$\frac{\mu_1 y+\mu_2(1-y)}{\mu_2 y(1-y)} \text{ and } \frac{\mu_1 y+\mu_2(1-y)}{\mu_1 y(1-y)} \text{ respectively.}$$

For the diffusion approximation

$$m(x) = (\mu_1-\mu_2)x(1-x)\left[N\left\{\mu_1 x+\mu_2(1-x)\right\}\right]^{-1}$$

$$= hx(1-x)N^{-2} + o(N^{-2})$$

and

$$v(x) = 2x(1-x)N^{-2} + o(N^{-2})$$

Thus (3.5) becomes, if unit time corresponds to N^2 birth-death events,

$$h\frac{dU(p)}{dp}+\frac{d^2U(p)}{dp^2} = p^{-1}(1-p)^{-1}$$

whose solution is (c.f. equation 3.55)

$$U(p) = C\int_0^p \exp(-hy)\,dy - \int_0^p \exp(-hy)\int^y z^{-1}(1-z)^{-1}\exp(hz)\,dz\,dy$$

where

$$C = \left[\int_0^1 \exp(-hy) \int^y z^{-1}(1-z)^{-1} \exp(hz)\,dz\,dy \right] \div \left[\int_0^1 \exp(-hz)\,dz \right]$$

... (4.22)

The similarity between (4.21) and (4.22) is not immediately obvious; however (4.22) may be simplified by an integration by parts and by a subsequent rearrangement of terms. The result of doing this leads to the formula

$$U(p) = \frac{N^2}{h} \left[\frac{e^{h(1-p)} - 1}{e^h - 1} \int_0^p y^{-1}(1-y)^{-1} (e^{hy} - 1)\,dy \right.$$

$$\left. + \frac{e^h - e^{h(1-p)}}{e^h - 1} \int_p^1 y^{-1}(1-y)^{-1} e^{-hy} (e^h - e^{hy})\,dy \right] \quad (4.23)$$

(where time has now been rescaled to birth-death events). In this form the close agreement between (4.21) and (4.23) is apparent. In fact it is clear that the entire pseudo-transient function is uniformly well-approximated by the diffusion expression.

In the case $\mu_1 = \mu_2$, equation (4.23) becomes

$$U(p) = -N^2 \left\{ p \ell n p + (1-p) \ell n (1-p) \right\} \quad (4.24)$$

which agrees closely with (4.17). In this case the diffusion approximation to the pseudo-transient function possesses the remarkable property of being exact for all i, if this function is taken at discrete points. In fact, using (3.27), the equation

$$f(x) = N(1-p)/(1-x) \qquad 0 < x \leq p$$

$$f(x) = N\,p/x \qquad p < x < 1$$

is found for the (normalized) pseudo-transient function, and at points $x = iN^{-1}$ this becomes

$$N(N-k)/(N-i) \qquad i \leq k$$

$$Nk/i \qquad i > k$$

which is exact for all i.

We turn now to the comparison between the exact value (given by (4.19)) and the diffusion approximation (found by solving (3.6)) for the variance of the time until absorption. Considering (4.19) first, the first and last terms are of order N^4, while the second is of order N^2 only. Therefore, to order N^3,

$$\sigma_k^2 = 2 \sum_{i=1}^{k} \theta_{ki}\, \mu(i) + 2 \sum_{i=k+1}^{N-1} \psi_{ki}\, \mu(i) - \mu^2(k)$$

since $\mu(i) = \sum_{j=1}^{i} \theta_{ij} + \sum_{j=i+1}^{N-1} \psi_{ij}$.

By putting $p = kN^{-1}$ and $x = iN^{-1}$, and $\mu(k) = U(p)$, this expression may be approximated by

$$2N^4\left[\frac{e^{h(1-p)}-1}{e^h-1}\int_0^p (e^{hy}-1)k_1(y)U(y)\,dy\right.$$

$$\left.+\frac{e^h-e^{h(1-p)}}{e^h-1}\int_p^1 \frac{e^h-e^{hy}}{e^h}k_2(y)U(y)dy\right]-\mu^2(k) \qquad (4.25)$$

For the diffusion approximation, the solution of the equation

$$-2U(p) = \tfrac{1}{2}v(p)\frac{d^2S(p)}{dp^2} + m(p)\frac{dS(p)}{dp}$$

satisfied by the expected value $S(p)$ of the square of the time taken until absorption, subject to boundary conditions $S(0) = S(1) = 0$, is found to be, after some simplification,

$$S(p) = N^4\left[\frac{e^{h(1-p)}-1}{e^h-1}\int_0^p (e^{hy}-1)\,U(y)\,[v(y)]^{-1}\,dy\right.$$

$$\left.+\frac{e^h-e^{h(1-p)}}{e^h-1}\int_p^1 \frac{e^h-e^{hy}}{e^h}\,U(y)\,[v(y)]^{-1}\,dy\right] \qquad (4.26)$$

Now $2k_1(y) \approx 2k_2(y) \approx [v(y)]^{-1}$, so that if the square of the diffusion approximation for $\mu(k)$, (known to be a close approximation) is subtracted from (4.26), and the resulting expression compared with (4.25), it is observed that the diffusion approximation is once more remarkably close to (4.25) and thus close to the true value.

CHAPTER 5.

ARBITRARY PROCESSES WITH CONTINUANT TRANSITION MATRICES

5.1 Introduction

In this chapter the results of Chapter 4 are generalized, and expressions are derived for absorption probabilities, mean absorption times, variances, pseudo-transient functions, etc. in the case where the transition matrix of the variate under consideration is an arbitrary continuant. By doing this exact results may be obtained for any population with overlapping generations. We begin by noting a result which is a generalization of the results of the previous chapter.

5.2 A Theorem on Continuants

Let P be the transition matrix of a Markov chain with N+1 states E_0, E_1, ..., E_N, for which E_0 and E_N are absorbing and the remaining states transient. Suppose also that P is a continuant, so that given the process is in state E_i, it may next move only to states E_{i-1}, E_i, and E_{i+1}, with probabilities Π_i $1-\Pi_i-\eta_i$, and η_i respectively. Let E_k be the initial state. Then the probability that the process is eventually absorbed in E_0 rather than E_N is

$$P(0,k) = \left\{ \sum_{i=k}^{N-1} \rho_i \right\} \div \left\{ \sum_{i=0}^{N-1} \rho_i \right\} \qquad (5.1)$$

where $\rho_0 = 1$, $\rho_i = \dfrac{\Pi_1 \Pi_2 \cdots \Pi_i}{\eta_1 \eta_2 \cdots \eta_i}$ $\qquad (i > 0)$.

If each transition takes place at unit time intervals, then the mean time before absorption in one or other absorbing state is

$$\mu(k) = \sum_{i=1}^{N-1} n_i \tag{5.2}$$

where n_i is the mean time the process is in state E_i and is given by

$$n_i = \frac{\sum_{j=k}^{N-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=0}^{i-1} \rho_j}{\Pi_i \rho_{i-1}} \qquad i = 1,2,..,k$$

$$n_i = \frac{\sum_{j=0}^{k-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=i}^{N-1} \rho_j}{\eta_i \rho_i} \qquad i = k+1,\ldots,N-1 \tag{5.3}$$

Further, the variance of the absorption time is given by equation (4.19) if we put

$$\theta_{ik} = \frac{\sum_{j=k}^{N-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=0}^{i-1} \rho_j}{\Pi_i \rho_{i-1}} \qquad (5.4)$$

and

$$\psi_{ik} = \frac{\sum_{j=0}^{k-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=i}^{N-1} \rho_j}{\eta_i \rho_i}$$

and $\mu(k)$ is defined by (5.2) and (5.3).

The result (5.1) is well-known. The mean number of steps before absorption, given jointly by (5.2) and (5.3) has been given (incorrectly) by Chung (1960), p.70, equation 8. The present derivation obtains all four quantities in question simultaneously.

Proof

As in Chapter 4 we consider the associated "return" process for which whenever the process enters either E_0 or E_N it is immediately restarted at E_k. If P* is the transition matrix of the return process, then

$$I-P^* = \begin{bmatrix} 1 & 0 & 0 & \cdots & -1 & \cdots & 0 \\ -\Pi_1 & \Pi_1+\eta_1 & -\eta_1 & \cdots & 0 & \cdots & 0 \\ 0 & -\Pi_2 & \Pi_2+\eta_2 & \cdots & 0 & & 0 \\ \vdots & \vdots & \vdots & & & & \\ 0 & 0 & 0 & \cdots & 0 & \cdots & -\eta_{N-1} \\ 0 & 0 & 0 & \cdots & -1 & \cdots & 1 \end{bmatrix}$$

The return process admits a stationary vector $\underset{\sim}{\lambda}'$ satisfying

$$\underset{\sim}{\lambda}' \left[I - P^* \right] = \underset{\sim}{0}' \qquad (5.5)$$

The elements in $\underset{\sim}{\lambda}'$ are now inflated so that the sum of the first and last elements (now written P_0 and P_N respectively), is unity. From the discussion in the previous chapter it is clear that P_0 is in fact the probability $P(0,k)$ that in the original non-return process, the state E_0 is eventually reached rather than E_N. Further, the remaining elements in $\underset{\sim}{\lambda}'$ are the mean number of times in the original process that the process is in each of the various transient states before being absorbed in either E_0 or E_N. Denoting these elements $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ it follows, using (5.5), that

$$P_0 = \Pi_1\lambda_1$$

$$(\Pi_1 + \eta_1)\,\lambda_1 = \Pi_2\lambda_2 \qquad (5.6)$$

and in general, for $2 \leq i \leq k-1$,

$$(\Pi_i + \eta_i)\lambda_i = \Pi_{i+1}\lambda_{i+1} + \eta_{i-1}\lambda_{i-1} \cdot \qquad (5.7)$$

(5.7) may be rewritten

$$\Pi_{i+1}\lambda_{i+1} - \eta_i\lambda_i = \Pi_i\lambda_i - \eta_{i-1}\lambda_{i-1}$$

or $$\Pi_{i+1}\lambda_{i+1} - \eta_i\lambda_i = \text{const.}, \quad 2 \leqq i \leqq k-1 \cdot \qquad (5.8)$$

By considering i = 2, it follows that the constant in (5.8) is P_0. Thus

$$\Pi_1\lambda_1 = P_0$$

$$\Pi_2\lambda_2 = P_0\left[1 + \frac{\eta_1}{\Pi_1}\right]$$

$$\Pi_3\lambda_3 = P_0\left[1 + \frac{\eta_2}{\Pi_2} + \frac{\eta_2\eta_1}{\Pi_2\Pi_1}\right]$$

and in general

$$\Pi_i\lambda_i = P_0\left[1 + \frac{\eta_{i-1}}{\Pi_{i-1}} + \frac{\eta_{i-1}\eta_{i-2}}{\Pi_{i-1}\Pi_{i-2}} + \ldots + \frac{\eta_{i-1}\eta_{i-2}\cdots\eta_1}{\Pi_{i-1}\Pi_{i-2}\cdots\Pi_1}\right]$$

$$\ldots \qquad (5.9)$$

Considering now values of i > k, we obtain, using (5.8)

$$\left.\begin{aligned} P_N &= \lambda_{N-1}\,\eta_{N-1} \\ \lambda_{N-1}(\eta_{N-1} + \Pi_{N-1}) &= \lambda_{N-2}\,\eta_{N-2} \end{aligned}\right\} \qquad (5.10)$$

Further, equations (5.7) and (5.8) both continue to hold provided

that in (5.8) a possibly different constant is used and the range of i is altered to $k \leq i \leq N-2$. The (new) constant in (5.8) may be evaluated by using the case i = N-2, from which it is found that the constant is $-P_N$. This gives

$$\lambda_{N-1}\,\Pi_{N-1} = P_N\,\frac{\Pi_{N-1}}{\eta_{N-1}}$$

$$\lambda_{N-2}\,\Pi_{N-2} = P_N\left[\frac{\Pi_{N-2}}{\eta_{N-2}} + \frac{\Pi_{N-2}\,\Pi_{N-1}}{\eta_{N-2}\,\eta_{N-1}}\right]$$

and in general

$$\lambda_i\Pi_i = P_N\left[\frac{\eta_i}{\eta_i} + \frac{\Pi_i\Pi_{i+1}}{\eta_i\eta_{i+1}} + \ldots + \frac{\Pi_i\Pi_{i+1}\cdots\Pi_{N-1}}{\eta_i\eta_{i+1}\cdots\eta_{N-1}}\right] \qquad (5.11)$$

$$\text{for } i = k+1,\ k+2,\ldots,N-1.$$

(5.9) and (5.11) give together

$$\lambda_i = \frac{P_0}{\Pi_i}\left[1 + \frac{\eta_{i-1}}{\Pi_{i-1}} + \frac{\eta_{i-1}\eta_{i-2}}{\Pi_{i-1}\Pi_{i-2}} + \ldots + \frac{\eta_{i-1}\eta_{i-2}\cdots\eta_1}{\Pi_{i-1}\Pi_{i-2}\cdots\Pi_1}\right] i=1,2,..k-1$$

$$\lambda_i = \frac{P_N}{\Pi_i}\left[\frac{\Pi_i}{\eta_i} + \frac{\Pi_i\Pi_{i+1}}{\eta_i\eta_{i+1}} + \ldots + \frac{\Pi_i\Pi_{i+1}\cdots\Pi_{N-1}}{\eta_i\eta_{i+1}\cdots\eta_{N-1}}\right] i=k+1,..,N-1$$

... (5.12)

Remembering the result in the paragraph preceding equation (4.11) in the previous chapter, it may be suspected that the two functions in (5.12) "dovetail" at i = k. If this were so, then we would have

$$P_0\left[1 + \frac{\eta_{k-1}}{\Pi_{k-1}} + \ldots + \frac{\eta_{k-1}\eta_{k-2}\cdots\eta_1}{\Pi_{k-1}\Pi_{k-2}\cdots\Pi_1}\right] = P_N\left[\frac{\Pi_k}{\eta_k} + \ldots + \frac{\Pi_k\Pi_{k+1}\cdots\Pi_{N-1}}{\eta_k\eta_{k+1}\cdots\eta_{N-1}}\right].$$

... (5.13)

$$P^{0}\left[1+\frac{\Pi^{K-1}}{J^{K-1}}+\cdots+\frac{\Pi^{K-1}\Pi_{K-2}\cdots\Pi^{1}}{J^{K-1}J_{K-2}\cdots J^{1}}\right]\cdot\left[\frac{J^{K}}{\Pi^{K}}+\cdots+\frac{J^{K}J_{K+1}\cdots J^{N-1}}{\Pi^{K}\Pi_{K+1}\cdots\Pi^{N-1}}\right]=P^{N} \quad \cdots \quad (5.15)$$

in (5.12) "dovetail" at $i = k$. If this were so, then we would have

Remembering the result in the paragraph preceding equation (4.11)

in the previous chapter, it may be suspected that the two functions

Putting $P_N = 1-P_0$ and solving for P_0, the equation

$$P_0 = \frac{\frac{\Pi_1 \cdots \Pi_k}{\eta_1 \cdots \eta_k} + \frac{\Pi_1 \cdots \Pi_{k+1}}{\eta_1 \cdots \eta_{k+1}} + \cdots + \frac{\Pi_1 \cdots \Pi_{N-1}}{\eta_1 \cdots \eta_{N-1}}}{1 + \frac{\Pi_1}{\eta_1} + \frac{\Pi_1 \Pi_2}{\eta_1 \eta_2} + \cdots + \frac{\Pi_1 \Pi_2 \cdots \Pi_{N-1}}{\eta_1 \eta_2 \cdots \eta_{N-1}}}$$

$$= \left\{ \sum_{i=k}^{N-1} \rho_i \right\} \div \left\{ \sum_{i=0}^{N-1} \rho_i \right\} \qquad (5.14)$$

where ρ_i has been defined previously, is obtained. To show that P_0, defined by (5.14), is the probability of absorption at zero, P_0 is written more fully as $P_{0,k}$. Then if $P_{0,k}$ satisfies the boundary conditions

$$P_{0,0} = 1 \quad , \quad P_{0,N} = 0$$

as well as the difference equation

$$P_{0,k} = \Pi_k P_{0,k-1} + (1-\Pi_k-\eta_k) P_{0,k} + \eta_k P_{0,k+1}$$

for $1 \leq k \leq N-1$, then $P_{0,k}$ is the required probability $P(0,k)$. Clearly the boundary conditions are satisfied, and by writing the difference equation in the form

$$\Pi_k(P_{0,k-1} - P_{0,k}) = \eta_k(P_{0,k} - P_{0,k+1})$$

it is easily verified that the difference equation is satisfied by (5.14). Therefore $P_{0,k} = P(0,k)$, the required probability.

Equations (5.12) and (5.14) may be used jointly to show that the mean time the process is in state E_i before absorption is

$$n_i = \frac{\sum_{j=k}^{N-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=0}^{i-1} \rho_j}{\Pi_i \rho_{i-1}} \qquad i=1,2,\ldots, k$$

$$n_i = \frac{\sum_{j=0}^{k-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=i}^{N-1} \rho_j}{\eta_i \rho_i} \qquad i=k+1,\ldots,N-1 \tag{5.15}$$

which verifies (5.3), and clearly the mean time until the process is absorbed at one or other barrier is

$$\sum_{i=1}^{N-1} n_i \tag{5.16}$$

Furthermore, the variance may be obtained immediately by noting that the various elements in (5.15) constitute the k^{th} row in $(I-Q)^{-1}$, where Q is the submatrix of P corresponding to transitions between transient states. Thus if we define

$$\theta_{ik} = \frac{\sum_{j=k}^{N-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=0}^{i-1} \rho_j}{\Pi_i \rho_{i-1}}$$

$$\text{and } \psi_{ik} = \frac{\sum_{j=0}^{k-1} \rho_j}{\sum_{j=0}^{N-1} \rho_j} \cdot \frac{\sum_{j=i}^{N-1} \rho_j}{\eta_i \rho_i} \qquad (5.17)$$

then the variance of the absorption time is given by equation (4.19), where $\mu(k)$ is given by (5.2) and (5.3).

The results (5.1), (5.2), (5.3), and (5.4) may now be applied directly to various genetical models.

5.3 A selection and dominance model

Consider first a haploid population of size N where the number of individuals which are A is a Markovian variate with transition probabilities

$$\begin{aligned} p_{i,i-1} &= i(N-i)N^{-2}\left\{1 - \tfrac{1}{2}s(iN^{-1} + h(N-2i)N^{-1})\right\} = \Pi_i \\ p_{i,i+1} &= i(N-i)N^{-2}\left\{1 + s(iN^{-1} + h(N-2i)N^{-1})\right\} = \eta_i \\ p_{i,i} &= 1 - p_{i,i-1} - p_{i,i+1} \end{aligned} \qquad (5.18)$$

Here s and h are constants for which s is $O(N^{-1})$ and h

is in (0,1). This haploid population approximates to a diploid population with genotypes AA, Aa, and aa having selective advantages 1+s, 1+sh, and 1 respectively, with the three genotypes occurring with frequencies specified by the Hardy-Weinberg law. Thus s is a measure of selection and h a measure of dominance. This model generalizes a case considered by Moran (1963) which is discussed later. Then using the previous results it may be stated immediately that the probability $P(0,k)$ that the process is eventually absorbed at 0, given initially k A individuals, is given by

$$P(0,k) = \frac{\sum_{i=k}^{N-1} \prod_{j=1}^{i} \left[\frac{1-\frac{1}{2}s(jN^{-1}+h(N-2j)N^{-1})}{1+\frac{1}{2}s(jN^{-1}+h(N-2j)N^{-1})} \right]}{\sum_{i=k}^{N-1} \prod_{j=1}^{i} \left[\frac{1-\frac{1}{2}s(jN^{-1}+h(N-2j)N^{-1})}{1+\frac{1}{2}s(jN^{-1}+h(N-2j)N^{-1})} \right]} \qquad (5.19)$$

with empty products conventionally defined as unity (c.f. the definition of ρ_0).

Further, the mean number of times n_j the number of A individuals is j is given by (5.15) where

$$\rho_i = \prod_{j=1}^{i} \frac{1-\frac{1}{2}s\left\{jN^{-1} + h(N-2j)N^{-1}\right\}}{1+\frac{1}{2}s\left\{jN^{-1} + h(N-2j)N^{-1}\right\}} \qquad (5.20)$$

The mean time until absorption will be the sum of the n_i (i=1,2,...,N-1), while the variance of the absorption time is given

by (4.19) and (5.17), if ρ_j is defined as in (5.20).

Now the above quantities, while being exact, are awkward and it would be useful to find approximations for them. Since s is $O(N^{-1})$ and therefore small,

$$1 \mp \tfrac{1}{2}s \left\{ j\, N^{-1} + h(N-2j)N^{-1} \right\}$$

may be approximated by

$$\exp\left[\mp \tfrac{1}{2}s \left\{ jN^{-1} + h(N-2j)N^{-1} \right\} \right]$$

Thus ρ_i may be approximated by

$$\exp\left[-s \left\{ \tfrac{1}{2}i(i+1)N^{-1} + hi - hi(i+1)N^{-1} \right\} \right]$$

Therefore $P(0,k)$ may be approximated by

$$\frac{\int_k^{N-1} \exp\left[-2\left\{ \tfrac{1}{2}i(i+1)N^{-1} + hi - hi(i+1)N^{-1} \right\} \right] di}{\int_0^{N-1} \exp\left[-2\left\{ \tfrac{1}{2}i(i+1)N^{-1} + hi - hi(i+1)N^{+1} \right\} \right] di}$$

$$\approx \frac{\int_p^1 \exp\left[-\alpha h x - \tfrac{1}{2}\alpha x^2(1-2h) \right] dx}{\int_0^1 \exp\left[-\alpha h x - \tfrac{1}{2}\alpha x^2(1-2h) \right] dx} \qquad (5.21)$$

$= \bar{P}(0,k)$, say, where $\alpha = sN$ and $k = Np$.

Similarly, it may be shown after some algebra that the exact mean absorption time may be approximated by the expression

$$N^2\Big[\bar{P}(0,k)\int_0^p 2y^{-1}(1-y)^{-1}\exp\Big\{h\alpha y+\tfrac{1}{2}\alpha y^2(1-2h)\Big\}\int_0^y \exp\ -h\alpha z-\tfrac{1}{2}\alpha z^2(1-2h)\,dz\ dy$$

$$+\Big\{1-\bar{P}(0,k)\Big\}\int_p^1 2y^{-1}(1-y)^{-1}\exp\Big\{h\alpha y+\tfrac{1}{2}\alpha y^2(1-2h)\int_y^1 \exp(-h\alpha z-\tfrac{1}{2}\alpha z^2(1-2h)\,dz\ dy\ \Big] \qquad \ldots \quad (5.22)$$

where the terms under the outer integrals constitute an approximation for the pseudo-transient function. Similarly an approximation for the variance can be made. We now wish to compare these approximate values with those given by diffusion methods, since with s being $O(N^{-1})$ the drift and diffusion coefficients $m(x)$ and $v(x)$ are of the same order of magnitude. Using the transition probabilities (5.18), we have, in the notation of Chapter 3, when time is measured in units of N^2 birth-death events,

$$m(x) = \alpha x(1-x)\Big\{x + h(1-2x)\Big\}$$

$$v(x) = 2x(1-x)$$

Then using equations (3.9) and (3.10), the diffusion approximation to the probability $P(0,k)$ is given by

$$\frac{\displaystyle\int_p^1 \exp\Big[-h\alpha x - \tfrac{1}{2}\alpha x^2(1-2h)\Big]\,dx}{\displaystyle\int_0^1 \exp\Big[-h\alpha x - \tfrac{1}{2}\alpha x^2(1-2h)\Big]\,dx}$$

which is identical to the approximation $\bar{P}(0,k)$, given by equation (5.21), and is therefore close to the true value (5.19). Similarly, it may be shown that the diffusion approximation (c.f. equation (3.30)) for the mean absorption time is identical to the approximation (5.22) and hence close to the true value, and that the complete pseudo-transient function is similarly closely approximated. Further, the diffusion approximation to the variance is also close to the true value.

The above model generalizes a model considered by Moran (1963) where the selective advantages are 1, 1+d, 1. Moran has found the mean absorption time in this case by using the symmetry of these selective advantages and considering an associated process with a single absorbing boundary (at zero) and a reflecting boundary at $\frac{1}{2}N$. This allows the mean absorption time to be found by convolutions. The present method uses entirely different methods, since the symmetry property is absent.

5.4 Haploid populations with selection depending on gene frequency.

A second use of the methods given in section 5.2 is in the case of an overlapping generation haploid population of size N with selection depending on gene frequency. It is supposed that the two individuals A and a have selective advantages

$$1 + sw(x), \quad 1 \tag{5.23}$$

respectively, where x is the proportion of A individuals,

s is $O(N^{-1})$, and $w(x)$ is an arbitrary polynomial which is $O(1)$. Individuals are chosen at random to die and are replaced immediately by a new individual whose probability of being A is proportional to $x\{1 + s\, w(x)\}$ and whose probability of being a is proportional to $(1-x)$. Then if the number of A individuals is $i = xN$, then i is Markovian with transition matrix

$$\begin{aligned} p_{i,i-1} &= x(1-x)/\left\{1 + sx\, w(x)\right\} \\ p_{i,i+1} &= x(1-x)\left\{1 + s\, w(x)\right\}/\left\{1 + sx\, w(x)\right\} \\ p_{i,i} &= 1 - p_{i,i-1} - p_{i,i+1} \end{aligned} \tag{5.24}$$

Then using the values (5.24) in equation (5.14), the probability that the A individuals are eventually lost from the population, given initially k A individuals, is given by (5.14) with

$$\rho_i = \prod_{j=1}^{i} \left[1 + s\, w(j/N)\right]^{-1} \tag{5.25}$$

Now ρ_i may be approximated by

$$\begin{aligned} &\prod_{j=1}^{i} \exp\left[-s\, w(j/N)\right] \\ \approx\ &\exp\left[-\int_0^x \alpha\, w(y)\, dy\right] \end{aligned} \tag{5.26}$$

where $\alpha = Ns$ and $x = iN^{-1}$. Thus (5.14) is approximated by

$$\bar{P}(0,k) \qquad \frac{\int_p^1 \exp\left[-\alpha W(x)\right] dx}{\int_0^1 \exp\left[-\alpha W(x)\right] dx} \tag{5.27}$$

where $W(x)$ is the indefinite integral of $w(x)$ and $p = kN^{-1}$.

In order to find the diffusion approximation to $P(0,k)$, it is noted that (5.24) implies, in the notation of Chapter 3,

$$m(x) = \alpha x(1-x)\, w(x)$$

$$v(x) = 2x(1-x)$$

where time is measured in units of N^2 birth-death events. Therefore, using (3.9), the diffusion approximation to $P(0,k)$ is

$$\frac{\int_p^1 \exp\left[-\alpha W(x)\right] dx}{\int_0^1 \exp\left[-\alpha W(x)\right] dx}$$

which is identical to (5.27), so that the diffusion approximation is close to the true value given by (5.14) and (5.25).

Further, the mean time until homozygosity is given exactly by (5.15) and (5.25). Using the approximation (5.26) for ρ_i, it follows that this mean time is approximated by

$$N^2\Big[\bar{P}(0,k)\int_0^p x^{-1}(1-x)^{-1}\exp[\alpha W(x)]\int_0^x \exp[-\alpha W(y)]\,dy\,dx$$

$$+\,(1-\bar{P}(0,k))\int_p^1 x^{-1}(1-x)^{-1}\exp[\alpha W(x)]\int_x^1 \exp[\alpha W(y)]\,dy\,dx\Big] \qquad (5.28)$$

and using (3.30), this is found to be identical to the diffusion approximation, which is therefore close to the true value. Similarly the pseudo-transient function and the variance can be shown to be closely approximated by diffusion expressions.

Finally, it is interesting to note the various forms that (5.27) assumes in the important particular case when $w(x)$ is a linear function of x. We put $\alpha W(x) = \gamma x + \frac{1}{2}\beta x^2$, and note immediately that in the case $\gamma = \beta = 0$ both the diffusion approximation and the true probability both give $P(0,k) = 1-kN^{-1}$. It will in fact be more convenient for our purposes to consider the complimentary function $P(N,k) = 1 - P(0,k)$. Then for $\gamma = \beta = 0$, $P(N,k) = kN^{-1}$, and this value provides a standard against which the various values of

$$\bar{P}(N,k) = \frac{\displaystyle\int_0^p \exp\left[-\gamma y - \tfrac{1}{2}\beta y^2\right] dy}{\displaystyle\int_0^1 \exp\left[-\gamma y - \tfrac{1}{2}\beta y^2\right] dy} \qquad (5.29)$$

may be compared.

For $\beta = 0$, (5.29) gives

$$\bar{P}(N,k) = \frac{1 - \exp(-\gamma p)}{1 - \exp(-\gamma)}$$

which may be compared with the exact value found in Chapter 2.

When $\beta \neq 0$, it is useful to investigate the various forms that (5.29) takes for various γ.

(i) $\gamma = 0$

Equation (5.29) gives

$$\bar{P}(N,k) = \frac{\int_0^p \exp(-\frac{1}{2}\beta y^2)\, dy}{\int_0^1 \exp(-\frac{1}{2}\beta y^2)\, dy}$$

For $\beta > 0$, $\bar{P}(N,k) > kN^{-1}$ while for $\beta < 0$, $\bar{P}(N,k) < kN^{-1}$.

(ii) $\gamma = -\frac{1}{4}\beta$

This case corresponds to selective advantages

$$1 + t(x-\tfrac{1}{4}),\ 1$$

where $\beta = tN$. Equation (5.29) gives

$$\bar{P}(N,k) = \frac{\int_0^p \exp\left[\frac{1}{4}\beta y - \frac{1}{2}\beta y^2\right] dy}{\int_0^1 \exp\left[\frac{1}{4}\beta y - \frac{1}{2}\beta y^2\right] dy}$$

The nature of the curve of $\bar{P}(N,k)$ against p may be of different forms. Suppose initially that $\beta > 0$. Then

$$\left[\frac{d\bar{P}(N,k)}{dp}\right]_{p=1} < 1$$

so that in the neighbourhood of $p = 1$, $P(kN^{-1}) > kN^{-1}$. However

$$\left[\frac{d\bar{P}(N,k)}{dp}\right]_{p=0} = \frac{\exp(-\beta/32)}{\int_{-\frac{1}{4}}^{3/4} \exp(-\frac{1}{2}\beta y^2)\,dy}$$

and numerically it is found that the solution of the equation $\bar{P}'(0) = 1$ is approximately $\beta = 26.25$, and that for β less than this value $\bar{P}'(0)$ is greater than unity, while for β greater than this value $\bar{P}'(0)$ is less than unity. Thus for $\beta < 26.25$ the curve of $\bar{P}(N,k)$ lies entirely above the line $\bar{P}(N,k) = kN^{-1}$, while for $\beta > 26.25$ the curve is initially under this line, but subsequently crosses it and then remains above the line. Thus if the coefficient of the selective advantage is large and the initial value p is small to moderate, the large initial selective disadvantage has a strong effect (despite large selective advantages for large kN^{-1}). It is easily checked that the "crossing-point", i.e. the solution of $\bar{P}(N,k) = kN^{-1}$, lies in $(0,\frac{1}{4})$ for all $\beta > 26.25$. In the case $\beta < 0$ it also follows that the curve lies below the line $\bar{P}(N,k) = kN^{-1}$.

(iii) $\gamma = -\frac{1}{2}\beta$

This corresponds to selective advantages

$$1 + t(x-\tfrac{1}{2}),\ 1$$

and we find

$$\bar{P}(N,k) = \frac{\int_0^p \exp\left[\tfrac{1}{2}\beta y(1-y)\right] dy}{\int_0^1 \exp\left[\tfrac{1}{2}\beta y(1-y)\right] dy}$$

From symmetry, $\bar{P}(N,k) + \bar{P}(N,N-k) = 1$

and in particular $\bar{P}(N,\tfrac{1}{2}N) = \tfrac{1}{2}$.

For $\beta > 0$ the curve is initially below the line $\bar{P}(N,k) = kN^{-1}$, crosses the line at $\tfrac{1}{2}$, and then lies above the line, while for $\beta < 0$ the reverse is the case.

(iv) $\gamma = -\tfrac{3}{4}\beta$

Here the selective advantages are

$$1 + t(x - {}^3/_4) \ , \ 1$$

and we obtain

$$\bar{P}(N,k) = \frac{\int_0^p \exp\left[\tfrac{3}{4}\beta y - \tfrac{1}{2}\beta y^2\right] dy}{\int_0^1 \exp\left[\tfrac{3}{4}\beta y - \tfrac{1}{2}\beta y^2\right] dy}$$

Considering first the case $\beta > 0$, it is found that $\bar{P}'(0) < 1$, and $\bar{P}'(1) < 1$ for $\beta < 26.25$, $\bar{P}'(1) > 1$ for $\beta > 26.25$. Thus for $\beta < 26.25$ the curve lies entirely below the line $\bar{P}(N,k) = kN^{-1}$, for $\beta > 26.25$, it is initially less, but crosses over and is subsequently above the line. For $\beta < 0$ it may be checked that the

curve always lies above the line.

(v) $\gamma = -\beta$

It is easily verified that $\bar{P}(N,k) < kN^{-1}$ for negative β, but $\bar{P}(N,k) > kN^{-1}$ for positive β.

(vi)

When γ is of the same sign as β the nature of the curves is obvious.

It is interesting in the present case to try to find bounds for $P(0,k)$ which are simple functions, despite the fact that an exact expression for $P(0,k)$ is known. This is because the exact expressions are unwieldy, and also since the bounds will be of a form similar to the diffusion approximation, thus giving a measure of the error of the latter. As an example we consider the case $\gamma = 0$, $\beta > 0$. Then the diffusion approximation for $P(0,k)$ is

$$\bar{P}(0,k) = \left[\int_p^1 \exp(-\tfrac{1}{2}\beta y^2)\,dy\right] \div \left[\int_0^1 \exp(-\tfrac{1}{2}\beta y^2)\,dy\right]$$

It is simpler to consider the associated process for which $i \to i$ transitions are ignored. This will not affect $P(0,k)$ and gives as new transition probabilities

$$p_{i,i-1} = (2 + tx)^{-1}$$

$$p_{i,i+1} = (1 + tx)(2 + tx)^{-1}$$

where $x = iN^{-1}$ and $t = \beta N^{-1}$. Then in order to find a strict lower

bound for $P(0,k)$ it is necessary (c.f. Chapter 3) to find an increasing function $\phi^{**}(x_t)$ such that

$$E\left[\ \phi^{**}(x_{t+1}) - \phi^{**}(x_t)\,|\,x_t\ \right] \geq 0 \tag{5.30}$$

We try to find a $\phi^{**}(x_t)$ which is of the form

$$\phi^{**}(x_t) = \int_{-\infty}^{x_t} \exp(-\tfrac{1}{2}\beta y^2 + \epsilon y)\ dy \tag{5.31}$$

where ϵ is $O(N^{-1})$ and x_t is the value of x after the t^{th} transition. Then (5.30) and (5.31) give

$$E\int_{x_t}^{x_{t+1}} \exp(-\tfrac{1}{2}\beta y^2 + \epsilon y)\ dy \geq 0 \tag{5.32}$$

where all expectations are now to be conditional on x_t. Putting $z = y - x_t$, $\delta_{t+1} = x_{t+1} - x_t$, $x_t = x$, (5.32) gives

$$E\int_{0}^{\delta_{t+1}} \exp(-\tfrac{1}{2}\beta y^2 - \beta x z + \epsilon z)\ dz \geq 0 \tag{5.33}$$

If the integral in (5.33) is now written as $\psi(\delta_{t+1})$, then

$$\psi(0) = 0,\quad \psi'(\delta) = \exp(-\tfrac{1}{2}\beta\delta^2 - \beta x\delta + \epsilon\delta)$$

$$\psi''(\delta) = (-\beta\delta - \beta x + \epsilon)\exp(-\tfrac{1}{2}\beta\delta^2 - \beta x\delta + \epsilon\delta)$$

$$\psi'''(\lambda\delta) = \left[-\beta + (\beta\lambda\delta + \beta x - \epsilon)^2\right]\exp(-\tfrac{1}{2}\beta\lambda^2\delta^2 - \beta x\lambda\delta + \epsilon\lambda\delta)$$

for any λ in $(0,1)$.

Thus $\psi'(0) = 1$, $\psi''(0) \doteq \epsilon - \beta x$

$$|\psi'''(\lambda\delta)| \leq (\beta + 4\beta^2) \exp\left\{(1+\beta)N^{-1}\right\}$$

Hence

$$E\left[\psi(\delta)\right] = tx\, N^{-1}(2+tx)^{-1} + \tfrac{1}{2}(\epsilon - \beta x)N^{-2} + R \qquad (5.34)$$

where $R \leq \frac{1}{6}\, txN^{-3}\, (\beta + 4\beta^2) \exp\left\{(1+\beta)N^{-1}\right\} = R_1$ say.

Thus (5.34) is ≥ 0 if

$$\tfrac{1}{2}\, \epsilon N^{-2} \geq \tfrac{1}{2}\beta x N^{-2} - \tfrac{1}{2}\beta x N^{-2}(1+\tfrac{1}{2}tx)^{-1} + R_1$$

and this will be so for $\epsilon = \beta^2 N^{-1}$.

Similarly (5.34) ≤ 0 if $\epsilon = -\beta^2 N^{-1}$.

Thus using the results of Chapter 3, bounds for the exact value $P(0,k)$ are given by

$$\frac{\int_p^1 \exp(-\tfrac{1}{2}\beta y^2 + \beta^2 y N^{-1})\, dy}{\int_0^1 \exp(-\tfrac{1}{2}\beta y^2 + \beta^2 y N^{-1})\, dy} \leq P(0,k) \leq \frac{\int_p^1 \exp(-\tfrac{1}{2}\beta y^2 - \beta^2 y N^{-1})\, dy}{\int_0^1 \exp(-\tfrac{1}{2}\beta y^2 - \beta^2 y N^{-1})\, dy}$$

which may in fact be more convenient than the exact value.

5.5 Mutation and stationary distribution

If a small amount of mutation in both directions is allowed (at rates λ_1 for $A \to a$ and λ_2 for $a \to A$), then no absorption takes

place and a stationary distribution may be obtained. This will not be a pseudo-transient distribution and the diffusion approximation to it is given by the Wright equation

$$f(x) = \text{const } [v(x)]^{-1} \exp\left[2\int^{x} m(y)/v(y)\,dy \right] \qquad (5.35)$$

where the notation of Chapter 3 has been used. Note that this corresponds to equation (3.16) with $C_1 = 0$, i.e. no asymptotic probability flux. It is now shown that under the usual assumptions equation (5.35) provides a close approximation to the true stationary distribution in the case where the transition matrix is a continuant with $p_{i,i-1} = \Pi_i$ and $p_{i,i+1} = \eta_i$ and diffusion approximations are allowable. Then with mutation, the probabilities η_0, Π_N are positive. Also, if $m(x)$ and $v(x)$ are the drift on diffusion coefficients per birth-death event, then putting $i = Nx$, Π_i and η_i may be written

$$\Pi_i = \tfrac{1}{2}\left\{ N^2 v(x) - Nm(x) \right\}$$

$$\eta_i = \tfrac{1}{2}\left\{ N^2 v(x) + Nm(x) \right\}$$

Thus

$$\frac{\Pi_i}{\eta_i} = \frac{N^2 v(x) - Nm(x)}{N^2 v(x) + Nm(x)}$$

$$\approx \exp\left[-2m(x) \left\{ v(x) \right\}^{-1} N^{-1} \right] \qquad (5.36)$$

Now the stationary distribution $\underset{\sim}{\lambda}'$ satisfies

$$\underset{\sim}{\lambda}' \left[I - P \right] = \underset{\sim}{0}'$$

the typical equation being

$$-\lambda_{i-1}\,\eta_{i-1} + (\Pi_i + \eta_i)\,\lambda_i - \lambda_{i+1}\,\Pi_{i+1} = 0 \quad (i = 1,2,\ldots N-1)$$

giving

$$\lambda_i\,\Pi_i - \lambda_{i-1}\,\eta_{i-1} = \text{constant} \quad (i = 1,2,\ldots,N) \qquad (5.37)$$

It is easy to see, by using the particular case i = 1, that the constant in (5.37) is zero. This is in contrast to the allocation of the constant in (5.8) where a pseudo-transient distribution is considered, and in fact this leads to the interpretation of

$$\lambda_i\,\Pi_i - \lambda_{i-1}\,\Pi_{i-1}$$

as an asymptotic probability flux. This follows immediately by drawing an analogy between the two second order equations

$$\tfrac{1}{2}\frac{d^2}{dx^2}\left\{\, v(x)f(x) \,\right\} - \frac{d}{dx}\left\{\, m(x)f(x) \,\right\} = 0$$

and

$$-\lambda_{i-1}\,\eta_{i-1} + (\Pi_i+\eta_i)\lambda_i - \lambda_{i+1}\,\Pi_{i+1} = 0$$

The first is reduced to

$$\tfrac{1}{2}\frac{d}{dx}\left\{v(x)f(x)\right\} - m(x)f(x) = C_1$$

and the second to

$$\lambda_i\,\Pi_i - \lambda_{i-1}\,\eta_{i-1} = C_2.$$

If a true stationary distribution exists, then C_1 must be put equal to zero, and it has just been found that in the discrete case

C_2 must also be put equal to zero if a stationary distribution exists. On the other hand, if no non-trivial stationary distribution exists, then in the diffusion case C_1 has the interpretation of an asymptotic probability flux in the "return" process, so that by analogy a similar interpretation may be attached to C_2 in the "return" discrete process.

Returning now to the equation

$$\lambda_i \Pi_i - \lambda_{i-1} \eta_{i-1} = 0$$

we obtain

$$\lambda_i = \lambda_0 \left(\frac{\eta_0 \cdots \eta_{i-1}}{\Pi_1 \cdots \Pi_i} \right) \qquad i = 1,2,\ldots,N.$$

as the stationary vector of the process, where the constant λ_0 may be obtained by normalization. Using equation (5.36), λ_i may be approximated by

$$\bar{\lambda}_i = \frac{\text{const}}{\Pi_i} \exp \left[2 \sum_{j=0}^{i-1} \left\{ m(j/N) \right\} / \left\{ N\, v(j/N) \right\} \right]$$

$$\approx \frac{\text{const}}{v(x)} \exp \left[2 \int_0^x m(y)/v(y)\, dy \right] \qquad (5.38)$$

where $y = jN^{-1}$ and $x = iN^{-1}$. Equation (5.38) is identical to the Wright equation (5.35), which may therefore be taken as being a close approximation to the exact stationary distribution. It is, in fact, possible to use (5.36) to show similarly that when the assumptions necessary to use diffusion methods are satisfied, then the diffusion

approximation to the pseudo-transient distribution is close to the true distribution.

It follows easily from (5.38) that in the case where mutation is allowed, (at the rates specified above) the diffusion approximation to the stationary distribution is

$$f(x) = \text{const. } x^{\lambda_2 N-1} (1-x)^{\lambda_1 N-1} \exp\left\{ - \xi(x) \right\}$$

where $\xi(x)$ is that function for which

$$\bar{P}(N,Np) = \frac{\int_0^p \xi(x)\, dx}{\int_0^1 \xi(x)\, dx}$$

in the corresponding process with no mutation. Thus the previous discussion on the various forms $\bar{P}(N,Np)$ may take may be used in discussing $f(x)$. For example, if, in the case $\lambda_1 N = \lambda_2 N = 1$, the stationary distribution is greatest for small values of x, then in the process without mutation the probability of reaching the upper boundary $x = 1$ rather than $x = 0$ tends to be small, as would be expected.

5.6 Mutation in one direction only

The above methods may be used to derive exact pseudo-transient distributions in the case where mutation in one direction only is allowed. Consider, for example, the haploid population discussed by Moran (1962), p.132, with N individuals which are either A or a. If at any time the number of A individuals is i, then after the next birth-death event it will be i-1, i, or i+1, with probabilities given by

$$p_{i,i-1} = iN^{-1} q_i \quad = \Pi_i$$

$$p_{i,i+1} = (N-i)N^{-1} p_i = \eta_i \qquad (5.39)$$

$$p_{i,i} = 1 - p_{i,i-1} - p_{i,i+1}$$

where $p_i = i(1-\lambda)N^{-1}$

$$q_i = i\lambda N^{-1} + (N-i)N^{-1}.$$

These probabilities correspond to mutation at rate λ from A to a but no mutation from a to A. Thus it is certain that eventual elimination of A individuals will take place and the only question is how long this may be expected to take. The transition matrix of the Markovian variate i, the number of A individuals, is

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \Pi_1 & 1-\Pi_1-\eta_1 & \eta_1 & \cdots & 0 & 0 \\ 0 & \Pi_2 & 1-\Pi_2-\eta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \Pi_N & 1-\Pi_N \end{pmatrix}$$

Now this matrix is identical to the general form of a transition matrix considered in Chapter 9. We may therefore use the results (9.9) and (9.12) found in that chapter to state that the pseudo-transient function is given by

$$\lambda_i = \frac{1}{\Pi_i}\left[1 + \frac{\eta_{i-1}}{\Pi_{i-1}} + \frac{\eta_{i-1}\eta_{i-2}}{\Pi_{i-1}\Pi_{i-2}} + \ldots + \frac{\eta_{i-1}\eta_{i-2}\cdots\eta_1}{\Pi_{i-1}\Pi_{i-2}\cdots\Pi_1}\right] \qquad (5.40)$$

for i = 1, 2, ..., k

and

$$\lambda_i = \frac{1}{\Pi_i}\left[\frac{\eta_k\cdots\eta_{i-1}}{\Pi_k\cdots\Pi_{i-1}} + \frac{\eta_{k-1}\cdots\eta_{i-1}}{\Pi_{k-1}\cdots\Pi_{i-1}} + \ldots + \frac{\eta_1\cdots\eta_{i-1}}{\Pi_1\cdots\Pi_{i-1}}\right] \qquad (5.41)$$

$$\text{for } i = k+1,\ k+2,\ \ldots,\ N$$

where k is the initial number of A individuals. It will now be shown that (5.40) and (5.41) are closely approximated by the expressions derived from (3.40),(3.41), and (3.42). Using (5.39),

$$\frac{\eta_i}{\Pi_i} = \frac{(1-\lambda)(N-i)}{\lambda i + N - i}$$

and assuming that λ is $O(N^{-1})$, so that $\gamma = \lambda N$ is $O(1)$, then

$$\frac{\eta_i}{\Pi_i} \approx 1 - \lambda N(N-i)^{-1}$$

$$\approx \exp\left\{-\gamma(N-i)^{-1}\right\}$$

Therefore

$$\frac{\eta_{i-1}\cdots\eta_{i-\ell}}{\Pi_{i-1}\cdots\Pi_{i-\ell}} \approx \exp\left[-\gamma\left\{(N-i+1)^{-1} + \ldots + (N-i+\ell)^{-1}\right\}\right]$$

$$\approx \exp\left[-\gamma\int_1^{\ell}(N-i+n)^{-1}dn\right]$$

$$= \left[(N-i+1)(N-i+\ell)^{-1}\right]^{\gamma} \qquad (5.42)$$

Then putting $i = Nx$, $\Pi_i \approx x(1-x)$, so that for $1 \leqq i \leqq k$,

$$\lambda_i \approx x^{-1}(1-x)^{-1} \int_0^i \left[(N-i+1)(N-i+\ell)^{-1} \right]^{\gamma} d\ell$$

$$\approx Nx^{-1}(1-x)^{-1} \int_0^x \left[(1-x)(1-x+y)^{-1} \right]^{\gamma} dy$$

$$= Nx^{-1}(1-\gamma)^{-1} \left[(1-x)^{\gamma-1} - 1 \right] \qquad (5.43)$$

For comparison with (3.40), (3.41) and (3.42) it is necessary to put $a = 2$ in the latter, since in the above model the diffusion approximation to the variance is $2x(1-x)N^{-2}$, and similarly put $c = \gamma$. Then the constant d in (3.40) and (3.41) becomes $1-\gamma$ and the identity between (3.40) and (5.43) becomes obvious, when the different time scales are taken into account.

Similarly, for $i > k$, λ_i may be approximated by using (5.42) and (5.41). We have

$$\lambda_i \approx x^{-1}(1-x)^{-1} \int_{i-k}^{i-1} \left[(N-i+1)(N-i+\ell)^{-1} \right]^{\gamma} d\ell$$

(where again $i = Nx$). This leads eventually to

$$\lambda_i \approx Nx^{-1}(1-\gamma)^{-1}(1-x)^{\gamma-1} \left[1 - (1-p)^{1-\gamma} \right] \qquad (5.44)$$

where $k = Np$. By putting $a = 2$ and $c = \gamma$ in (3.41) and (3.42), the identity with (5.44) is clear. Thus the diffusion method gives a close approximation for the complete pseudo-transient function, and therefore the mean time until A individuals are lost from the population is well-approximated by (3.43).

CHAPTER 6.

SOME NUMERICAL AND DIFFUSION RESULTS

6.1 Introduction

In the overlapping generation models discussed in the previous two chapters, explicit expressions have been found for nearly all quantities of genetical interest (an exception being that in some cases it appears difficult to derive explicit expressions for latent roots of transition matrices).

In another genetical model (Wright, (1931)) all the members of a haploid population die simultaneously and are replaced by a new generation of the same size as the old. If states correspond to the number of A individuals in any generation, then in this model transition between all states in one generation is possible (except, of course, when mutation is absent, so that there are two absorbing states). This model is much harder to deal with than those for which transition to neighbouring states only is possible. In fact the only results known explicitly for this model occur when selection is not allowed. However it is easy to derive diffusion approximations for most quantities of interest, and it may be conjectured, after the previous two chapters, that the diffusion approximations will be reasonably accurate. In this chapter numerical results are given which support this conjecture but which present other problems.

The four quantities examined are (i) the probability that a given allele is eventually lost from the population; (ii) the mean time for elimination of one or other allele; (iii) the probability that a given allele is lost by the n^{th} generation;

(iv) the dominant non-unit latent root of the transition matrix.

The numerical results were obtained by using an electronic computor and are therefore subject to rounding error. Most of the results were obtained by successive powering of a transition matrix, and a check sum of the elements of each row for each powered matrix was obtained. This check sum differed from unity at the most in the sixth decimal place, so that the numerical results given here may be taken as being correct to the order in which they are given. In view of the size and speed of the computor the population size (twelve) considered is extremely small, but is in fact sufficient to draw useful conclusions.

It has been thought by some writers (Fisher)(1930), Kolmogorov (1959)) that diffusion methods break down near the boundaries within which the variate under consideration lies and that branching process techniques are necessary to examine the behaviour of the process near such points. This has led, amongst other things, to the necessity for "fusing" the branching process results with the diffusion results in the neighbourhood of the boundaries. However, it is difficult to see from the nature and derivation of the diffusion equation why diffusion results should not hold down to the boundaries and the numerical results obtained here suggest that in fact the diffusion results may be more accurate, in an absolute sense, near the boundaries than in the interior of the interval. Such a result has also been found numerically by

Knox (1962), some of whose results are abstracted in this chapter. In fact arguments at the end of the chapter suggest theoretically that in the case of the mean absorption time the diffusion results will be more accurate in an absolute sense near the boundaries, but on the other hand less accurate relatively than in the interior. This is observed here and in Knox's numerical results.

6.2 Absorption probabilities

We consider first the probability that the A individuals are eventually fixed in the population, given that they have selective advantage 1+s, where s is $O(N^{-1})$ and N is the population size. The transition matrix is

$$P = \left\{ p_{ij} \right\} = \left\{ \binom{N}{j} p_i^j \left(1-p_i\right)^{N-j} \right\}$$

where $p_i = (1+s)i(N+si)^{-1}$. Bounds for absorption probabilities in this model have been given in Chapter 2 (as case (3)). To obtain the numerical values the transition matrix is raised to the $1,2,4,\ldots,128^{th}$ power, by which time, for N = 12, the probability that both alleles A and a are still present is very small (of the order 10^{-6}). The diffusion approximation is obtained by putting

$$\left.\begin{aligned} m(x) &= \alpha x(1-x) \\ v(x) &= x(1-x) \end{aligned}\right\} \qquad (6.1)$$

where $\alpha = Ns$, in equation (3.8), with $P_0(p)$ replaced by $P_1(p)$ and with boundary conditions $P_1(0) = 0$, $P_1(1) = 1$. Thus the

diffusion approximation $P_1(p)$ to the probability of eventual fixation of A individuals in the population, given initial proportion p, is given by

$$P_1(p) = \frac{1 - \exp(-2\alpha p)}{1 - \exp(-2\alpha)} \qquad (6.2)$$

For s = 0, diffusion methods give $P_1(p) = p$, which is the exact value. By giving s various (positive) values we are able to compute a set of exact probabilities and diffusion approximations for the probability of fixation of A individuals, for various values of p. These are tabled below (Table 6.1).

It will be noted that the diffusion approximations are close to the true values and also always exceed the true values, as was shown would happen in Chapter 2. The bounds in Chapter 2 may be applied and are reasonably sharp; for instance in the case s = .04, 12p = 6, the bounds derived from (2.7) and (2.9) are .6134 and .6178, and the arithmetic mean of the bounds gives the exact value to the order of accuracy considered.

Table 6.1

Exact values and diffusion approximations (D.A.) for the probability of fixation of a gene having selective advantage 1 + s in a population of size 12, p being the initial proportion.

12p	Exact	D.A.	Error	Exact	D.A.	Error	Exact	D.A.	Error
	s = 0			s = .02			s = .04		
1	.0833	.0833	0	.1027	.1029	.0002	.1238	.1246	.0008
2	.1667	.1667	0	.2013	.2017	.0004	.2382	.2396	.0014
3	.2500	.2500	0	.2962	.2966	.0004	.3439	.3458	.0019
4	.3333	.3333	0	.3873	.3879	.0006	.4417	.4438	.0021
5	.4167	.4167	0	.4749	.4755	.0006	.5320	.5342	.0022
6	.5000	.5000	0	.5591	.5597	.0006	.6156	.6177	.0021
7	.5833	.5833	0	.6401	.6406	.0005	.6928	.6948	.0020
8	.6667	.6667	0	.7178	.7183	.0005	.7642	.7660	.0018
9	.7500	.7500	0	.7926	.7930	.0004	.8302	.8317	.0015
10	.8333	.8333	0	.8645	.8648	.0003	.8913	.8923	.0010
11	.9167	.9167	0	.9336	.9338	.0002	.9478	.9483	.0005
	s = .06			s = .08			s = .10		
1	.1463	.1482	.0019	.1699	.1733	.0034	.1940	.1994	.0054
2	.2764	.2796	.0032	.3153	.3209	.0056	.3539	.3626	.0087
3	.3921	.3962	.0041	.4397	.4467	.0070	.4858	.4962	.0104
4	.4951	.4996	.0045	.5463	.5539	.0076	.5946	.6056	.0110
5	.5866	.5913	.0047	.6376	.6453	.0077	.6844	.6952	.0108
6	.6681	.6726	.0045	.7159	.7231	.0072	.7586	.7685	.0099
7	.7406	.7447	.0043	.7830	.7895	.0065	.8200	.8286	.0086
8	.8051	.8087	.0036	.8406	.8460	.0054	.8707	.8777	.0070
9	.8626	.8655	.0029	.8900	.8942	.0042	.9127	.9180	.0053
10	.9138	.9158	.0020	.9324	.9352	.0028	.9474	.9509	.0035
11	.9594	.9604	.0010	.9688	.9702	.0014	.9762	.9779	.0017

6.3 Mean time for homozygosity

The vector of $\underset{\sim}{\mu}$ of mean times until homozygosity for various initial states, is given by

$$\underset{\sim}{\mu} = (I - Q)^{-1} \underline{\Psi} \tag{6.3}$$

where Q is the submatrix of P corresponding to transitions between transient states and $\underline{\Psi}$ is a column vector of unities. The diffusion approximation is found by solving (3.5) with the coefficients (6.1). The solution obtained is

$$U(p) = \frac{N}{\alpha}\left[\frac{e^{2\alpha(1-p)} - 1}{e^{2\alpha} - 1}\int_0^p \frac{e^{2\alpha y} - 1}{y(1-y)}\,dy\right.$$

$$\left. + \frac{e^{2\alpha} - e^{2\alpha(1-p)}}{e^{2\alpha} - 1}\int_p^1 \frac{e^{2\alpha} - e^{2\alpha y}}{e^{2\alpha}\, y(1-y)}\,dy\right] \tag{6.4}$$

where time has been rescaled to generations. The integrals have to be evaluated numerically by using Simpson's rule. By using (6.3) for exact results and (6.4) for the diffusion approximation, Table 6.2 below may be drawn up. Note that the mean times given are in terms of generations.

Table 6.2

Exact values and diffusion approximations (D.A.) for the mean time until homozygosity, p being the initial proportion of A genes and s the selective advantage.

12p	Exact	D.A.	Error	Exact	D.A.	Error	Exact	D.A.	Error
	s = 0			s = .02			s = .04		
1	6.147	6.884	.737	6.443	7.200	.757	6.715	7.490	.775
2	9.766	10.813	1.047	10.164	11.231	1.067	10.505	11.590	1.085
3	12.306	13.496	1.190	12.680	13.886	1.206	12.963	14.181	1.218
4	14.004	15.276	1.272	14.271	15.553	1.282	14.424	15.707	1.283
5	14.984	16.300	1.316	15.095	16.414	1.319	15.082	16.391	1.309
6	15.305	16.636	1.331	15.240	16.566	1.326	15.056	16.359	1.303
7	14.984	16.300	1.316	14.752	16.054	1.302	14.416	15.686	1.270
8	14.004	15.276	1.272	13.636	14.887	1.251	13.190	14.402	1.212
9	12.306	13.496	1.190	11.860	13.025	1.165	11.367	12.490	1.123
10	9.766	10.813	1.047	9.328	10.351	1.023	8.873	9.857	.984
11	6.147	6.884	.737	5.840	6.559	.719	5.535	6.226	.691
	s = .06			s = .08			s = .10		
1	6.959	7.752	.793	7.173	7.979	.806	7.356	8.171	.815
2	10.783	11.881	1.098	10.998	12.100	1.102	11.153	12.248	1.095
3	13.156	14.375	1.219	13.262	14.468	1.206	13.289	14.465	1.176
4	14.469	15.739	1.270	14.417	15.655	1.238	14.280	15.466	1.186
5	14.958	16.238	1.280	14.739	15.968	1.229	14.442	15.602	1.160
6	14.770	16.030	1.260	14.403	15.599	1.196	13.976	15.089	1.113
7	13.998	15.215	1.217	13.521	14.666	1.145	13.006	14.062	1.056
8	12.689	13.843	1.156	12.154	13.232	1.078	11.606	12.595	.989
9	10.847	11.912	1.065	10.318	11.312	.994	9.795	10.708	.913
10	8.415	9.347	.932	7.966	8.839	.873	7.535	8.343	.808
11	5.239	5.895	.656	4.958	5.575	.617	4.695	5.270	.575

In the first place it may be noted that the diffusion approximation always exceeds the true value. That this will always be so for sufficiently large N may be proved in the case s = 0 as follows. Firstly we note

Lemma 6.1

Let x be the proportion of successes observed in N Bernoulli trials, with constant probability p of success. Then

$$E\left\{ x \ln x + (1-x) \ln(1-x) \right\} \geq p \ln p + (1-p)\ln(1-p) + (2N)^{-1}$$

whenever $N^{-1} \leq p \leq (N-1)N^{-1}$ and $N \geq N_0$ (N_0 a suitable constant).

Proof

(a) It is sufficient to prove the lemma for $N^{-1} \leq p \leq \frac{1}{2}$, by symmetry, and it is therefore assumed that p lies in this range.

(b) If

$$\psi(x) = x \ln x + (1-x) \ln(1-x)$$

then the lemma asserts that

$$E\left\{ \psi(x) - \psi(p) \right\} \geq (2N)^{-1} \qquad (6.5)$$

Put $\delta = x - p$, the deviation of x from its expected value. Then the left-hand side in (6.5) becomes

$$E\left\{ \delta\psi'(p) + \tfrac{1}{2}\delta^2\psi''(p) + \tfrac{1}{6}\delta^3\psi'''(p) + \tfrac{1}{24}\delta^4\psi^{(iv)}(p) \right.$$

$$\left. + \tfrac{1}{120}\delta^5\psi^{(v)}(p) + \tfrac{1}{720}\delta^6\psi^{(vi)}(\theta x + (1-\theta)p) \right\}$$

where $\theta = \theta(p)$ and lies in $[0,1]$. Now from the definition of $\psi(p)$,

$$\psi''(p) = p^{-1}(1-p)^{-1}, \quad \psi'''(p) = p^{-2}(1-p)^{-2}(2p-1),$$

$$\psi^{(iv)}(p) = 2p^{-3}(1-p)^{-3}(1-3p+3p^2),$$

$$\psi^{(v)}(p) = 6p^{-4}(1-p)^{-4}(2p-1)(1-2p+2p^2), \quad \psi^{(vi)}(y) > 0 \text{ for } y \text{ in } (0,1)$$

Also, from the central moments of the binomial distribution,

$$E(\delta) = 0, \quad E(\delta^2) = p(1-p)N^{-1}, \quad E(\delta^3) = p(1-p)(1-2p)N^{-2},$$

$$E(\delta^4) = 3p^2(1-p)^2N^{-2} + p(1-p)(6p^2-6p+1)N^{-3},$$

$$E(\delta^5) = 10p^2(1-p)^2(1-2p)N^{-3} + p(1-p)(1-2p)(1-12p+12p^2)N^{-4}$$

and $E(\delta^6) > 0$. Thus the left-hand side in (6.5) may be written

$$(2N)^{-1} + (1-p+p^2)\left\{ 12p(1-p)N^2 \right\}^{-1}$$

$$+ \left\{ -5+27(p-p^2)-30(p-p^2)^2 \right\} \left\{ 12p^2(1-p)^2 N^3 \right\}^{-1}$$

$$+ \left\{ -(1-2p)^2(1-2p+2p^2)(1-12p+12p^2) \right\} \left\{ 20p^3 (1-p)^3 N^4 \right\}^{-1} + + ^{ve}.$$

Now for $0 \leq p \leq \frac{1}{2}$, $1-p+p^2 \geq 3/4$

$$-5+27(p-p^2) - 30(p-p^2)^2 \geq -5$$

and $$-(1-2p)^2(1-2p+2p^2)(1-12p+12p^2) \geq -1$$

Thus the left-hand side in (6.5) is greater than or equal to

$$(2N)^{-1} + \left\{ 16p(1-p)N^2 \right\}^{-1} - 5\left\{ 12p^2(1-p)^2 N^3 \right\}^{-1} - \left\{ 20p^3(1-p)^3 N^4 \right\}^{-1}$$

$$= (2N)^{-1} + \left\{ 16p(1-p)N^2 \right\}^{-1} \left[1 - 20\left\{ 3p(1-p)N \right\}^{-1} - 4\left\{ 5p^2(1-p)^2 N^2 \right\}^{-1} \right]$$

It is easy to show that the expression in square brackets is always positive for $N \geq 100$ and $p \geq 8N^{-1}$, so that the required inequality holds for these values. (6.6)

To show that the inequality holds for $p = N^{-1}$, $2N^{-1}, \ldots, 7N^{-1}$ for sufficiently large N, we note that since $(1-x)\ln(1-x)$ is a convex function of x, then

$$E\left\{ (1-x)\ln(1-x) \right\} \geq (1-p)\ln(1-p)$$

so that it is sufficient to prove

$$E\left\{ x \ln x \right\} \geq p \ln p + (2N)^{-1} .$$

Equivalently, it is sufficient to prove

$$E\left\{ y \ln y \right\} \geq \mu \ln \mu + \tfrac{1}{2}$$

where y is the number of successes in N Bernoulli trials with parameter μN^{-1}, $(\mu = 1,2,\ldots,7)$, and N is sufficiently large. Now it may be verified that if y is the value observed in a Poisson distribution with parameter μ, that the following table is true. We define $T(i;\mu)$ by

$$T(i;\mu) = \sum_{y=0}^{i} (y \ln y)\mu^y \exp(-\mu)/y!$$

Table 6.3

μ	i	$T(i;\mu)$	$\mu \ln \mu$	3rd col.- 4th col.
1	9	.5734	0	.5734
2	12	1.9560	1.3863	.5697
3	14	3.8431	3.2958	.5473
4	17	6.0774	5.5452	.5322
5	19	8.5705	8,0472	.5233
6	20	11.2685	10,7506	.5179
7	22	14.1358	13.5214	.5144

Thus $\sum_{y=0}^{22} (y \ln y)\, e^{-\mu}\mu^y/y!$ exceeds $\mu \ln \mu$ by at least .5144 for $\mu = 1,2,\ldots,7$. But there is only a finite number of terms in the above sum, so that since the terms in a binomial distribution with index N and parameter μN^{-1} converge to those of a Poisson distribution with parameter μ as $N \to \infty$, there exists an N_0' for which

$$\sum_{y=0}^{22} (y \ln y) \binom{N}{y} (\mu N^{-1})^y (1-\mu N^{-1})^{N-y} > \tfrac{1}{2} + \mu \ln \mu$$

for $\mu = 1,2,\ldots,7$ and $N \geq N_0'$.

Using this result and (6.6) the lemma has been proved for $N > \max(100, N_0') = N_0$. It may be noted also that numerical results suggest that the lemma is true for all positive integral N.

We may now use the lemma and the methods of section 3.6 to prove the assertion that the diffusion approximation always exceeds the true value for sufficiently large N. In the case

$s = 0$ (i.e. $m(x) = 0$), equation (3.54) yields

$$\phi(x) = -2\left\{ x \ell n\, x + (1-x)\ell n(1-x) \right\} \tag{6.7}$$

Now the proportion x_{t+1} of A individuals in the $(t+1)^{th}$ generation is obtained by binomial sampling from a distribution with parameter x_t. Thus the lemma shows that for N sufficiently large

$$E\left\{ \phi(x_{t+1}) - \phi(x_t) \right\} \geqq -N^{-1} \tag{6.8}$$.

By iterating in (6.8) until homozygosity is reached, we find, since $\phi(0) = \phi(1) = 0$, that

$$-\phi(p) \leqq -U(p)N^{-1}$$

or $U(p) \leqq \phi(p)$

$$= -2N\left\{ p \ell n\, p + (1-p)\ell n(1-p) \right\} \tag{6.9}$$

Since (6.9) is the diffusion approximation to the mean absorption time, the assertion that the diffusion approximation overestimates the true mean time for N sufficiently large is proved. Numerical results (c.f. Table 6.4) suggest that this assertion is true for all N.

In the paper (Ewens (1963 b)) in which the above results were first given, an attempt was made to explain the value of the discrepancy between the true value and the diffusion approximation along the following lines. Considering the analogous model discussed in Chapter 4, the exact expression (4.11) for the mean time until homozygosity is approximated by a formula (4.21) which

is close to the diffusion approximation (4.23). However (4.21) and (4.23) would be closer approximations to (4.11) if the terminals in the integrals were replaced by $\left((2N)^{-1},\ p\right)$ and $\left(p,\ (2N-1)(2N)^{-1}\right)$ respectively. If a similar change is made in (6.4) the value of the right-hand side would decrease, to a very close approximation, by

$$\frac{N}{\alpha}\left[\frac{e^{2\alpha(1-p)}-1}{e^{2\alpha}-1}\cdot\frac{2\alpha}{2N}+\frac{e^{2\alpha}-e^{2\alpha(1-p)}}{e^{2\alpha}-1}\cdot\frac{2\alpha}{2N}\right]=1$$

Since a discrepancy of about unity is observed in Table 6.2 it was conjectured that subtracting unity from the diffusion approximation would increase the accuracy of the approximation. This conjecture seems to be true in the "analysis of variance" sense that a significant proportion of the discrepancy could be removed by such a subtraction, but the following argument suggests that a more accurate statement about the discrepancy can be made.

The argument uses both the concept of the pseudo-transient distribution as well as the almost-invariant function given in Chapter 3. The equation

$$E\left\{\phi(x_{t+1})-\phi(x_t)\right\}\approx N^{-1} \qquad (6.10)$$

where $\phi(x)=2\left\{x\,\ell n\,x-(1-x)\,\ell n(1-x)\right\}$

was obtained by expanding $\phi(x_{t+1})$ about x_t in a Taylor series and ignoring terms which are $o(N^{-1})$. Iteration in (6.10) gives the diffusion approximation for the mean time. If the next term in

the Taylor series is included we obtain more accurately

$$E\left\{ \phi(x_{t+1})-\phi(x_t)\right\} = N^{-1}+(1-x_t+x_t^2)\left\{ 6x_t(1-x_t)N^2\right\}^{-1} + O(N^{-3}) \qquad (6.11)$$

Suppose that the mean number m_j of generations (before homozygosity) that the proportion of A individuals assumes the value jN^{-1} $(j = 1,2,\ldots,N-1)$ is known or can be approximated. Then iteration in (6.11) gives

$$-\phi(p) \approx \mu(k)N^{-1} + \sum_{j=1}^{N-1} m_j(1-jN^{-1}+j^2N^{-2}) \left\{ 6j(N-j) \right\}^{-1}$$

so that a more accurate formula than

$$\mu(k) \approx -2N\left\{ p\, \ell n\, p + (1-p)\, \ell n(1-p) \right\} \qquad (6.12)$$

is

$$\mu(k) \approx -2N\left\{p\, \ell n\, p + (1-p)\, \ell n(1-p)\right\} - N^{-1}\sum_{j=1}^{N-1} m_j(1-jN^{-1}+j^2N^{-2})\left\{6j(N-j)\right\}^{-1}$$

$$\ldots \qquad (6.13)$$

Now using pseudo-transient function methods it is possible to find close approximations for m_j. In fact since $p = kN^{-1}$ the diffusion approximations for the m_j are

$$\left.\begin{aligned} m_j &= 2(N-k)(N-j)^{-1} && j = 1,2,..,k \\ m_j &= 2kj^{-1} && j = k+1,..,N-1 \end{aligned}\right\} \qquad (6.14)$$

Using (6.13) and (6.14) we obtain as a more accurate formula than (6.12) the equation

$$\mu(k) \approx -2N\left\{ p \ln p + (1-p) \ln(1-p) \right\}$$

$$-(3N)^{-1}\left[\sum_{j=1}^{k} (N-k)(N^2-Nj+j^2)(N-j)^{-2}j^{-1} \right.$$

$$\left. + \sum_{j=k+1}^{N-1} k(N^2-Nj+j^2)j^{-2}(N-j)^{-1} \right] \qquad (6.15)$$

It is interesting to examine this more accurate formula in the cases $k = 1$, $k = \frac{1}{2}N$. For $k = 1$ it is found that (6.15) reduces approximately to

$$\mu(1) \approx -2\left\{ \ln N^{-1} + (N-1) \ln(1-N^{-1}) \right\} - \pi^2/18 - (3N)^{-1}\ln(N-1) - .1924N^{-1}$$

... (6.16)

The first term on the right-hand side of (6.16) is the diffusion approximation, so that (6.16) suggests that asymptotically the diffusion approximation overestimates the true mean time by $\pi^2/18 \approx .55$, and further that the absolute value of the error of the diffusion approximation decreases as N increases. Also the relative error is approximately $\pi^2/(36 \ln N)$ which decreases very slowly with N.

In the case $k = \frac{1}{2}N$, (6.15) becomes, after some reduction

$$\mu(\tfrac{1}{2}N) \approx -2N \ln \tfrac{1}{2} - \tfrac{1}{3} \ln N - \tfrac{1}{3}(1 - \ln 2) \qquad (6.17)$$

Since the first term on the right-hand side of (6.17) is the diffusion approximation for $\mu(\frac{1}{2}N)$, (6.17) suggests that the diffusion approximation overestimates the true mean time by a term whose leading element is $\frac{1}{3}\ln N$, and which consequently increases with N,

(in contrast to the case k = 1). Further, the relative error is approximately

$$(\ell n\ N)(6N\ \ell n\ 2)^{-1}$$

These results may now be compared with numerical values found by Knox (1962), who has obtained exact results for N = 10, 20, 30, 40, 50, using a high-speed computor. The relevant values given by Knox are as follows.

Table 6.4

N	True value	Diffusion Approx.	Error	Rel.Error %
		k=1		
10	5.75328	6.50166	.74838	13.01
20	7.23122	7.94061	.70939	9.81
30	8.07599	8.76868	.69269	8.58
40	8.66930	9.35255	.68325	7.88
50	9.12677	9.80391	.67714	7.42
		$k=\frac{1}{2}N$		
10	12.5905	13.8629	1.2724	10.11
20	26.2295	27.7259	1.4964	5.70
30	39.9595	41.5888	1.6293	4.08
40	53.7280	55.4518	1.7238	3.21
50	67.5169	69.3147	1.7978	2.66

It may be noted that all the results suggested by the above theory are verified. For k = 1 the absolute error decreases and appears to approach a limit between .5 and .6. The relative error decreases slowly, as suggested. For $k = \frac{1}{2}N$ the absolute error increases at a rate roughly proportional to $\ell n\ N$, as suggested, while the relative error decreases at a rate approximately $(\ell n\ N)N^{-1}$.

The above arguments and numerical values make it plausible that for all N the diffusion approximation is most accurate near the boundaries, and hence it is unnecessary, as mentioned at the beginning of the chapter, to use branching process methods to estimate the mean absorption time when the initial value k is close to 0 or close to N.

6.4 Probability of absorption by the n^{th} generation

The probability that by the n^{th} generation all the individuals in the population are A can be obtained numerically by examining the n^{th} power of the transition matrix. For the diffusion approximation attention is restricted to the case s = 0, since the diffusion approximations for non-zero s are very complicated. In the case s = 0 Kimura (1955a) has found the diffusion approximation for the probability that the A individuals have become fixed in a population of size N by the $(Nn)^{th}$ generation, given initial proportion p. His expression is

$$\text{Prob}\left\{\text{ fixation by } (Nx)^{th} \text{ generation }\right\}$$

$$= p + \sum_{i=1}^{\infty} (-1)^i \frac{2(2i+1)p(1-p)}{i(i+1)} T_{i-1}^{1}(1-2p) \exp\left\{-\tfrac{1}{2}i(i+1)n\right\} \qquad (6.18)$$

where $T_{i-1}^{1}(z)$ is a Gegenbauer polynomial defined by

$$T_{i-1}^{1}(z) = \tfrac{1}{2}i(i+1)\, F(i+2,\ 1-i,\ 2,\ \tfrac{1}{2}(1-z)) .$$

To make a comparison with the numerical results it is necessary to put N = 12 and n = j/12 (j = 1,2,4,...,128). If this is done it is

found that the diffusion approximation becomes more accurate as j increases. For j = 128 the approximation is very close to the true value, due to the fact that the right-hand side in (6.18) is very close to p, as is the true probability. For j less than eight the diffusion approximation is poor. It may be noted from Table 6.5 below that the diffusion approximation underestimates the true probabilities for all values of p and j considered. This corresponds to the fact that the diffusion approximation overestimates the mean time to homozygosity.

6.5 Latent roots

For s = 0 the latent roots are known (Feller (1951)) to be

$$\lambda_r = \binom{N}{r} r!\, N^{-r} \qquad (r = 0,1,\ldots,N) \tag{6.19}$$

so that the largest non-unit latent root is $1-N^{-1}$ and is 11/12 in our case. For $s \neq 0$ the latent roots are unknown, and the largest non-unit latent root, which determines asymptotically the rate of approach to homozygosity, may be estimated as follows. By writing out the matrix P in spectral form we obtain

$$P^n = C + \sum_{i=2}^{N} \lambda_i^n D_i \qquad 1 > |\lambda_2| > \ldots > |\lambda_N| \tag{6.20}$$

where the D_i are spectral matrices corresponding to the λ_i ($\equiv\lambda_i(s)$) and C is a matrix having positive elements only in the first and last columns. These entries are elimination and fixation

Table 6.5

Exact values and diffusion approximations (D.A.) for the probability that the A genes have become fixed in a population of size 12 by the j^{th} generation, given initial proportion of A genes = p.

12p	j = 8			j = 16		
	Exact	D.A.	Error	Exact	D.A.	Error
1	.0058	.0030	-.0028	.0351	.0286	-.0065
2	.0175	.0100	-.0075	.0778	.0652	-.0126
3	.0371	.0228	-.0143	.1284	.1103	-.0181
4	.0665	.0442	-.0223	.1874	.1644	-.0230
5	.1083	.0761	-.0321	.2550	.2283	-.0267
6	.1649	.1230	-.0419	.3316	.3024	-.0292
7	.2394	.1886	-.0508	.4175	.3875	-.0300
8	.3348	.2776	-.0572	.5132	.4842	-.0290
9	.4547	.3966	-.0581	.6189	.5931	-.0258
10	.6027	.5513	-.0514	.7350	.7149	-.0201
11	.7830	.7495	-.0335	.8619	.8503	-.0116
	j = 32			j = 64		
1	.0706	.0675	-.0031	.0825	.0822	-.0003
2	.1435	.1379	-.0056	.1652	.1647	-.0005
3	.2187	.2111	-.0076	.2481	.2473	-.0008
4	.2962	.2871	-.0091	.3310	.3301	-.0009
5	.3760	.3661	-.0099	.4142	.4131	-.0011
6	.4581	.4479	-.0102	.4974	.4964	-.0010
7	.5426	.5326	-.0100	.5808	.5798	-.0010
8	.6294	.6202	-.0092	.6644	.6634	-.0010
9	.7185	.7108	-.0077	.7481	.7473	-.0008
10	.8100	.8043	-.0057	.8319	.8313	-.0006
11	.9038	.9006	-.0032	.9159	.9156	-.0003

probabilities for various initial values of the number of A individuals and their sum by rows is unity. For large n we have

$$P^n \approx C + \lambda_2^n D_2$$

If $\underset{\sim}{\xi} = (100..01)'$ then $C\,\underset{\sim}{\xi} = (111...1)' = \underset{\sim}{\psi}$

and also $D_2\,\underset{\sim}{\xi}' = \underline{d} = (d_0\ d_1\ \ldots\ d_N)'$ say. Thus

$$P^n\,\underset{\sim}{\xi} \approx \underset{\sim}{\psi} - \lambda_2^n\,\underline{d}$$

Letting $\underset{\sim}{\eta}_j = (00..010..0)'$, where the 1 occurs in the j^{th} position $(j = 0,1,2,..,N)$ we obtain

$$\underset{\sim}{\eta}_j'\,(\underset{\sim}{\psi} - P^n\,\underset{\sim}{\xi}) \approx \lambda_2^n\,d_j = h(n,j) \text{ say} \qquad (6.21)$$

Thus an estimator of λ_2 $(\equiv\lambda_2(s))$ is

$$\hat{\lambda}_2(s) = \left[\, h(m,j)/h(n,j) \,\right]^{\frac{1}{m-n}} \qquad (6.22)$$

Ideally, values of m and n as large as possible would be used to estimate $\lambda_2(s)$, since the effect of the other latent roots will then be relatively smaller. In the numerical example considered the best values for m and n are found to be 64 and 32 respectively, since if the value 128 is used for m the expression (6.21) becomes very small and of the same order of magnitude as the rounding error in the computor. The values 64 and 32 are sufficiently large

to make the effects of the other latent roots negligible but not so large that the rounding error becomes important.

By varying j in (6.22) a set of estimators may be found for $\lambda_2(s)$ for each s. The results are summarized in Table 6.6 below. In the table the arithmetic mean $\bar{\lambda}_2(s)$ of the estimates and the upper and lower bounds attained for various j are given.

Table 6.6

Arithmetic mean $\bar{\lambda}_2(s)$ and bounds obtained from (5.3).

s	0	.02	.04	.06	.08	.10
$\bar{\lambda}_2(s)$	.91664	.91618	.91490	.91283	.91001	.90652
Range Lower	.916625	.916157	.914855	.912771	.909949	.906437
Range Upper	.916647	.916206	.914935	.912862	.910040	.906562

As a check, it is known that $\lambda_2(0) = .916667$ so that the error in $\bar{\lambda}_2(s)$ is in the fifth significant figure, so that errors in $\bar{\lambda}_2(s)$ for non-zero s may also be expected to be of this order.

The values in Table 6.6 may be used to estimate the relation between $\lambda_2(s)$ and s. If $\bar{\lambda}_2(s)$ is regressed jointly on s and s^2 the sum of squares removed by the regression on s is small. Similarly regressing $\bar{\lambda}_2(s)$ on s^2 and s^4 yields a small sum of squares for s^4. If $\bar{\lambda}_2(s)$ is regressed on s^2 alone (the regression being constrained to pass through (0,11/12)) the estimated curve is

$$\bar{\lambda}_2(s) = 11/12 - 1.028s^2 \qquad (6.23)$$

and this curve gives a very close fit to the observed values.

To derive for comparison the diffusion approximation for $\lambda_2(s)$ it is necessary to solve equation (3.1), when $m(x)$ and $v(x)$ are given by (6.1), for the distribution $f(x;t)$ in the interval $(0,1)$ of the proportion x of A individuals at time $t(= tN$ generations). This has been done by Kimura (1955b) who obtains the solution

$$f(x;t) = \sum_{i=0}^{\infty} c_i \exp(-\lambda_i t)\, V_{1i}(z) \exp(-\tfrac{1}{2}sz) \qquad (6.24)$$

where $z = 1-2x$, the $V_{1i}(z)$ are linear sums of Gegenbauer polynomials, the c_i are constants and the λ_i are certain (latent root) constants. For our purposes only the λ_i are of interest and in particular the smallest (absolute) latent root λ_0. Clearly the probability that the proportion x lies in the open interval decreases asymptotically at the rate $\exp(-\lambda_0)$ per N generations on $\exp(-\lambda_0/N)$ per generation. Kimura showed that this smallest latent root is given by

$$\lambda_0 = 1 + \frac{N^2s^2}{10} - \frac{N^4s^4}{7000} - \frac{N^6s^6}{1,050,000} - \cdots.$$

so that $\exp(-\lambda_0/N)$ is approximately

$$(N-1)N^{-1} - Ns^2/10 \qquad (6.25)$$

for the values we are considering. In the case N=12, (6.25) becomes

$$11/12 - 1.2s^2 \qquad (6.26)$$

which agrees reasonably well with the numerically-derived result (6.23).

It is possible to find explicitly the largest non-unit latent root in the cases N = 2, N = 3. If terms of order s^4 are ignored, these are

$$N = 2 : \lambda_2(s) = \tfrac{1}{2} - s^2/8$$

$$N = 3 : \lambda_2(s) = \tfrac{2}{3} - 2s^2/9$$

The diffusion approximations for these values are

$$N = 2 : \lambda_2(s) = \tfrac{1}{2} - s^2/5$$

$$N = 3 : \lambda_2(s) = \tfrac{2}{3} - 3s^2/10$$

These are used to draw up Table 6.7 below, which gives the proportionate error of the diffusion approximation for the coefficient of s^2 compared with the true value. For this purpose the numerical result (6.23) is considered to be exact. Clearly the diffusion approximations are always too large (in absolute value), but the proportionate error decreases steadily. For N greater than about 50 the diffusion approximation should be reasonably accurate.

Table 6.7

True value and diffusion approximation for coefficient of $-s^2$ in $\lambda_2(s)$, with proportionate error of diffusion approximation.

N	2	3	....	12
True Value	.1250	.2222	....	1.028
Diffusion Approximation	.2000	.3000	....	1.200
Proportionate Error of Diffusion Approximation	60.00%	35.00%	....	16.73%

CHAPTER 7

DIPLOID POPULATIONS

7.1 Introduction

The next two chapters consider monoecious diploid populations, which admit selection, as for haploid populations, but also dominance, which is meaningless for haploid populations (except for artificial examples similar to that considered in Chapter 5, section 2). The population size is fixed at N, giving 2N alleles at the locus under consideration. The possible genotypes are AA, Aa, and aa, with selective advantages 1+s, 1+sh, and 1 respectively. Thus s is a measure of selection (which for our purposes is assumed $O(N^{-1})$ and h is a measure of dominance. If $h = \frac{1}{2}$ the selective advantage of the heterozygote is intermediate between those of the two homozygotes and there is no dominance.

If at the t^{th} generation the numbers of AA, Aa, and aa individuals are k_t, $N-k_t-\ell_t$, and ℓ_t respectively, then if the composition of the $(t+1)^{th}$ generation is determined only by that of the t^{th} generation, the pair (k_t, ℓ_t) will be Markovian, so that in theory a transition matrix could be set up and all results derived from it. In practice this is not possible and all that can be done is to make further assumptions and/or to use diffusion methods.

It has been shown by Watterson (1962) that diffusion methods may be used if a variate can be found satisfying certain quasi-Markovian conditions, but it is necessary for the purposes of finding bounds along the lines indicated in Chapter 3 that

that variate under consideration be strictly Markovian. For any overlapping-generation model analogous to those considered in Chapters 4 and 5, a single Markovian variate to replace the Markovian pair (k_t, ℓ_t) could be found only by making further assumptions. One such assumption would be that the numbers of the three genotypes at any time t occur in the Hardy-Weinberg proportions p_t^2, $2p_tq_t$, and q_t^2 say. In this case p_t would be Markovian and diffusion methods could be used. However it is in general impossible that the Hardy-Weinberg relations should hold both before and after a birth-death event, and in any case it would be preferable to use a model which does not necessitate such a restriction. In this chapter it is shown that in the non-overlapping generation model analogous to that of Chapter 6 it is possible to find a single Markovian variate, which is a function of k_t and ℓ_t, which allows diffusion approximations and strict bounds to be made for various functions of interest.

7.2 Absorption Probabilities

The model considered (c.f. Moran (1958b)) is analogous to that considered in Chapter 6. With the selective advantages given above the relative outputs of the three genotypes are $k_t(1+s)$, $(N-k_t-\ell_t)(1+sh)$, and ℓ_t respectively, so that the expected proportion of A genes will be

$$p(A) = \frac{(1+s)k_t + \frac{1}{2}(1+sh)(N-k_t-\ell_t)}{(1+s)k_t + (1+sh)(N-k_t-\ell_t) + \ell_t} \qquad (7.1)$$

$$= \frac{(1+s)a_t + \frac{1}{2}(1+sh)(1-a_t-b_t)}{1 + sa_t + sh(1-a_t-b_t)} \qquad (7.2)$$

where $a_t = k_t N^{-1}$, $b_t = \ell_t N^{-1}$.

It is supposed that each of the N offspring will be AA, Aa, or aa with respective probabilities $p^2(A)$, $2p(A)\{1-p(A)\}$, $\{1-p(A)\}^2$. It follows that the number of A genes in the next generation is a binomial variate with index 2N and parameter p(A).

The next step is to find a single Markovian variate describing the process. Watterson (1962) in considering an analogous problem for dioecious populations chose a "quasi-Markovian" variate which in the present case would reduce to $z_t = (2N)^{-1}(N+k_t-\ell_t)$. This is the proportion of A genes in the t^{th} generation but is not a Markovian variate. This occurs because it is necessary to know not only how many A genes there are but also what proportion of these are in homozygous (AA) individuals, since the selective advantage of these differs from that of the heterozygotes. Thus z_t does not provide complete information and it is more relevant to consider a weighted proportion of A genes, the weights being proportional to the selective advantages. This leads to the consideration of the variate

$$x_t = \frac{(1+s)a_t + \frac{1}{2}(1+sh)(1-a_t-b_t)}{1+sa_t+sh(1-a_t-b_t)} \qquad (7.3)$$

which is identical to p(A) above, and which is seen then to be

Markovian. The difference between z_t and x_t is of order s and is therefore small; it is this fact which allows z_t to be treated as a "quasi-Markovian" variate rather than x_t.

Having found a Markovian variate it is possible to use diffusion methods to approximate to the probability that the A genes eventually become fixed in the population. By expanding the denominator in (7.3) in series and ignoring for the moment terms of order s^2 we obtain

$$E(x_{t+1}) = E\tfrac{1}{2}(1+a_{t+1}-b_{t+1}) + \tfrac{1}{2}sE\Big\{ a_{t+1} - a_{t+1}^2 + a_{t+1}\, b_{t+1} + \\ + (b_{t+1} - a_{t+1})\, h(1 - a_{t+1} - b_{t+1}) \Big\}$$

$$= x_t + \tfrac{1}{2}x\left[x_t^2 - x_t^4 + x_t^2(1-x_t)^2 + \left\{(1-x_t)^2 - x_t^2\right\} h\left\{1 - x_t^2 - (1-x_t)^2\right\} \right]$$

Therefore if $\delta_{t+1} = x_{t+1} - x_t$, to order s

$$E(\delta_{t+1}) = sx_t(1-x_t)\left\{h + (1-2h)\, x_t\right\} \qquad (7.4)$$

and also

$$\mathrm{Var}(\delta_{t+1}) = (2N)^{-1}\, x_t(1-x_t) \qquad (7.5)$$

to the same order of accuracy.

Equations (7.4), (7.5), (3.9) and (3.10) give the formula

$$\bar{P}(N, Nx_0) = \frac{\displaystyle\int_0^{x_0} \exp\left[-2\alpha h t + \alpha D t^2 \right] dt}{\displaystyle\int_0^{1} \exp\left[-2\alpha h t + \alpha D t^2 \right] dt} \qquad (7.6)$$

as the diffusion approximation for the probability of eventual fixation of A genes, where $\alpha = 2Ns$, $D = 2h-1$, and x_0 the initial value of x_t. Note that in general x_0 is not the initial proportion of A genes but differs from this proportion by a term of order s. Equation (7.6) has been derived elsewhere (Watterson (1962), Kimura (1957)), except that in these cases x_0 is to be interpreted as the initial proportion of A genes.

Similarly the diffusion approximation to the mean absorption time and to the pseudo- transient distribution may be obtained directly by using (7.4) and (7.5) in equations (3.14) and (3.27).

It may be noted that in the case $h = \frac{1}{2}$, (no dominance), equation (7.6) reduces to

$$\bar{P}(N, Nx_0) = \frac{1 - \exp(-\alpha x_0)}{1 - \exp(-\alpha)}$$

which is the same as equation (6.2) with α replaced by 2α. This replacement corresponds to the fact that with the present model a diploid population with selective advantages $1+s$, $1+\frac{1}{2}s$, 1 acts in the same way as a haploid population with selective advantages $1+\frac{1}{2}s$, 1. Equation (7.6) also corresponds, in the case $h=\frac{1}{2}$, to the value obtained in diploid populations with gametic selection and selective advantages $1+s$, 1, but it should be noted that while this is so it does not follow that a population with gametic selection is equivalent to a population with zygotic selection with the heterozygote exactly intermediate in selective

advantage between the two homozygotes. This is most easily seen in the case where in any generation all individuals are heterozygotes; in the former selection still exists but in the latter the two alleles have equal selective advantages.

7.3 Bounds

The method outlined in Chapter 3 may be used to find bounds for the true fixation probability approximated by (7.6). Now (7.6) is essentially derived by noting that the function

$$\int^{x_t} \exp\left\{ -2\alpha ht + \alpha Dt^2 \right\} dt$$

possess the "almost-invariant" property

$$E\left[\int^{x_{t+1}} \exp\left\{-2\alpha ht + \alpha Dt^2\right\} dt \right] \approx \int^{x_t} \exp\left\{-2\alpha ht + \alpha Dt^2 \right\} dt \qquad (7.7)$$

where expectations are conditional on x_t. Thus in order to find bounds we seek a function of the form

$$\int^{x_t} \exp\left\{-2\alpha ht + vN^{-1} + \alpha Dt^2 \right\} dt$$

where v is O(1), for which

$$E\left[\int^{x_{t+1}} \exp\left\{-2\alpha ht+vN^{-1}t+\alpha Dt^2\right\} dt \right] \leq \int^{x_t} \exp\left\{-2\alpha ht+vN^{-1}t+\alpha Dt^2\right\} dt$$

or

$$E\left[\int_{x_t}^{x_{t+1}} \exp\left\{-2\alpha ht+vN^{-1}t+\alpha Dt^2 \right\} dt \leq 0 \right. \qquad (7.8)$$

Then by iteration it follows, using (3.51), that the true fixation

probability $P(N,Nx_0)$ satisfies

$$P(N,Nx_0) \leqq \frac{\int_0^{x_0} \exp\left\{-2\alpha ht+vN^{-1}t+\alpha Dt^2\right\} dt}{\int_0^{1} \exp\left\{-2\alpha ht+vN^{-1}t+\alpha Dt^2\right\} dt} \tag{7.9}$$

Now equation (7.8) may be rewritten

$$E\left[\int_0^{\delta_{t+1}} \exp\left\{-2\alpha hu+N^{-1}vu+\alpha Du^2+2\alpha Dux_t\right\} du \leqq 0\right. \tag{7.10}$$

or

$$E\left\{\theta(\delta_{t+1})\right\} \leqq 0 \tag{7.11}$$

Putting $c = 2\alpha Dx_i - 2\alpha h + N^{-1}v$, we get

$$\theta(0) = 0,\ \theta'(0) = 1,\ \theta''(0) = c,\ \theta'''(0) = 2\alpha D+c^2$$

$$\theta^{(iv)}(\lambda\delta) = \left\{6\alpha D(2\alpha D\lambda\delta+c) + (2\alpha D\lambda\delta+c)^3\right\} \exp(\alpha D\lambda^2\delta^2+c\lambda\delta)$$

Since v is $O(1)$ we have

$$|\theta'''(0)| \leqq 4\alpha+8\alpha^2,\ |\theta^{(iv)}(\lambda\delta)| \leqq (24\alpha^2+72\alpha^3)\exp(3\alpha) \tag{7.12}$$

Now

$$E(x_{t+1}) = \tfrac{1}{2}E(1+a_{t+1}-b_{t+1}) + \tfrac{1}{2}sE\left\{a_{t+1}-a_{t+1}^2+a_{t+1}b_{t+1}+\right.$$

$$\left.+(b_{t+1}-a_{t+1})h(1-a_{t+1}-b_{t+1})\right\} + s^2E\left\{\xi(a_{t+1},b_{t+1})\right\}$$

... (7.13)

where the function $E(\xi)$ is rational in x_t, bounded in absolute

value by 2, and vanishes at $x_t = 0,1$. Thus

$$E(\delta_{t+1}) = sx_t(1-x_t)\left\{h + (1-2h)x_t\right\} + R_1\, x_t(1-x_t)$$

$$\text{where } |R_1| \leq 8s^2 + 2sN^{-1}$$

the latter terms in the bound for R_1 arising from covariance terms in the second factor on the right-hand side of (7.13). Similarly

$$\text{Var}(\delta_{t+1}) = x_t(1-x_t)(2N)^{-1} + R_2\, x_t(1-x_t)$$

$$\text{where } |R_2| \leq 2sN^{-1}.$$

Also the third and fourth moments of δ_{t+1} about its mean are bounded in absolute value by $x_t(1-x_t)(2N^2)^{-1}$ and $x_t(1-x_t)(4N^2)^{-1}$ respectively, and the fourth moment of δ_{t+1} about zero is also bounded by the latter bound. Therefore

$$|E(\delta_{t+1}^2)| \leq \text{var}(\delta_{t+1}) + \alpha^2\, x_t(1-x_t)(8N^2)^{-1},$$

$$|E(\delta_{t+1}^3)| \leq x_t(1-x_t)(2N^2)^{-1} + \alpha x_t(1-x_t)(4N^2)^{-1}.$$

Thus when $\theta(\delta_{t+1})$ is expanded in a truncated Taylor series as far as the fourth derivative term and expectations taken, with the above bounds being used, then it is found that the last two terms in the expansion are bounded by

$$x_t(1-x_t)(4N^2)^{-1}\left\{ 2\alpha+4\alpha^2+\tfrac{4}{3}\alpha^3+(\alpha^2+3\alpha^3)\exp(3\alpha) \right\}$$

Similarly the first two terms are greater than

$$x_t(1-x_t)(4N^2)^{-1}\left\{ \tfrac{1}{2}v - \tfrac{1}{2}\alpha^3 - 4\alpha - 12\alpha^2 \right\}$$

Therefore the right-hand side in (7.11) will be positive if we put

$$v = v^* = 2\left\{ 6\alpha + 16\alpha^2 + 2\alpha^3 + (\alpha^2+3\alpha^3)\exp(3\alpha)\right\}$$

and will be negative if $v = -v^*$. Thus by using equation (7.9) and a similar equation for a lower bound we have

$$\frac{\int_0^{x_0} \exp\left\{-2\alpha ht+v^*N^{-1}+\alpha Dt^2\right\} dt}{\int_0^1 \exp\left\{-2\alpha ht+v^*N^{-1}+\alpha Dt^2\right\} dt} \leqq P(N,Nx_0) \leqq \frac{\int_0^{x_0} \exp\left\{-2\alpha ht-v^*N^{-1}+\alpha Dt^2\right\} dt}{\int_0^1 \exp\left\{-2\alpha ht-v^*N^{-1}+\alpha Dt^2\right\} dt}$$

It may be noted that the complications in the above model arise from the nature of the Markovian variate, which is a rational function of a_t and b_t rather than a polynomial. Therefore for the purposes of obtaining bounds it would be simpler to consider an equivalent model where $p(A)$, defined in (7.1), is replaced by

$$p^*(A) = \tfrac{1}{2}(1+a_t-b_t) + \tfrac{1}{2}s\left\{a_t-a_t^2+a_tb_t+(b_t-a_t)h(1-a_t-b_t)\right\}$$

which differs from $p(A)$ by terms of order s^2. This difference is very small and of the same order of magnitude as approximations already implicit in (7.1), for example replacing a selective advantage of $(1+s)^h$ by $1+sh$. This is so because as has been remarked in Chapter 2, it is the ratio between selective advantages which matters and not the difference. Thus a selective advantage mid-way between $(1+s)$ and 1 is not $1+\frac{1}{2}s$ but $(1+s)^{\frac{1}{2}}$.

By using the above model with $p(A)$ replaced by $p^*(A)$ much more precise bounds could be found.

CHAPTER 8

DIPLOID POPULATIONS WITH SELECTION DEPENDING ON GENE FREQUENCY

8.1 Introduction

The starting point in the previous chapter was provided by the recognition that in the diploid population model it was possible to find a single Markovian variate which can be used to describe the behaviour of the population. It is possible to extend this model to the situation where selective advantages are no longer constant but depend themselves on the gene frequencies in the previous generation. Thus populations can be discussed for which the presence of similar genotypes hinders any individual, as would happen for example if similar genotypes competed for scarce food or shelter peculiar to them. The behaviour of such a population can be approximated by allowing the selective advantage of any individual to decrease as the proportion of individuals of similar genotype increases. Conversely, the case where the presence of similar genotypes favours any individual can be considered by allowing the selective advantage to increase with the proportion of individuals of similar genotype.

As in the previous chapter the existence of a Markovian variate is used solely to justify the use of diffusion methods, and for the derivation of bounds. Exact results would be extremely difficult to derive, due to the size and complexity of the transition matrix, and in any case the diffusion approximations may not only be expected to be close but are also relatively simple functions allowing simple interpretation of the effects of the frequency-dependent factors.

The simplest and perhaps most useful case is where the selective advantages depend linearly on gene frequencies, and this is the case considered at length here. The case of more general selective advantages is considered more briefly and a general formula only, without any specific examples, is given.

Since only ratios of selective advantages are important it is always possible to let one of these be unity, and for convenience the heterozygote is taken as having unit selective advantage from now on.

The quantities discussed are survival probabilities, mean absorption times, pseudo-transient distributions and stationary distributions in the case where a small amount of mutation in both directions is allowed. In the case where all genotypes have unit selective advantage the exact probability that the population eventually consists entirely of individuals being (say) AA is equal to the initial proportion of A genes, a result which is also given by the diffusion approximation. This provides a standard against which survival probabilities may be measured.

Before considering finite populations it is interesting to consider the (deterministic) behaviour of gene frequencies in infinite populations with selection depending on gene frequency. Suppose that the three genotypes AA, Aa, and aa occur in proportions given by the Hardy-Weinberg law. We are interested in the proportion p of A genes, and consider first the case discussed by Wright (1948) and Moran (1962) where the selective advantages

are $1-s+tq$, 1, $1+s-tq$ respectively, where $q = 1-p$, and s and t are both taken to be small and positive. Then the increase in p from one generation to the next is

$$\Delta p = - \frac{pq(s-tq)}{1-(s-tq)(p-q)}$$

If $s > t$, then $p \to 0$, but if $s < t$, then $p \to 1-st^{-1}$ and this latter point is one of stable equilibrium. On the other hand if the selective advantages are $1-s+tk$, 1, $1+s-tk$ respectively, where k is any constant, then there are no equilibrium points except $p = 0,1$. Thus the introduction of frequency-dependent selective advantages may lead to a non-trivial equilibrium point where no such point can exist for corresponding fixed selective advantages. However this will not necessarily happen, as is demonstrated in the situation where the selective advantages are $1+s+tp$, 1, $1+s+tq$, respectively. Here it is found that $\Delta p = 0$ for $p = 0$, $\frac{1}{2}$ or 1, but that the equilibrium at the point $p = \frac{1}{2}$ is unstable.

8.2 Absorption probabilities

For the finite population case we suppose that the population size is constant and equal to N (N large). Suppose that in generation i there are k_i AA individuals and ℓ_i aa individuals, and we put $a_i = k_i N^{-1}$ and $b_i = \ell_i N^{-1}$. We suppose also that the selective advantages of AA, Aa, and aa individuals are

$$1 - s_1 + t_1 q_i,\ 1,\ 1 + s_2 - t_2 q_i$$

respectively, where s_1, s_2, t_1 and t_2 are small (i.e. of order N^{-1}) and q_i is the proportion of a genes in generation i. Then by considering the model analogous to that of the previous section, the number of A genes in the $(i+1)^{th}$ generation is a binomial variate with index 2N and parameter

$$\Pi_i = \frac{a_i\left\{1 - s_1 + \frac{1}{2}t_1(1+b_i-a_i)\right\} + \frac{1}{2}\left\{1-a_i-b_i\right\}}{a_i\left\{1-s_1+\frac{1}{2}t_1(1+b_i-a_i)\right\} + (1-a_i-b_i)+b_i\left\{1+s_2-\frac{1}{2}t_2(1+b_i-a_i)\right\}}$$

$$= \frac{\frac{1}{2}(1+a_i-b_i) - s_1a_i + \frac{1}{2}t_1a_i(1+b_i-a_i)}{1-s_1a_i+s_2b_i+\frac{1}{2}t_1a_i(1+b_i-a_i) - \frac{1}{2}t_2b_i(1+b_i-a_i)}$$

Π_i will be called the "effective" proportion of A genes in constant to $p_i = \frac{1}{2}(1+a_i-b_i)$, the actual proportion. Π_i and p_i differ by terms of order N^{-1}, but as in the previous chapter Π_i is a Markovian variate whereas p_i is not, so that attention is concentrated on Π_i. In order to use the methods of Chapter 3 we expand the denominator of Π_i in series, and ignoring terms which are $o(N^{-1})$ we obtain

$$\Pi_i = \tfrac{1}{2}(1+a_i-b_i) - s_1a_i + \tfrac{1}{2}t_1a_i(1+b_i-a_i)$$

$$- \tfrac{1}{2}(1+a_i-b_i)\left\{-s_1a_i+s_2b_i+\tfrac{1}{2}t_1a_i(1+b_i-a_i)-\tfrac{1}{2}t_2b_i(1+b_i-a_i)\right\}$$

Therefore, since $E(a_{i+1}) = \Pi_i^2$, $E(b_{i+1}) = (1-\Pi_i)^2$

$$V(a_{i+1}),\ V(b_{i+1}) \text{ are } o(N^{-1}),$$

(where all expectations are conditional on Π_i),

we have, to the same order of accuracy

$$E(\Pi_{i+1}) = \Pi_i - s_1\Pi_i^2 + t_1\Pi_i^2(1-\Pi_i) - \Pi_i\left\{-s_1\Pi_i^2 + s_2(1-\Pi_i)^2 + t_1\Pi_i^2(1-\Pi_i) - t_2(1-\Pi_i)^3\right\}$$

Thus if $\delta_{i+1} = \Pi_{i+1} - \Pi_i$, to order N^{-1}

$$E(\delta_{i+1}) = \Pi_i(1-\Pi_i)\left\{-s_1\Pi_i + t_1\Pi_i(1-\Pi_i) - s_2(1-\Pi_i) + t_2(1-\Pi_i)^2\right\}$$

Also $V(\delta_{i+1}) = V(\Pi_{i+1})$

$$= \Pi_i(1-\Pi_i)(2N)^{-1}$$

to the same order of accuracy. We now use equation (3.9) to obtain for the diffusion approximation $P(\Pi_0)$ to the probability that the whole population eventually consists of AA individuals the expression

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{\alpha_1 x^2 - \alpha_2(1-x)^2 - 2\beta_1(\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3) + \tfrac{2}{3}\beta_2(1-x)^3\right\} dx}{\int_0^1 \exp\left\{\alpha_1 x^2 - \alpha_2(1-x)^2 - 2\beta_1(\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3) + \tfrac{2}{3}\beta_2(1-x)^3\right\} dx} \qquad (8.1)$$

where $\alpha_1 = 2Ns_1$, $\beta_1 = 2Nt_1$, $\alpha_2 = 2Ns_2$, $\beta_2 = 2Nt_2$.

It is worth noting that in the particular case $\beta_1 = \beta_2 = \beta$ say, i.e. when the coefficients of the frequency-dependent factors are the same for both homozygotes,

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{\alpha_1 x^2 - \alpha_2(1-x)^2 + \beta(1-x)^2\right\} dx}{\int_0^1 \exp\left\{\alpha_1 x^2 - \alpha_2(1-x)^2 + \beta(1-x)^2\right\} dx} \qquad (8.2)$$

Thus $P(\Pi_0)$ is similar in form to the values obtained in the previous chapter where the selective advantages are fixed. In fact it will be shown later that it will frequently be possible to identify, so far as diffusion approximations are concerned, populations with selection depending linearly on gene frequency with some population having fixed selective advantages.

We may now consider the various forms that (8.1) and (8.2) may take for various selective advantages.

8.3 Particular Cases

Case 1.

The selective advantages

$$1+tq,\ 1,\ 1+tp$$

correspond to $\alpha_1 = 0$, $\alpha_2 = \beta_1 = \beta_2 = \beta$. For positive t these selective advantages could be used to consider the behaviour of a population for which like genotypes compete with like genotypes, since the selective advantage of each homozygote decreases steadily as the frequency of the corresponding gene increases. By inserting the above values in (8.1) or (8.2) it is found that

$$P(\Pi_0) = \Pi_0$$

so that survival probabilities are independent of t and are the same as for the case when no selection operates. This is a particular example of a more general case considered later. The result is, of course, true only to the order of magnitude provided

by diffusion methods, but it would be possible to derive exact bounds along the lines considered in the next section. It is often of interest to find survival probabilities for a single initial (mutant) gene, and taking this to be A in the present example the survival probability is clearly $(2N)^{-1}$.

Case 2

If we put

$$\alpha_1 = -\alpha,\ \alpha_2 = 0,\ \beta_1 = -\alpha,\ \beta_2 = \beta \tag{8.3}$$

the selective advantages become

$$1+sp,\ 1,\ 1+tq$$

If $s, t > 0$ the presence of like genotypes favours like genotypes. If $s,t < 0$ the selective advantage of each genotype decreases as the proportion of the corresponding gene increases, but it will turn out that the results are not the same as those of Case 1 above. Inserting the values (8.3) in (8.1) we obtain

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{-\frac{2}{3}\alpha x^3 - \frac{2}{3}\beta(1-x)^3\right\}}{\int_0^1 \exp\left\{-\frac{2}{3}\alpha x^3 - \frac{2}{3}\beta(1-x)^3\right\}} \tag{8.4}$$

In the case $\alpha \neq \beta$ a qualitative examination of (8.4) shows that whenever $\alpha > \beta$, then $P(\Pi_0) > \Pi_0$, for all Π_0, while for the case $\alpha < \beta$ the opposite is true. Both these results would be expected qualitatively but numerical values would be harder to obtain.

In the case of a single initial mutant, to a close approximation,

$$P\left\{(2N)^{-1}\right\} = \frac{(2N)^{-1}\ \exp\left(-2/3\ \beta\right)}{\int_0^1 \exp\left\{-\frac{2}{3}\alpha x^3 - \frac{2}{3}\beta(1-x)^3\right\}}\ .$$

The particular case $\alpha = \beta$ corresponds to selective advantages 1+sp, 1, 1+sq, and equation (8.4) reduces to

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{2\alpha x(1-x)\right\}\ dx}{\int_0^1 \exp\left\{2\alpha x(1-x)\right\}\ dx} \qquad (8.5)$$

so that $P(1-\Pi_0) = 1-P(\Pi_0)$ and $P(\frac{1}{2}) = \frac{1}{2}$ for all α. It is possible to calculate $P(\Pi_0)$ for various α and Π_0, and some typical values are given below in Table 8.1. Because of the symmetry only values of $\Pi_0 \geq \frac{1}{2}$ are considered.

Table 8.1

Values of $P(\Pi_0)$ (Equation 8.5) for various Π_0 and α

Π_0	α =16	α= 8	α= 4	α= 2	α=.5	α=-.5	α=-2	α=-4	α=-8
1.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.95	.9999	.9969	.9862	.9728	.9568	.9425	.9154	.8704	.7687
.90	.9993	.9905	.9664	.9424	.9115	.8875	.8449	.7803	.6571
.85	.9975	.9784	.9392	.9022	.8644	.8347	.7848	.7148	.5984
.80	.9918	.9593	.9033	.8599	.8157	.7838	.7326	.6668	.5650
.75	.9773	.9233	.8576	.8088	.7654	.7342	.6863	.6264	.5444
.70	.9452	.8728	.8019	.7554	.7139	.6860	.6444	.5946	.5306
.65	.8850	.8033	.7365	.6959	.6613	.6386	.6057	.5674	.5207
.60	.7882	.7152	.6628	.6328	.6080	.5920	.5693	.5434	.5128
.55	.6554	.6119	.5830	.5670	.5542	.5459	.5343	.5213	.5062
.50	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

If the values in this table are graphed it is noted that the curve of $P(\Pi_0)$ against Π_0 tends to be very flat for intermediate values of Π_0 and steep at the ends for negative α. For positive α, $P(\Pi_0)$ is very sensitive to Π_0 for intermediate values but is flat in the extremities. This behaviour is expected; when the selective advantage decreases as the number of corresponding genes increases it does not matter much what the initial value of Π is, so long as it is not too near either boundary and so long as $-\alpha$ is at all large (greater than 10 or 12). This is so because there will be a strong tendency for Π to drift to $\frac{1}{2}$. In this case a large absorption time would also be expected. When α is positive the initial value of Π is important since there will be a tendency for Π to drift towards the closer boundary. Further, once Π is reasonably near one or other boundary, then it is very likely that that boundary will be reached rather than the other, so that the curve of $P(\Pi_0)$ tends to be flat and close to the values 0,1, in the neighbourhood of the boundaries $x = 0$, $x = 1$ respectively. We further expect a relatively smaller mean absorption time. If there is only a single initial mutant, then

$$P(\Pi_0) = \frac{(2N)^{-1}}{\int_0^1 \exp\left\{ 2\alpha x(1-x) \right\} dx} \qquad (8.6)$$

This is less than $(2N)^{-1}$ for positive α, being $.0413(2N)^{-1}$ for $\alpha = 8$. This shows the importance of the initial value of Π, for in this example there is a symmetrical relation between A and a

and yet the absorption probability differs greatly from $(2N)^{-1}$ even in the example considered. The large initial selective disadvantage thus influences the probability markedly.

Case 3.

If $t_1 = t_2 = 0$ there are no frequency-dependent selective advantages and the case of selection with dominance is obtained. Putting $s_1 = s(h-1)$ and $s_2 = -sh$ the selective advantages are equivalent to (to order s^2)

$$1+s,\ 1+sh,\ 1$$

and inserting in (8.1) or (8.2) we obtain

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp(-2\alpha hx + \alpha Dx^2)\ dx}{\int_0^1 \exp(-2\alpha hx + \alpha Dx^2)\ dx} \qquad D = 2h-1 \qquad (8.7)$$

which agrees with the result of the previous chapter.

Case 4.

The case of complete dominance is obtained by putting $s_1 = t_1 = 0$. Using (8.1) we obtain eventually

$$P(\Pi_0) = \frac{\int_{1-\Pi_0}^1 \exp\left\{-\alpha y^2 + \frac{2}{3}\beta y^3\right\}\ dy}{\int_0^1 \exp\left\{-\alpha y^2 + \frac{2}{3}\beta y^3\right\}\ dy}$$

where $y = 1-x$, $\alpha_2 = \alpha$, $\beta_2 = \beta$ for convenience. The nature of the curve of $P(\Pi_0)$ against Π_0 can take various forms and is best

examined by examples.

Example (i): ($\alpha=0$)

In this case

$$P(\Pi_0) = \frac{\int_{1-\Pi_0}^{1} \exp\left(\frac{2}{3}\beta y^3\right) dy}{\int_{0}^{1} \exp\left(\frac{2}{3}\beta y^3\right) dy}$$

and clearly $P(\Pi_0) > \Pi_0$ for $\beta > 0$, while $P(\Pi_0) < \Pi_0$ for $\beta < 0$.

Example (ii): ($\alpha = \frac{1}{3}\beta$)

Here we have

$$P(\Pi_0) = \frac{\int_{1-\Pi_0}^{1} \exp\left\{\frac{1}{3}\beta(2y^3-y^2)\right\} dy}{\int_{0}^{1} \exp\left\{\frac{1}{3}\beta(2y^3-y^2)\right\} dy}$$

and it may be shown that again $P(\Pi_0) > \Pi_0$ for $\beta>0$ while $P(\Pi_0) < \Pi_0$ for $\beta < 0$.

Example (iii): ($\alpha=\frac{2}{3}\beta$).

Here

$$P(\Pi_0) = \frac{\int_{1-\Pi_0}^{1} \exp\left\{-\frac{2}{3}\beta y^2(1-y)\right\} dy}{\int_{0}^{1} \exp\left\{-\frac{2}{3}\beta y^2(1-y)\right\} dy}$$

$P(\Pi_0)$ has the property that for positive β, $P(\Pi_0)$ is greater than Π_0 for small Π_0 but less than Π_0 for larger Π_0. For negative β the converse holds. Thus the selective factors have a marked effect on the curve.

Example (iv): $(\alpha = \beta)$.

In this case

$$P(\Pi_0) = \frac{\int_{1-\Pi_0}^{1} \exp\left\{ \beta(\tfrac{2}{3} y^3 - y^2) \right\} dy}{\int_{0}^{1} \exp\left\{ \beta(\tfrac{2}{3} y^3 - y^2) \right\} dy}$$

and it is readily shown that $P(\Pi_0) < \Pi_0$ for $\beta > 0$ while $P(\Pi_0) > \Pi_0$ for $\beta < 0$. This contrasts markedly with the behaviour of $P(\Pi_0)$ in Examples (i) and (ii) and Example (iii) is one of transition between the two types. In all examples it is easy to approximate to $P\left\{ (2N)^{-1} \right\}$. The value of α (for fixed β) for which $P\left\{(2N)^{-1}\right\} = (2N)^{-1}$ is very close to the solution of the equation

$$\exp\left(-\alpha + \tfrac{2}{3}\beta\right) = \int_0^1 \exp(-\alpha y^2 + \tfrac{2}{3}\beta y^3)\, dy$$

which has been shown to lie in $(\frac{2}{3}\beta,\ \beta)$.

Case 5.

By putting $s_1 = s_2 = t_1 = t_2 = s$, so that $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = \alpha$, the selective advantages become

$$1-sp,\ 1,\ 1+sp$$

$$\text{Then} \quad P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp(\alpha x^2)\, dx}{\int_0^1 \exp(\alpha x^2)\, dx} \qquad (8.8)$$

Here it is expected that if s is positive, then $P(\Pi_0) < \Pi_0$ since the allele A will have a smaller selective advantage than that of a irrespective of p. Similarly, for s negative it is expected that $P(\Pi_0) > \Pi_0$. Specific values may be found easily and are tabulated below for typical values of α and Π_0 (Table 8.2)

Table 8.2

Values of $P(\Pi_0)$ (Equation 8.8) for various Π_0 and α

Π_0	α=-4	α=-1	α= 1	α= 4
.95	.9975	.9741	.9104	.7250
.90	.9937	.9457	.8308	.5373
.85	.9884	.9145	.7573	.4070
.80	.9809	.8806	.6897	.3141
.75	.9706	.8439	.6274	.2467
.70	.9568	.8043	.5696	.1968
.65	.9384	.7619	.5757	.1591
.60	.9146	.7166	.4651	.1300
.55	.8843	.6687	.4176	.1072
.50	.8467	.6177	.3725	.0888
.45	.8007	.5642	.3297	.0738
.40	.7456	.5084	.2887	.0613
.35	.6810	.4502	.2494	.0506
.30	.6067	.3900	.2113	.0413
.25	.5230	.3279	.1745	.0331
.20	.4304	.2643	.1386	.0256
.15	.3302	.1994	.1033	.0188
.10	.2238	.1335	.0685	.0123
.05	.1130	.0669	.0342	.0061

It may also be noted from the table that the effect of negative values of α seems to be stronger than that of the corresponding values.

Case 6.

If we put $-s_1 = s_2 = s$, $t_1 = t_2 = t$, the selective advantages become

$$1+s+tq,\ 1,\ 1+s-tq$$

For $s > t > 0$ the homozygotes are favoured, for $0 > t > s$ the heterozygotes are favoured. Using (8.2) it follows that

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{ 2\alpha x(1-x) + \beta(1-x)^2 \right\} dx}{\int_0^1 \exp\left\{ 2\alpha x(1-x) + \beta(1-x)^2 \right\} dx}$$

An interesting case is where $s = \frac{1}{2}t$ for which the selective advantages are

$$1+t(q+\tfrac{1}{2}),\ 1,\ 1+t(q-\tfrac{1}{2})$$

and for which

$$P(\Pi_0) = \frac{1-\exp(-\beta\Pi_0)}{1-\exp(-\beta)}$$

which is remarkably similar to probabilities obtained for haploid populations, (c.f. equation 6.2). In the case $s = t$ we obtain

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp(-\beta x^2)\, dx}{\int_0^1 \exp(-\beta x^2)\, dx}$$

which is similar to the probability obtained for case 5, with β replaced by $-\beta$.

Case 7.

If $s_1 = -s-t$, $s_2 = s$, $t_1 = t_2 = -t$, then the selective advantages become

$$1+s+tp,\ 1,\ 1+s+tq$$

so that for $t > 0$, each genotype is favoured by the presence of like genotypes. Substituting in (8.2),

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{2(\alpha+\beta)\, x(1-x)\right\} dx}{\int_0^1 \exp\left\{2(\alpha+\beta)\, x(1-x)\right\} dx}$$

For $\alpha = -\beta$, $P(\Pi_0) = \Pi_0$ as in case 1, as is otherwise obvious. Note that $P(\Pi_0)$ depends on α and β only through their sum, so that by putting $\alpha + \beta = \gamma$ and $s + t = c$, so that $s = c-t$, the selective advantages are of the type

$$1 + c - tq,\ 1,\ 1 + c - tp$$

and for these selective advantages $P(\Pi_0)$ is independent of t. This

result extends that of Case 1 and enables a range of situations to be covered simultaneously. In particular by putting $t = c$ Case 2 is recovered and by putting $t = 0$ a particular case of non-frequency-dependent selection is obtained. This example shows how density-dependent selective advantages may sometimes be ignored or treated as being constant.

8.4 Identifications with fixed selective advantages

It was noted at the beginning of the last section that when $t_1 = t_2$ the survival probabilities are similar to those found in the previous chapter for selection with dominance or alternatively to those of Case 3 above. This makes it possible in some cases, to identify, so far as survival probabilities are concerned, some populations with selection depending on gene-frequency with populations with fixed selective advantages. It is convenient to write (8.7) in the form

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{ - 2\alpha^* h^* t + \alpha^*(2h^*-1)t^2 \right\} dt}{\int_0^1 \exp\left\{ - 2\alpha^* h^* t + \alpha^*(2h^*-1)t^2 \right\} dt} \qquad (8.9)$$

for convenience in identification. Identifying this with the value found for Case 1 it is necessary to put $\alpha^* = 0$ giving all selective advantages unity, as has been observed. To identify (8.9) with (8.5) it is necessary to put

$$2\alpha = -2\alpha^* h^*$$

$$2\alpha = -\alpha^*(2h^*-1)$$

and clearly for $\alpha \neq 0$ these equations cannot be solved for α^*, h^*, so that identification is not possible.

In case 5 identification is possible if we put

$$\alpha = \alpha^*(2h^*-1)$$

$$0 = -2\alpha^* h^*$$

Thus $h^* = 0$, $\alpha = -\alpha^*$, so that it is possible to identify a population having selective advantages 1-s, 1, 1 with one having selective advantages 1-sp, 1, 1+sp.

For Case 6 the identification gives

$$2\alpha - 2\beta = -2\alpha^* h^*$$

$$-2\alpha + \beta = \alpha^*(2h^*-1)$$

These equations give $\alpha^* = \beta$, $h^* = 1 - \alpha\beta^{-1}$, so that a population with selective advantages 1+s+tq, 1, 1+s-tq may be identified with a population having selective advantages 1+t, 1+t-s, 1.

The reason why such an identification is not always possible is that with selective advantages of the form 1+s, 1+sh, 1, whenever the selective advantage of AA individuals is unity the selective advantage of Aa individuals is also unity.

8.5 Bounds

Since terms of order N^{-2} have been ignored, the previous results are only close approximations, and bounds may be found within which the true fixation probability lies. The main result

is that these bounds may be obtained by adding a term of order N^{-1} to the diffusion approximation. As an example, consider Case 5 where it is supposed that $\alpha > 0$. Suppose it is possible to find a function $\phi^*(\Pi_i)$ of the form

$$\phi^*(\Pi_i) = \int^{\Pi_i} \exp(\alpha x^2 + \epsilon x)\, dx \tag{8.10}$$

where ϵ is $O(N^{-1})$, for which

$$E\phi^*(\Pi_{i+1}) \leqq \phi^*(\Pi_i). \tag{8.11}$$

Then from equation (3.51), such a function may be used to provide an upper bound for the true fixation probability. Note that all expectations are conditional on Π_i. Then (8.11) may be rewritten

$$E \int_{\Pi_i}^{\Pi_{i+1}} \exp(\alpha x^2 + \epsilon x)\, dx < 0$$

or

$$E \int_0^{\delta_{i+1}} \exp(\alpha y^2 + 2\alpha\Pi_i y + \epsilon y)\, dy < 0 \tag{8.12}$$

Letting

$$\psi(\delta) = \int_0^{\delta} \exp\,(\alpha y^2 + 2\alpha\Pi_i y + \epsilon y)dy,$$

we obtain

$$\psi'(\delta) = \exp\,(\alpha\delta^2 + 2\alpha\Pi_i\delta + \epsilon\delta)$$

$$\psi''(\delta) = (\epsilon + 2\alpha\delta + 2\alpha\Pi_i)\, \exp\,(\epsilon\delta + \alpha\delta^2 + 2\alpha\Pi_i\delta)$$

$$\psi'''(\delta) = \left[(\epsilon+2\alpha\delta+2\alpha\Pi_i)^2 + 2\alpha \right] \exp\ (\epsilon\delta+\alpha\delta^2+2\alpha\Pi_i\delta)$$

$$\psi^{(iv)}(\lambda\delta) = \left[(\epsilon+2\alpha\lambda\delta+2\alpha\Pi_i)^3 + 6\alpha(\epsilon+2\alpha\lambda\delta+2\alpha\Pi_i) \right] \exp(\epsilon\delta+\alpha\delta^2+2\alpha\Pi_i\delta)$$

so that $\psi(0) = 0$, $\psi'(0) = 1$, $\psi''(0) = \epsilon+2\alpha\Pi_i$, $\psi'''(0) = (\epsilon+2\alpha\Pi_i)^2 + 2\alpha$

The left-hand side in (8.12) may be written

$$E\left[\psi(0) + \delta\psi'(0) + \frac{\delta^2}{2}\psi''(0) + \frac{\delta^3}{6}\psi'''(0) + \frac{\delta^4}{24}\psi^{(iv)}(\lambda\delta) \right] \qquad (8.13)$$

where λ is a function of δ_{i+1} and lies in $(0,1)$.

Since ϵ is $O(N^{-1})$ we have

$$| \psi'''(0) | < 8\alpha^2 + 2\alpha$$

$$| \psi^{(iv)}(\lambda\delta_{i+1}) | < (72\alpha^3 + 24\alpha^2) \exp(3\alpha).$$

By expanding out Π_{i+1} we find

$$\Pi_{i+1} = \tfrac{1}{2}(1+a_{i+1}-b_{i+1}) - \tfrac{1}{2}sa_{i+1}(1+a_{i+1}-b_{i+1})$$

$$- \tfrac{1}{4}s(b_{i+1}-a_{i+1})(1+a_{i+1}-b_{i+1})^2 + s^2\,\xi(a_{i+1},b_{i+1}) \qquad (8.14)$$

where $|\xi(a_{i+1},b_{i+1})| < 4$

and $E\left[\xi(a_{i+1},b_{i+1})\right]$ is a rational function of Π_i and vanishes at $\Pi_i = 0,1$.

$$\text{Hence } E(\delta_{i+1}) = -s\Pi_i^2(1-\Pi_i) + R_1\Pi_i(1-\Pi_i)$$

where $|R_1| < 4\alpha^2 N^{-2} + \alpha N^{-2}$

the latter term in the bound for R_1 arising from covariance terms when expectations are taken in (8.14). Similarly

$$V(\delta_{i+1}) = \Pi_i(1-\Pi_i)(2N)^{-1} + R_2\Pi_i(1-\Pi_i)$$

where

$$|R_2| < 2\alpha N^{-2}$$

so that

$$E(\delta_{i+1}^2) = \Pi_i(1-\Pi_i)(2N)^{-1} + R_3\,\Pi_i(1-\Pi_i)$$

where

$$|R_3| < \tfrac{1}{4}\alpha^2 N^{-2} + 2\alpha N^{-2}$$

Also $E(\delta_{i+1}^3) = \tfrac{1}{2}\Pi_i(1-\Pi_i)N^{-2} + R_4\Pi_i(1-\Pi_i)$

where $|R_4| < \alpha N^{-2}$

and $|E(\delta_{i+1}^4)| < \tfrac{1}{4}\Pi_i(1-\Pi_i)N^{-2}$

Thus (8.13) may be written

$$\Pi_i(1-\Pi_i)\left[\ \epsilon(4N)^{-1} + R_5\right]$$

where $|R_5| < N^{-2}\left[2\alpha^3 + 10\alpha^2 + 2\alpha + \tfrac{1}{4}(3\alpha^3 + \alpha^2)\exp(3\alpha)\ \right]$

Thus by choosing ϵ equal to

$$N^{-1}\left[8\alpha^3 + 40\alpha^2 + 8\alpha + (3\alpha^3 + \alpha^2)\exp(3\alpha)\ \right] = \epsilon^* \qquad (8.15)$$

equation (8.11) will be satisfied, and similarly by putting $\epsilon = -\epsilon^*$ the equation will hold with the inequality sign reversed. Thus an upper bound for the exact probability is

$$\frac{\int_0^{\Pi_0} \exp\left(-\epsilon^* x + \alpha x^2\right) dx}{\int_0^1 \exp\left(-\epsilon^* x + \alpha x^2\right) dx}$$

and a lower bound is

$$\frac{\int_0^{\Pi_0} \exp\left(\epsilon^* x + \alpha x^2\right) dx}{\int_0^1 \exp\left(\epsilon^* x + \alpha x^2\right) dx}$$

8.6 Arbitrary Selective Advantages

The above methods may be generalized immediately to the situation where the selective advantages are arbitrary functions of gene frequency, provided that these functions are sufficiently well-behaved in some sense. We suppose that the selective advantages are $1+s\xi_1(p)$, 1, $1+s\xi_2(p)$. Then the probability that a gene chosen at random from the $(i+1)^{th}$ generation will be A is

$$\Pi_i = \frac{a_i\left\{1 + s\,\xi_1(p_i)\right\} + \frac{1}{2}(1 - a_i - b_i}{1 + sa_i\,\xi_1(p_i) + sb_i\,\xi_2(p)}$$

Ignoring terms of order s^2, it follows that

$$\Pi_{i+1} = a_{i+1}\left\{1+s\xi_1(p_{i+1})\right\} + \tfrac{1}{2}(1-a_{i+1}-b_{i+1})$$

$$- \tfrac{1}{2}s(1+a_{i+1}-b_{i+1})\left\{ a_{i+1}\xi_1(p_{i+1}) + b_{i+1}\,\xi_2(p_{i+1})\right\}$$

Now terms of order s^2 are ignored for the moment, so that we may put, for instance,

$$E\left\{ s\, a_{i+1}\, \xi_1(p_{i+1}) \right\} = sE(a_{i+1})\, \xi_1(Ep_{i+1})$$

This gives

$$E(\delta_{i+1}) = s\Pi_i(1-\Pi_i)\left\{ \Pi_i\, \xi_1(\Pi_i) - (1-\Pi_i)\, \xi_2(\Pi_i) \right\}$$

and $V(\delta_{i+1}) = \Pi_i(1-\Pi_i)(2N)^{-1}$ to the same order of accuracy.

Inserting these values in (3.9) we find as the diffusion approximation for the probability of fixation of A genes the expression

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left[- 2\alpha \int^x t\, \xi_1(t) - (1-t)\, \xi_2(t)\, dt \right] dx}{\int_0^1 \exp\left[- 2\alpha \int^x t\, \xi_1(t) - (1-t)\, \xi_2(t)\, dt \right] dx} \qquad (8.16)$$

From this it follows immediately that if there exists a function $g(p)$ for which

$$\xi_1(p) = (1-p)\, g(p)$$

$$\xi_2(p) = pg(p)$$

Then $P(\Pi_0) = \Pi_0$. This extends the result of Case 1 which has

$g(p)$ = constant. Finally $\xi_1(p)$ and $\xi_2(p)$ will be "sufficiently well behaved" at least if $\xi_1(p)$, $\xi_2(p)$ and their derivatives are all $O(1)$ for p in $(0,1)$.

8.7 Stationary Distributions

When a small amount of mutation in each direction is allowed a stationary distribution will result, given closely by equation (5.35). It is supposed that the effect of mutation is such that the probability that a gene chosen at random in the $(i+1)^{th}$ generation is A is no longer Π_i but $\Pi_i - \lambda \Pi_i + \mu(1 - \Pi_i)$, ($\lambda$, μ of order N^{-1}). This corresponds to mutation at rate λ from A to a and μ from a to A. This gives

$$f(\Pi) = \text{const } \Pi^{4N\mu-1} (1-\Pi)^{4N\lambda-1} \exp\left[- \xi(\Pi) \right] \qquad (8.17)$$

where $\xi(\Pi)$ is that function for which

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \exp\left\{ \xi(x) \right\} dx}{\int_0^1 \exp\left\{ \xi(x) \right\} dx}$$

This shows immediately that if $P(\Pi_0)$ is independent of frequency-dependent factors, then so is $f(\Pi)$. Also, when $4N\lambda = 4N\mu = 1$,

$$P(\Pi_0) = \frac{\int_0^{\Pi_0} \left\{ f(\Pi) \right\}^{-1} d\Pi}{\int_0^1 \left\{ f(\Pi) \right\}^{-1} d\Pi}$$

From this it follows that if the proportion of A genes in the stationary distribution tends to be small, then the probability of fixation of A genes in the corresponding case without mutation tends to be high, and vice versa. Equation (8.17) may now be applied directly to derive the stationary distributions for the cases considered in section 8.3. In the case $4N\lambda = 4N\mu = 1$ it may also be noted that the stationary distributions exhibit the feature that if the presence of like genes favours like genes, the curves tend to concentrate in the extremities, while if the presence of like genes hinders like genes the curves tend to concentrate in the centre of the interval.

8.8 Mean absorption times

To discuss the mean absorption time it will be sufficient to consider only the pseudo-transient function, since the mean time is the integral of this function over (0,1), and this function gives more information by showing the transient behaviour of the process. Again exact results are difficult and it is necessary to use the diffusion approximations (3.27) and (3.28). In applying these results it should be remembered that time is measured in units of N generations.

When the selective advantages are of the form $1+s(1-p)g(p)$, 1, $1+spg(p)$, then $m(p) = 0$ and equation (3.29) holds. The mean time until homozygosity is reached is given, in terms of generations, by

$$U(\Pi_0) = -2N \left\{ \Pi_0 \ell n\, \Pi_0 + (1-\Pi_0)\, \ell n(1-\Pi_0) \right\}$$

For the other cases it is simply a matter of substitution in (3.27) and (3.28) to find the mean time. Unfortunately in most cases the resulting expressions cannot be expressed in a very simple form and numerical methods would be necessary for their evaluation.

CHAPTER 9

SELF-STERILITY POPULATIONS.

9.1 Introduction

In the preceding chapters it has been possible to use diffusion approximations for various quantities which are difficult to find exactly. However such methods may be used only under certain restrictions, and for our purposes self-sterility populations provide an example of a case where such methods cannot be applied. However some authors, in particular Fisher (1958) and Wright (1960), have used diffusion methods for these populations, and in this chapter it is shown why these methods are inapplicable and an alternative approach and alternative problem are developed.

An example of a self-sterility population sufficient for our purposes is the plant species Oenothera organensis. This species occurs in an isolated area and is not thought to total more than five hundred individuals in all. The number of alleles known is 45 (Lewis,(1948)), and it appears typical for self-sterility populations to have a large number of different alleles even in small populations. The peculiar breeding mechanism of these plants means that pollen of type A or B is unacceptable on AB styles, so that the offspring from such styles will be AX with probability $\frac{1}{2}$ and BX with probability $\frac{1}{2}$, where X is any allele other than A and B and is derived from pollen.

The problem considered by Fisher and Wright is to find how much mutation is necessary in order to maintain a given number of alleles in a population of given size. The mutation rates indicated by their analyses are of the order 2.8 per thousand for the Oenothera

population, but this high mutation rate seems to be ruled out by the failure of Lewis (1948) to find a single mutant in 220×10^6 cell divisions. Wright (1960) and Fisher (1961a) have attempted to explain this anomoly, and Fisher (1961b) has further discussed Wright's explanation.

In order to examine the applicability of diffusion methods and to discuss the previous treatments it is sufficient to discuss Wright's analysis.

9.2 Wright's Analysis

Wright (1960) supposed that the population is of size N and denoted the alleles $S_1, S_2, \ldots S_k$ with respective frequencies $q_1, q_2, \ldots q_k$. The method consists of considering the frequency q of any particular allele (say S_1). An expression was found for the mean change $\overline{\Delta q}$ of q from one generation to the next and also for the variance $\sigma^2_{\Delta q}$ of this change. By substitution in his steady-state formula (5.35) he obtained a steady-state distribution for the frequency q of any allele. This continuous approximation breaks down at q = 0, so that Wright restricted attention to "the probability distribution of alleles when present", i.e. to the distribution taken at discrete points $(2N)^{-1}, 2(2N)^{-1}, \ldots,$ and suitably normalized to give unit total probability. From this the mean number n of alleles was given by the formula $n = \bar{q}^{-1}$, where $\bar{q}$ is the mean of the discrete distribution. Bounds for the mutation rate v necessary to maintain this number of alleles were found from a formula connecting the probabilities that 0 or 1 genes of the allele in question are present and the mutation rates.

The methods and assumptions used by Wright have been questioned by Bennett (1956) and Moran (1962), p.163. The criticisms given by these authors are not exhaustive, and to them may be added the following. The "exact" value for $\sigma^2_{\Delta q}$ given by Fisher (1958) and used by Wright is $q(1-2q)(2N)^{-1}$. This is incorrect since it implies determinate behaviour when $q = \frac{1}{2}$. But it is clear that in this case the number of individuals in the next generation having S_1 as one allele is a binomial variate with parameter $\frac{1}{2}$ and index N. This follows since $q = \frac{1}{2}$ implies that each individual has exactly one allele which is S_1. Therefore S_1 pollen is not effective and the allele S_1 may only be derived from the ova, with probability $\frac{1}{2}$ for each individual.

The above criticism is not too important since the error may be corrected readily. However the next two criticisms are fundamental and show that the results derived from an analysis such as that of Wright are not meaningful. The first of these major criticisms is that diffusion methods may not be used for self-sterility populations, so that equation (5.35), which is derived by diffusion methods, is not applicable. The reason for this is that diffusion methods assume, for instance, that for all q, terms like $(\overline{\Delta q})^2$ are small in comparison with $\overline{\Delta q}$ and $\sigma^2_{\Delta q}$ and may be ignored for large N. However in self-sterility populations this assumption is not true (in contrast to populations discussed in previous chapters). The reasoning given above to show that $q(1-2q)(2N)^{-1}$ cannot be the true variance indicates this clearly. If $q = \frac{1}{2}$

then the number of individuals containing an S_1 allele is binomial with index N and parameter $\frac{1}{2}$. This leads to $\overline{\Delta q} = \frac{1}{4}$, $\sigma^2_{\Delta q} = (16N)^{-1}$. Thus $(\overline{\Delta q})^2$ is 1/16 and is certainly not small compared with either $\overline{\Delta q}$ or $\sigma^2_{\Delta q}$; in fact it exceeds $\sigma^2_{\Delta q}$ by a factor of N. It will be shown later that certain results obtained formally by using diffusion methods when these methods should not be used are completely inaccurate; thus any result for self-sterility populations derived by diffusion methods may also be expected to be inaccurate.

The second major criticism is that even if diffusion methods were allowable, then equation (5.35) would still not be relevant or meaningful, since no stationary distribution for self-sterility populations of fixed size can exist, even when mutation is allowed, contrasting with the populations considered in previous chapters. The fact that no stationary distribution can exist follows from the impossibility of forming homozygous individuals; in any generation there is positive probability that all individuals are of one genotype (S_1 S_2 say) and that no mutation occurs. After such a generation the population will automatically die out. Thus eventual extinction is certain and stationarity has no meaning. It is more meaningful to discuss pseudo-transient distributions, the mean time until the population dies out, the relevance of the values obtained and to consider the effect of mutation on these quantities.

If there are more than three alleles in the population,

mathematical discussion is very difficult since no single Markovian variate exists, and we therefore consider here only the case of three possible alleles in any detail.

9.3 A Non-Overlapping generation model

We consider in this section a population of fixed size N and admitting three alleles A,B,C. Suppose the population eventually dies out because in the final generation all individuals are BC, and suppose also that at any moment the number of AB or AC (briefly AX) individuals is i ($i > 0$). Then using a model analogous to that of Chapters 6 and 7, the number of offspring from the styles of AX individuals is a binomial variate with parameter iN^{-1} and index N. The probability that any one of these is AX is $\frac{1}{2}$, so that conditional on there being M offspring from AX styles, the number of such offspring which are AX is a binomial variate with parameter $\frac{1}{2}$ and index M. All offspring from BC styles must be AX, since A is the only allele effective on such styles. From this it follows that the probability p_{ij} that altogether there will be j AX individuals in the next generation is the coefficient of θ^j in

$$\sum_{M=0}^{N} \binom{N}{M} (iN^{-1})^N (1-iN^{-1})^{N-M} (\tfrac{1}{2}+\tfrac{1}{2}\theta)^M \theta^{N-M}$$

i.e. in $$\left[\left\{ 1-iN^{-1} \right\} \theta + i(2N)^{-1} + i\theta(2N)^{-1} \right]^N$$

which is

$$\binom{N}{j}\left\{1 - i(2N)^{-1}\right\}^{j}\left\{i(2N)^{-1}\right\}^{N-j} \tag{9.1}$$

For i = 0 the population dies out at the next generation. Thus the number of AX individuals is Markovian with transition matrix P given by

$$\begin{aligned} p_{00} &= 1 \\ p_{ij} &= \binom{N}{j}\left\{1 - i(2N)^{-1}\right\}^{j}\left\{i(2N)^{-1}\right\}^{N-j} \end{aligned} \tag{9.2}$$

Since $p_{iO} = \left\{ i(2N)^{-1} \right\}^{N}$ for $i > 0$ it follows that whatever the initial number of AX individuals, the mean time until the population dies out due to random elimination of the allele A is $\geq 2^{N}$, since each $p_{iO}^{-1} \geq 2^{N}$.

Suppose further that the initial number of AX individuals is N. Then a pseudo-transient distribution may be obtained from (9.2) by putting $p_{ON} = 1$. If the pseudo-transient distribution is $\underline{\lambda}' = (\lambda_0\ \lambda_1\ \ldots\ \lambda_N)$ then λ_0^{-1} is the mean time until the population dies out, given initially N AX individuals. If P* is the transition matrix of the amended process, then $\underline{\lambda}'$ satisfies $\underline{\lambda}' = \underline{\lambda}'P^*$, which gives in particular

$$\lambda_0 = \sum_{i=0}^{N} \lambda_i \left\{ i(2N)^{-1} \right\}^{N}$$

Thus λ_0 is $(2N)^{-N}$ times the N^{th} moment of the stationary distribution, whose mean is $\frac{2}{3}$. It follows (c.f. Loeve (1960), p.156) that $\lambda_0 \geq 3^{-N}$ so that $\lambda_0^{-1} \leq 3^N$. Thus $\lambda_0^{-1/N}$ lies in [2,3].

The method of Gani (1961) may be used to show that the latent roots θ_r of P* are given by

$$\theta_r = (-\tfrac{1}{2})^r \binom{N}{r} r!\, N^{-r} \qquad r = 0, 1, \ldots, N \qquad (9.3)$$

Thus if Q* is the submatrix of P* obtained by striking out the first row and first column of P*, we obtain

$$\lambda_0^{-1} = \left\{ |I - Q^*| \right\}^{-1} \prod_{i=1}^{N} (1-\theta_i) \qquad (9.4)$$

as a closed expression for the mean time. However this expression is not very helpful and in particular the asymptotic behaviour of $\lambda_0^{-1/N}$ does not follow readily from it. The numerical results in Table 9.1 below suggest that $\lambda_0^{-1/N}$ approaches a limit near 2.45.

Table 9.1

Population size (N)	λ_0^{-1}	$\lambda_0^{-1/N}$
2	7.0000	2.6457
3	17.3437	2.5885
4	43.0830	2.5620
5	107.1627	2.5469
6	266.7205	2.5371
7	664.0886	2.5303
8	1,653.4802	2.5252
9	4,116.6843	2.5213
10	10,247.21	2.5180
11	25,391.10	2.5143
12	62,846.63	2.5111
13	152,573.06	2.5046
14	358,806.70	2.4933

The above model has not led to simple explicit results and it is useful to consider a model allowing transitions only to neighbouring states (as in Chapters 4 and 5) for which explicit results may be found.

9.4 An overlapping-generation model

If individuals die and are replaced one by one rather than a generation at a time a transition matrix of the form considered in section 5.5 is obtained.

We suppose that individuals die at random one by one, and that the dying individual is replaced by a new individual derived from a style chosen at random from the population immediately before the birth-death event. If there are $i (i > 0)$ AX individuals before the birth-death event, then Prob. { AX individual chosen to die} = Prob. {AX style chosen for pollination} = iN^{-1}. If an AX style is pollinated the A gene is transmitted to the offspring with probability $\frac{1}{2}$. By considering the various contingencies it follows that after the birth-death event the number of AX individuals will be i-1, i, or i+1 with respective probabilities

$$\begin{aligned} p_{i,i-1} &= i^2(2N^2)^{-1} \\ p_{i,i} &= i(3N-2i)(2N^2)^{-1} \\ p_{i,i+1} &= (N-i)(2N-i)(2N^2)^{-1} \end{aligned} \tag{9.5}$$

We also have $p_{00} = 1$. The mean time taken until the population dies out takes a simple form in the case where the initial number of AX

individuals is unity, and we suppose for the moment that this is the case. Then the pseudo-transient distribution will be obtained by putting $p_{01} = 1$. In this case (9.5) holds for all i. If the new matrix is denoted P*, then the pseudo-transient function $\underline{\lambda}'$ satisfies $\lambda_0 = 1$, $\underline{\lambda}' = \underline{\lambda}'P^*$. It is readily verified from (9.5) that the solution of these equations is

$$\lambda_i = \binom{N}{i}\binom{2N}{i} \qquad i = 0, 1, \ldots, N \tag{9.6}$$

Thus since $\sum_{i=0}^{N} \lambda_i = \binom{3N}{N}$, the mean time until the population dies out is also $\binom{3N}{N}$. This is of the order $(27/4)^N$ birth-death events and is consequently extremely large even for moderate N.

When the initial number of AX individuals is arbitrary the mean time may be found by using the methods considered in Chapter 5, although in this case the expressions are more involved. By writing for convenience

$$p_{i,i-1} = \Pi_i \quad , \quad p_{i,i+1} = \eta_i$$

the transition matrix P is given by

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \Pi_1 & 1-\Pi_1-\eta_1 & \eta_1 & \cdots & 0 \\ 0 & \Pi_2 & 1-\Pi_2-\eta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\Pi_N \end{pmatrix}$$

If the initial number of AX individuals is k, then the modified matrix P* corresponding to the return process will satisfy

$$I-P^* = \begin{pmatrix} 1 & 0 & 0 & \cdots & -1 & \cdots & 0 \\ -\Pi_1 & \Pi_1+\eta_1 & -\eta_1 & \cdots & 0 & \cdots & 0 \\ 0 & -\Pi_2 & \Pi_2+\eta_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & & 0 & & \Pi_N \end{pmatrix}$$

where the -1 in the first row is in the column corresponding to k AX individuals. The pseudo- transient function $\underline{\lambda}'$ satisfies $\lambda_0 = 1$ and the typical equation

$$-\eta_{i-1}\lambda_{i-1} + (\Pi_i+\eta_i)\lambda_i - \Pi_{i+1}\lambda_{i+1} = 0 \qquad (9.7)$$

This equation holds for i = 2,3,..., k-1, k+1, ..., N-1.

Now (9.7) gives

$$\Pi_i \lambda_i - \eta_{i-1} \lambda_{i-1} = \text{constant} \qquad (9.8).$$

Since $\lambda_1 = \Pi_1^{-1}$ it follows that the constant in (9.8) is unity for i = 1, 2, ..., k. This leads to

$$\lambda_i = \Pi_i^{-1}\left[1 + \frac{\eta_{i-1}}{\Pi_{i-1}} + \frac{\eta_{i-1}\eta_{i-2}}{\Pi_{i-1}\Pi_{i-2}} + \ldots + \frac{\eta_{i-1}\eta_{i-2}\cdots\eta_1}{\Pi_{i-1}\Pi_{i-2}\cdots\Pi_1}\right] \qquad (9.9)$$

for i = 1, 2, ..., k.

For i > k, equation 9.8 still holds, but by using the equation

$$\Pi_N \lambda_N = \eta_{N-1} \lambda_{N-1}$$

it follows that the constant in (9.8) is zero for these values of i. Thus

$$\lambda_{N-1} = \lambda_N \frac{\Pi_N}{\eta_{N-1}}$$

$$\lambda_{N-2} = \lambda_N \frac{\Pi_N \Pi_{N-1}}{\eta_{N-1}\eta_{N-2}}$$

and in general

$$\lambda_i = \lambda_N \frac{\Pi_N \Pi_{N-1} \cdots \Pi_{i+1}}{\eta_{N-1}\eta_{N-2} \cdots \eta_i} \tag{9.10}$$

This holds for i = k, k+1, ..., N-1, so that identifying (9.9) and (9.10) for i = k we obtain

$$\lambda_N = \Pi_N^{-1} \left[\frac{\eta_k \cdots \eta_{N-1}}{\Pi_k \cdots \Pi_{N-1}} + \frac{\eta_{k-1} \cdots \eta_{N-1}}{\Pi_{k-1} \cdots \Pi_{N-1}} + \cdots + \frac{\eta_1 \eta_2 \cdots \eta_{N-1}}{\Pi_1 \Pi_2 \cdots \Pi_{N-1}} \right] \tag{9.11}$$

so that from (9.10) and (9.11),

$$\lambda_i = \Pi_i^{-1} \left[\frac{\eta_k \cdots \eta_{i-1}}{\Pi_k \cdots \Pi_{i-1}} + \frac{\eta_{k-1} \cdots \eta_{i-1}}{\Pi_{k-1} \cdots \Pi_{i-1}} + \cdots + \frac{\eta_1 \eta_2 \cdots \eta_{i-1}}{\Pi_1 \Pi_2 \cdots \Pi_{i-1}} \right] \tag{9.12}$$

for i = k, k+1, ..., N

Therefore the mean number of birth-death events until the population dies out is

$$\sum_{i=0}^{N} \lambda_i \tag{9.13}$$

where λ_i is defined by (9.9) for $i \leq k$ and by (9.12) for $i \geq k$.

In the case k = 1, this reduces to

$$\sum_{i=0}^{N} \binom{2N}{i}\binom{N}{i}$$

$$= \binom{3N}{N} \text{ as expected.}$$

A lower bound for (9.13) is given immediately by this value, since to get to zero from any $k > 1$ to process must pass through $k = 1$. The maximum of (9.13) occurs when $k = N$, but the corresponding value does not seem to simplify so readily.

9.5 Mutation to existing alleles

Even when mutation exists, self-sterility populations will eventually die out, and the effect of mutation on the mean dying-out time may be considered. For simplicity we suppose that all mutation rates are equal; thus the probability that the A allele in a newly formed AB individual mutates to C is α, and so on. If the number of AX individuals before a birth-death event is i ($i > 0$) it will subsequently be i - 1 if either of the following contingencies occurs; (i) AX dies, BC born, no mutation, or (ii) AX dies, AX born, A mutates. The probability of this is

$$p_{i,i-1} = i\left\{ (2N-i)\alpha + i(1-2\alpha) \right\} (2N^2)^{-1} \qquad (9.14)$$

Similarly

$$p_{i,i} = i(3N-2i)(2N^2)^{-1} + \alpha\left\{2N^2-7Ni+6i^2\right\} (2N^2)^{-1} \qquad (9.15)$$

and

$$p_{i,i+1} = (N-i)(2N-i)(2N^2)^{-1} - \alpha(N-i)(2N-3i)(2N^2)^{-1} \qquad (9.16)$$

and $p_{00} = 1$. In the case where there is a single initial A gene we amend the transition matrix by putting

$$\left.\begin{aligned} p_{00} &= \alpha \\ p_{01} &= 1-\alpha \end{aligned}\right\} \tag{9.17}$$

so that (9.14), (9.15) and (9.16) hold for all i. The pseudo-transient distribution is readily found to be

$$\lambda_0 = \frac{\Gamma\left\{N + A + 1\right\}\Gamma\left\{B + 1\right\}}{\Gamma\left\{A + 1\right\}\Gamma\left\{N + B + 1\right\}} \tag{9.18}$$

and

$$\lambda_i = \lambda_0 \binom{N}{i} \frac{\Gamma\left\{A + 1\right\}\Gamma\left\{B - A + 1\right\}}{\Gamma\left\{A + i + 1\right\}\Gamma\left\{B - A - i + 1\right\}} \tag{9.19}$$

where

$$A = 2N\alpha(1-3\alpha)^{-1}, \quad B = 2N(1-3\alpha)^{-1} .$$

Thus the mean recurrence time of the state "no AX individuals" is

$$\lambda_0^{-1} = \frac{\Gamma\left\{A + 1\right\}\Gamma\left\{N + B + 1\right\}}{\Gamma\left\{N + A + 1\right\}\Gamma\left\{B + 1\right\}} \tag{9.20}$$

Using (9.17), it is found that the mean time μ until the population dies out in the original process is the solution of

$$\lambda_0^{-1} = \mu(1-\alpha) + \alpha \ .$$

so that

$$\mu = (\lambda_0^{-1} - \alpha)(1-\alpha)^{-1} \ .$$

This follows since in the process defined by (9.17), when the state "no AX individuals" is entered, the process stays in this state with probability α or moves to the state "one AX individual" with probability $1-\alpha$. Alternatively, this formula could be derived directly by using (9.12), and using (9.20) it is found that the two methods give the same result. For the very small values of α expected, μ may be approximated by λ_0^{-1} . Using Stirling's formula,

$$\mu^{1/N} \approx \left[\frac{(2\alpha)^{2\alpha} (3-3\alpha)^{3-3\alpha}}{4(1-\alpha)^{1-\alpha}} \right]^{\frac{1}{1-3\alpha}}$$

which is very close to $27/4$ for small α . Thus mutation has a small effect on the mean time until the population dies out.

In the more general case when the initial number of AX individuals is greater than unity, the mean dying-out time is given by (9.9), (9.12) and (9.13), where Π_i and y_i are given by (9.14) and (9.16) respectively.

9.6 Comparison with haploid models

For the haploid models of Chapters 4 and 6 it is possible to introduce mutation in both directions and thus obtain stationary distributions. It will now be shown that if the mutation rates are sufficiently high, the resulting process, and thus the resulting

stationary distributions, become identical to processes and distributions considered earlier in this chapter.

The model of Chapter 6 is for non-overlapping generations with

$$P = \left\{ p_{ij} \right\} = \left\{ \binom{N}{j} p_i^j (1-p_i)^{N-j} \right\} \qquad (9.21)$$

where $p_i = iN^{-1}$ in the case of no selection. Suppose that mutation at rates α_1 from A to a and α_2 from a to A now takes place. It is then necessary to replace p_i by

$$p_i = i(1-\alpha_1)N^{-1} + (N-i)\alpha_2 N^{-1} \qquad (9.22)$$

and the process now admits a stationary distribution. Suppose the mutation rates are very high, viz. $\alpha_1 = \frac{1}{2}$, $\alpha_2 = 1$. Then

$$p_i = (2N-i)(2N)^{-1} \qquad (i = 0,1,2,..,N)$$

and the process is identical to the amended process of section 9.3; (c.f. equation (9.2) when we put $p_{ON} = 1$).

Similarly in the model of Chapter 4 (c.f. equation 4.1) the transition probabilities are

$$p_{i,i-1} = i(N-i)\, N^{-2} = p_{i,i+1}$$

$$p_{i,i} = 1-p_{i,i-1} - p_{i,i+1}$$

when there is no selection. If mutation at the above rates is allowed we will have

$$\begin{aligned} p_{i,i-1} &= iq_i N^{-1} \\ p_{i,i} &= ip_i N^{-1} + (N-i)q_i N^{-1} \\ p_{i,i+1} &= (N-i)p_i N^{-1} \end{aligned} \tag{9.23}$$

where $p_i = 1-q_i = i(1-\alpha_1)N^{-1} + (N-i)\alpha_2 N^{-1}$.

By putting $\alpha_1 = \frac{1}{2}$, $\alpha_2 = 1$ as before we obtain the amended transition matrix (9.5) of section 9.4 if we put $p_{01} = 1$. Moran (1962), p.132 has found the stationary distribution of the process defined by (9.23) for general α_1 and α_2 and has shown that if $\alpha_1 = \beta_1 N^{-1}$, $\alpha_2 = \beta_2 N^{-1}$, and if β_1 and β_2 are kept fixed as $N \to \infty$, then this stationary distribution is given asymptotically by

$$f(x) = \text{const } x^{\beta_2-1} (1-x)^{\beta_1-1}$$

This is essentially the distribution used by Wright for the "stationary distribution" of the frequency of one allele in a three-allele self-sterility population. But with $\alpha_1 = \frac{1}{2}$, $\alpha_1 = 1$ it is impossible simultaneously to keep β_1 and β_2 fixed and let $N \to \infty$. This is another way of stating that the conditions required for the application of diffusion methods are not met.

9.7 A formal pseudo-transient function

We consider the model of section 9.4 when there is a single initial AX individual. Then the methods there used show that the exact pseudo-transient function is

$$p(i) = \binom{N}{i}\binom{2N}{i} \qquad (9.24)$$

The pseudo-transient function obtained formally by diffusion methods may be found when $m(q)$ and $v(q)$ are known. It follows from (9.5) that

$$m(q) = \tfrac{1}{2}(2-3q)$$

$$v(q) = \tfrac{1}{2}N^{-1}(2-3q+2q^2)$$

These lead formally to a pseudo-transient function proportional to

$$(2-3q+2q^2)^{-1}\exp\left[-2N\int^{q}(2-3x)(2-3x+2x^2)^{-1}\,dx\right]x$$

$$\int_0^q \exp\left[2N\int^{x}(2-3y)(2-3y+2y^2)^{-1}\,dy\right]dx$$

which is not an approximation to (9.24). Therefore since the diffusion approximation to the pseudo-transient function is inappropriate, the value obtained formally by diffusion methods for the mean time until the population dies out will also be inappropriate, since this is the integral over (0,1) of the pseudo-transient function.

9.8 Extensions

Extensions to the case of more than three alleles are very difficult unless additional assumptions are made. As an example of the sort of assumptions necessary we consider a population with k alleles $S_1 \ldots S_k$ for which S_1 is the first allele lost by random

elimination. It is assumed that the alleles $S_2 \ldots S_k$ occur with equal frequencies and that each individual produces exactly one offspring. Then using a non-overlapping model similar to that of section 9.3 it is found that the mean number of generations before the allele S_1 is lost when initially each individual is of the type $S_1 S_i$ $(i = 2,\ldots,k)$, is

$$\left(\frac{k}{k-2}\right)^N$$

where N is the population size. This will be only an extremely rough approximation for the mean time derived without the above assumptions, but it is worth noting that in the Oenothera population with $N = 500$ and $k = 45$, the above expression is roughly 7×10^9 generations. The mean time until a second allele is lost, under similar assumptions, is $\left\{(k-1)(k-3)^{-1}\right\}^N$ generations, and so on. Thus an extremely long time may be expected to pass before the population dies out.

The results here obtained lead to a different explanation for the large number of alleles in self-sterility populations than that given by Fisher and Wright. It is clear that because of the peculiar breeding system in these populations, if initially a large number of different alleles existed, a large number of these alleles may still be expected to be present after an extremely long time. On the other hand it only requires a minute mutation rate to new alleles to build up a large number of different alleles, so that while mathematically the behaviour of these populations is transient,

any self-sterility population that is actually observed may be expected to contain a large number of different alleles.

REFERENCES

1. Bartlett, M.S. (1955)

An Introduction to Stochastic Processes C.U.P.

2. Barucha-Reid, A.T. (1960)

Elements of the Theory of Markov Processes and Their Applications. McGraw-Hill.

3. Bennett, J.H. (1956)

Review of C.C. Li's Population Genetics. Heredity. 4, 275-8.

4. Chung, K.L. (1960)

Markov Chains with Stationary Transition Probabilities. Springer-Verlag.

5. Doob, J.L. (1953)

Stochastic Processes. Wiley.

6. Ewens, W.J. (1963a)

The Diffusion Equation and a Pseudo-Distribution in Genetics. To appear in J.R. Statist. Soc. (B)

7. Ewens, W.J. (1963b)

Numerical Results and Diffusion Approximations in a Genetic Process. To appear in Biometrika.

8. Ewens, W.J. (1963c)

The Mean Time for Absorption in a process of Genetic Type. To appear in J.Aust.Math.Soc.

9. Ewens, W.J. (1963d)

Diploid Populations with Selection depending on Gene Frequency. To appear in J. Aust. Math. Soc.

10. Ewens, W.J. (1963e)

On the Problem of Self-Sterility Alleles. Submitted to Biometrics.

11. Ewens, W.J. and Gani, J.M. (1961)

Absorption Probabilities in some Genetic Processes. Revue Inst.Int. de Stat. 9:2 33-41.

12. Feller, W. (1951)

Diffusion Processes in Genetics. Proc. 2nd Berk. Symp. 227-246. Univ. of Cal. Press.

13. Feller, W. (1952)

The Parabolic Differential Equations and the Associated Semi-Group of Transformations. Ann. of Math. 55, 468-519.

14. Feller, W. (1954)

Diffusion Processes in One Dimension. Trans.Amer.Math.Soc. 77 (July-Dec) 1-31.

15. Feller, W. (1957)

An Introduction to Probability Theory and its Applications. Vol. 1 (2nd Ed.) Wiley.

16. Fisher, R.A. (1930)

The Genetical Theory of Natural Selection. O.U.P.

17. Fisher, R.A. (1958)

The Genetical Theory of Natural Selection
Second Edition. Dover.

18. Fisher, R.A. (1961a)

A Model for the Generation of Self-Sterility Alleles. J. Theor.Biol., 1, 411-14.

19. Fisher, R.A. (1961b)

Possible Differentiation in the Wild Population Oenothera Organensis. Aust. J. Biol.Sci., 14, 76-78.

20. Gani, J.M. (1961)

On the Stochastic Matrix in a Genetic Model of Moran. Biometrika, 48, 203-06.

21. Kemeny, J.G. and Snell, J.L. (1960)

Finite Markov Chains. Van Nostrand.

22. Kimura, M. (1955a)

Process of Random Drift with a Continuous Model. Proc.Nat.Acad.Sci., Wash. 41, 144-50.

23. Kimura, M. (1955b)

Stochastic Processes and Gene Frequencies. Cold Spr. Harb.Symp. Quant.Biol., Vol.XX. Population Genetics. 33-51.

24. Kimura, M. (1957)

Some Problems of Stochastic Processes in Genetics. Annals of Math.Stat. 28, 882-901.

25. Knox, S.R. (1962)

Ph.D. Thesis, Virginia Polytechnic Institute.

26. Kolmogorov, A.N. (1959)

The Transition of Branching Processes into Diffusion Processes and Associated Problems of Genetics. (In Russian). Teor.Veroyatnost i Primenen. 4, 233-236.

27. Lewis, D. (1948)

Structure of the Incompatibility Gene. I. Spontaneous Mutation Rate. Heredity 2, 219-36.

28. Loeve, M. (1960)

Probability Theory (Second Edition.) Van Nostrand.

29. Moran, P.A.P. (1958a)

Random Processes in Genetics. Proc. Camb. Phil. Soc., 54, 60-71.

30. Moran, P.A.P. (1958b)

A General Theory of the Distribution of Gene Frequencies. II. Non-Overlapping Generations. Proc. Roy. Soc. B, 149, 113-116.

31. Moran, P.A.P. (1960)

The Survival of a Mutant Gene under Selection, II. J.Aust. Math.Soc., 1, 485-491.

32. Moran, P.A.P. (1962)

The Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

33. Moran, P.A.P. (1963)

Some General Results on Random Walks, with Genetic Applications. To appear in J.Aust.Math.Soc.

34. Watterson, G.A. (1960)

Ph.D. Thesis, Australian National University.

35. Watterson, G.A. (1961)

Markov Chains with Absorbing States; A Genetic Example. Annals of Math.Stat., 32, 716-29.

36. Watterson, G.A. (1962)

Some Theoretical Aspects of Diffusion Theory in Population Genetics. Annals of Math.Stat., 33, 939-57.

37. Wright, S. (1931)

Evolution in Mendelian Populations. Genetics, 16, 97-159.

38. Wright, S. (1948)

On the Roles of Directed and Random Changes in Gene Frequency in the Genetics of Populations. Evolution, 2, 279-84.

39. Wright, S. (1960)

On the Number of Self-incompatibility Alleles maintained in Equilibrium by a given Mutation Rate in a Population of Given Size: a Re-examination. Biometrics, 16, 61-85.

These amendments have been made at the request of the examiners.

NOTE

I should like to thank Dr. G. A. Watterson for pointing out that the arguments on pages 60-62 and 78-80 can be simplified greatly, and a more satisfactory derivation of the results obtained. The simplifications are as follows.

Equation (4.7) holds for $i = 1,2,\ldots,k$ (rather than $i = 1,2,\ldots,k-1$ only, as given in this thesis). Also, equation (4.9) holds for $i = k, k+1,\ldots,N$ (rather than $i = k+1,\ldots N$). By equating (4.7) and (4.9) at $i = k$ we obtain directly the values for P_0 and P_N given on page 62, without reference to the results of Chapter 2. The derivation at the bottom of page 62 and the top of page 63 is now unnecessary.

Similarly, equation (5.8) also holds for $i = 2,\ldots,k$ (rather than $i = 2,\ldots,k-1$, as given in this thesis). Thus equation (5.9) holds for the same values of i. Also (5.11) holds for $i = k,\ldots,N-1$, so that the ranges in equation (5.12) are $i = 1,2,\ldots,k$ and $i = k,k+1,\ldots,N$. The two expressions in (5.12) can be equated for $i = k$, from which the values of P_0 and P_N follow immediately. The argument on page 80 is now unnecessary.

I should also like to thank both Dr. Watterson and Mr. J. E. Moyal for pointing out that the proof leading to formula (3.27) is not strict, and should be regarded as a heuristic derivation. The truth of the formula itself is not in doubt.